

*IBM SPSS Modeler 16
Benutzerhandbuch*

IBM

Hinweis

Vor Verwendung dieser Informationen und des darin beschriebenen Produkts sollten die Informationen unter „Bemerkungen“ auf Seite 187 gelesen werden.

Product Information

This edition applies to version 16, release 0, modification 0 of IBM(r) SPSS(r) Modeler and to all subsequent releases and modifications until otherwise indicated in new editions.

Diese Veröffentlichung ist eine Übersetzung des Handbuchs
IBM SPSS Modeler 16, User's Guide,
herausgegeben von International Business Machines Corporation, USA

© Copyright International Business Machines Corporation 2013

Informationen, die nur für bestimmte Länder Gültigkeit haben und für Deutschland, Österreich und die Schweiz nicht zutreffen, wurden in dieser Veröffentlichung im Originaltext übernommen.

Möglicherweise sind nicht alle in dieser Übersetzung aufgeführten Produkte in Deutschland angekündigt und verfügbar; vor Entscheidungen empfiehlt sich der Kontakt mit der zuständigen IBM Geschäftsstelle.

Änderung des Textes bleibt vorbehalten.

Herausgegeben von:
TSC Germany
Kst. 2877
Oktober 2013

Inhaltsverzeichnis

Kapitel 1. Informationen zu IBM SPSS

Modeler	1
IBM SPSS Modeler-Produkte	1
IBM SPSS Modeler	1
IBM SPSS Modeler Server	1
IBM SPSS Modeler Administration Console	2
IBM SPSS Modeler Batch	2
IBM SPSS Modeler Solution Publisher	2
IBM SPSS Modeler Server-Adapter für IBM SPSS Collaboration and Deployment Services	2
IBM SPSS Modeler-Editionen	2
IBM SPSS Modeler-Dokumentation	3
SPSS Modeler Professional-Dokumentation	3
SPSS Modeler Premium Documentation	4
Anwendungsbeispiele	5
Ordner "Demos"	5

Kapitel 2. Neue Funktionen in IBM SPSS

Modeler 16	7
-----------------------------	----------

Kapitel 3. IBM SPSS Modeler - Übersicht 11

Erste Schritte	11
Starten von IBM SPSS Modeler	11
Starten über die Befehlszeile	11
Verbinden mit IBM SPSS Modeler Server	12
Ändern des temporären Verzeichnisses	14
Starten mehrerer IBM SPSS Modeler-Sitzungen	14
Schnittstelle von IBM SPSS Modeler auf einen Blick	15
Streamerstellungsbereich von IBM SPSS Modeler	15
Knotenpalette.	16
IBM SPSS Modeler-Manager.	17
IBM SPSS Modeler-Projekte	19
IBM SPSS Modeler-Symbolleiste	19
Anpassen der Symbolleiste	20
Anpassen des IBM SPSS Modeler-Fensters	21
Ändern der Symbolgröße für einen Stream	22
Verwenden der Maus in IBM SPSS Modeler	22
Verwenden von Tastenkombinationen	23
Drucken	24
Automatisieren von IBM SPSS Modeler	24

Kapitel 4. Data-Mining 25

Data-Mining - Übersicht	25
Bewerten der Daten	26
Eine Strategie für Data-Mining	27
Prozessmodell "CRISP-DM"	28
Typen von Modellen	29
Data-Mining-Beispiele	35

Kapitel 5. Erstellen von Streams. 37

Erstellen von Streams - Übersicht	37
Erstellen von Datenstreams	37
Arbeiten mit Knoten	37

Arbeiten mit Streams	43
Streambeschreibungen.	55
Ausführen von Streams	57
Arbeiten mit Modellen	57
Hinzufügen von Kommentaren und Anmerkungen zu Knoten und Streams	58
Speichern von Datenstreams.	64
Laden von Dateien	65
Zuordnen von Datenstreams.	66
Tipps und Tastenkombinationen	68

Kapitel 6. Umgang mit fehlenden Werten 71

Fehlende Werte - Übersicht	71
Umgang mit fehlenden Werten	72
Behandeln von Datensätzen mit fehlenden Werten	72
Behandeln von Feldern mit fehlenden Werten	72
Imputieren oder Auffüllen fehlender Werte	73
CLEM-Funktionen für fehlende Werte	73

Kapitel 7. Erstellen von CLEM-Ausdrücken. 75

Informationen zu CLEM	75
CLEM - Beispiele	75
Werte und Datentypen	77
Ausdrücke und Bedingungen	78
Stream-, Sitzungs- und Superknotenparameter.	78
Arbeiten mit Zeichenfolgen	79
Umgang mit Leerzeichen und fehlenden Werten	80
Arbeiten mit Zahlen	80
Arbeiten mit Zeit- und Datumsangaben	80
Zusammenfassen mehrerer Felder	81
Arbeiten mit Mehrfachantwortdaten	82
Expression Builder	83
Zugriff auf Expression Builder	83
Erstellen von Ausdrücken	83
Auswählen von Funktionen	84
Auswählen von Feldern, Parametern und globalen Variablen	84
Anzeigen oder Auswählen von Werten	85
Überprüfen von CLEM-Ausdrücken	85
Suchen und Ersetzen	85

Kapitel 8. CLEM-Sprachreferenz 89

CLEM-Referenz - Überblick	89
CLEM-Datentypen	89
Ganze Zahlen	89
Reelle Zahlen.	90
Zeichen.	90
Zeichenfolgen	90
Listen	91
Felder	91
Datumsangaben	91
Zeit	92

CLEM-Operatoren	93
Funktionsreferenz	95
Konventionen bei Funktionsbeschreibungen	96
Informationsfunktionen	96
Konvertierungsfunktionen	97
Vergleichsfunktionen	98
Logische Funktionen	100
Numerische Funktionen	101
Trigonometrische Funktionen	102
Wahrscheinlichkeitsfunktionen	102
Bitweise Operationen mit ganzen Zahlen	103
Zufallsfunktionen	104
Zeichenfolgefunktionen	105
SoundEx-Funktionen	110
Datums- und Uhrzeitfunktionen	110
Sequenzfunktionen	115
Globale Funktionen	120
Funktionen zum Umgang mit Leerstellen und Nullwerten	121
Sonderfelder	122

Kapitel 9. Verwenden von IBM SPSS Modeler mit einem Repository 125

Informationen zu IBM SPSS Collaboration and Deployment Services Repository	125
Speichern und Bereitstellen von Repository-Objekten	126
Verbinden mit dem Repository	127
Eingabe der Berechtigungsnachweise für das Repository	127
Durchsuchen des Repository-Inhalts	127
Speichern von Objekten im Repository	128
Festlegen der Objekteigenschaften	128
Speichern von Streams	130
Temporäres Speichern von Projekten	131
Speichern von Knoten	131
Speichern von Ausgabeobjekten	131
Speichern von Modellen und Modellpaletten	132
Abrufen von Objekten aus dem Repository	132
Auswählen eines abzurufenden Objekts	133
Auswählen einer Objektversion	133
Suchen nach Objekten im Repository	134
Ändern von Repository-Objekten	135
Erstellen, Umbenennen und Löschen von Ordnern	135
Sperren und Entsperren von Repository-Objekten	135
Löschen von Repository-Objekten	136
Verwalten der Eigenschaften von Repository-Objekten	136
Anzeigen von Ordneigenschaften	136
Anzeigen und Bearbeiten von Objekteigenschaften	137
Verwalten von Objektversionsbeschriftungen	138
Bereitstellen von Streams	139
Bereitstellungsoptionen für den Stream	139
Scoring-Verzweigung	141

Kapitel 10. Exportieren in externe Anwendungen 145

Exportieren in externe Anwendungen	145
Öffnen eines Streams in IBM SPSS Modeler Advantage	145
Importieren und Exportieren von Modellen als PMML	146
Modelltypen, die PMML unterstützen	146

Kapitel 11. Projekte und Berichte 149

Einführung in Projekte	149
CRISP-DM-Ansicht	149
Klassenansicht	150
Erstellen von Projekten	150
Erstellen eines neuen Projekts	150
Hinzufügen zu einem Projekt	151
Übertragen von Projekten an das IBM SPSS Collaboration and Deployment Services Repository	152
Festlegen der Projekteigenschaften	152
Erstellen von Anmerkungen für ein Projekt	152
Objekteigenschaften	153
Schließen von Projekten	154
Generieren von Berichten	154
Speichern und Exportieren von generierten Berichten	156

Kapitel 12. Anpassen von IBM SPSS Modeler 157

Anpassen von IBM SPSS Modeler-Optionen	157
Festlegen von IBM SPSS Modeler-Optionen	157
Systemoptionen	157
Festlegen von Standardverzeichnissen	158
Festlegen von Benutzeroptionen	158
Anpassen der Knotenpalette	162
Anpassen des Palettenmanagers	162
Ändern der Ansicht einer Palettenregisterkarte	165
CEMI-Knotenverwaltung	165

Kapitel 13. Erläuterungen zur Leistung für Streams und Knoten 167

Knotenreihenfolge	167
Knotencaches	168
Leistung: Prozessknoten	169
Leistung: Modellierungsknoten	170
Leistung: CLEM-Ausdrücke	171

Kapitel 14. Eingabehilfen in IBM SPSS Modeler 173

Übersicht über die Eingabehilfen in IBM SPSS Modeler	173
Arten der Unterstützung von Eingabehilfen	173
Zugriffsmöglichkeiten für Sehbehinderte	173
Zugriffsmöglichkeiten für Blinde	174
Tastatureingabehilfen	174
Verwenden eines Sprachausgabeprogramms	182
Tipps zur Verwendung	183
Kompatibilität mit anderer Software	184
JAWS und Java	184
Verwenden von Diagrammen in IBM SPSS Modeler	184

Kapitel 15. Unicode-Unterstützung	185
Unicode-Unterstützung in IBM SPSS Modeler . . .	185
Bemerkungen	187

Marken	188
Index	189

Kapitel 1. Informationen zu IBM SPSS Modeler

IBM® SPSS Modeler ist ein Set von Data-Mining-Tools, mit dem Sie auf der Grundlage Ihres Fachwissens schnell und einfach Vorhersagemodelle erstellen und zur Erleichterung der Entscheidungsfindung in die Betriebsabläufe einbinden können. Das Produkt IBM SPSS Modeler, das auf der Grundlage des den Industrienormen entsprechenden Modells CRISP-DM entwickelt wurde, unterstützt den gesamten Data Mining-Prozess, von den Daten bis hin zu besseren Geschäftsergebnissen.

IBM SPSS Modeler bietet eine Vielzahl von Modellbildungsmethoden, die aus dem maschinellen Lernen, der künstlichen Intelligenz und der Statistik stammen. Mit den in der Modellierungspalette verfügbaren Methoden können Sie aus Ihren Daten neue Informationen ableiten und Vorhersagemodelle erstellen. Jede Methode hat ihre speziellen Stärken und eignet sich besonders für bestimmte Problemtypen.

SPSS Modeler kann als Standalone-Produkt oder als Client in Verbindung mit SPSS Modeler Server erworben werden. Außerdem ist eine Reihe von Zusatzoptionen verfügbar, die in den folgenden Abschnitten kurz zusammengefasst werden. For more information, see <http://www.ibm.com/software/analytics/spss/products/modeler/>.

IBM SPSS Modeler-Produkte

Zur IBM SPSS Modeler-Produktfamilie und der zugehörigen Software gehören folgende Elemente.

- IBM SPSS Modeler
- IBM SPSS Modeler Server
- IBM SPSS Modeler Administration Console
- IBM SPSS Modeler Batch
- IBM SPSS Modeler Solution Publisher
- IBM SPSS Modeler Server-Adapter für IBM SPSS Collaboration and Deployment Services

IBM SPSS Modeler

SPSS Modeler ist eine funktionell in sich abgeschlossene Produktversion, die Sie auf Ihrem PC installieren und ausführen können. Sie können SPSS Modeler im lokalen Modus als Standalone-Produkt oder im verteilten Modus zusammen mit IBM SPSS Modeler Server verwenden, um bei Datasets die Leistung zu verbessern.

Mit SPSS Modeler können Sie schnell und intuitiv genaue Vorhersagemodelle erstellen, und das ohne Programmierung. Mithilfe der speziellen visuellen Benutzerschnittstelle können Sie den Data Mining-Prozess auf einfache Weise visualisieren. Mit der Unterstützung der in das Produkt eingebetteten erweiterten Analyseprozesse können Sie zuvor verborgene Muster und Trends in Ihren Daten aufdecken. Sie können Ergebnisse modellieren und Einblick in die Faktoren gewinnen, die Einfluss auf diese Ergebnisse haben, wodurch Sie in die Lage versetzt werden, Geschäftschancen zu nutzen und Risiken zu mindern.

SPSS Modeler ist in zwei Editionen erhältlich: SPSS Modeler Professional und SPSS Modeler Premium. Weitere Informationen finden Sie unter im Thema „IBM SPSS Modeler-Editionen“ auf Seite 2.

IBM SPSS Modeler Server

SPSS Modeler verwendet eine Client/Server-Architektur zur Verteilung von Anforderungen für ressourcenintensive Vorgänge an leistungsstarke Serversoftware, wodurch bei größeren Datasets eine höhere Leistung erzielt werden kann.

SPSS Modeler Server ist ein separat lizenziertes Produkt, das durchgehend im verteilten Analysemodus auf einem Server-Host in Verbindung mit einer oder mehreren IBM SPSS Modeler-Installationen ausgeführt wird. Auf diese Weise bietet SPSS Modeler Server eine herausragende Leistung bei großen Datensets, da speicherintensive Vorgänge auf dem Server ausgeführt werden können, ohne Daten auf den Client-Computer herunterladen zu müssen. IBM SPSS Modeler Server bietet außerdem Unterstützung für SQL-Optimierung sowie Möglichkeiten zur Modellierung innerhalb der Datenbank, was weitere Vorteile hinsichtlich Leistung und Automatisierung mit sich bringt.

IBM SPSS Modeler Administration Console

Modeler Administration Console ist eine grafische Anwendung zur Verwaltung einer Vielzahl der SPSS Modeler Server-Konfigurationsoptionen, die auch mithilfe einer Optionsdatei konfiguriert werden können. Die Anwendung bietet eine Konsolenbenutzerschnittstelle zur Überwachung und Konfiguration der SPSS Modeler Server-Installationen und steht aktuellen SPSS Modeler Server-Kunden kostenlos zur Verfügung. Die Anwendung kann nur unter Windows installiert werden. Der von ihr verwaltete Server kann jedoch auf einer beliebigen unterstützten Plattform installiert sein.

IBM SPSS Modeler Batch

Das Data-Mining ist zwar in der Regel ein interaktiver Vorgang, es ist jedoch auch möglich, SPSS Modeler über eine Befehlszeile auszuführen, ohne dass die grafische Benutzerschnittstelle verwendet werden muss. Beispielsweise kann es sinnvoll sein, langwierige oder sich wiederholende Aufgaben ohne Eingreifen des Benutzers durchzuführen. SPSS Modeler Batch ist eine spezielle Version des Produkts, die die vollständigen Analysefunktionen von SPSS Modeler ohne Zugriff auf die reguläre Benutzerschnittstelle bietet. Zur Verwendung von SPSS Modeler Batch ist eine SPSS Modeler Server-Lizenz erforderlich.

IBM SPSS Modeler Solution Publisher

SPSS Modeler Solution Publisher ist ein Tool, mit dem Sie eine gepackte Version eines SPSS Modeler-Streams erstellen können, der durch eine externe Runtime-Engine ausgeführt oder in eine externe Anwendung eingebettet werden kann. Auf diese Weise können Sie vollständige SPSS Modeler-Streams für die Verwendung in Umgebungen veröffentlichen und bereitstellen, in denen SPSS Modeler nicht installiert ist. SPSS Modeler Solution Publisher wird als Teil des Diensts für IBM SPSS Collaboration and Deployment Services - Scoring verteilt, für den eine separate Lizenz erforderlich ist. Mit dieser Lizenz erhalten Sie SPSS Modeler Solution Publisher Runtime, womit Sie die veröffentlichten Streams ausführen können.

IBM SPSS Modeler Server-Adapter für IBM SPSS Collaboration and Deployment Services

Für IBM SPSS Collaboration and Deployment Services ist eine Reihe von Adaptern verfügbar, mit denen SPSS Modeler und SPSS Modeler Server mit einem IBM SPSS Collaboration and Deployment Services-Repository interagieren können. Auf diese Weise kann ein im Repository bereitgestellter SPSS Modeler-Stream von mehreren Benutzern gemeinsam verwendet werden. Auch der Zugriff über die Thin-Client-Anwendung IBM SPSS Modeler Advantage ist möglich. Sie installieren den Adapter auf dem System, das als Host für das Repository fungiert.

IBM SPSS Modeler-Editionen

SPSS Modeler ist in den folgenden Editionen erhältlich.

SPSS Modeler Professional

SPSS Modeler Professional bietet sämtliche Tools, die Sie für die Arbeit mit den meisten Typen von strukturierten Daten benötigen, beispielsweise in CRM-Systemen erfasste Verhaltensweisen und Interaktionen, demografische Daten, Kaufverhalten und Umsatzdaten.

SPSS Modeler Premium

SPSS Modeler Premium ist ein separat lizenziertes Produkt, das SPSS Modeler Professional für die Arbeit mit spezialisierten Daten, wie beispielsweise Daten, die für Entitätsanalysen oder soziale Netze verwendet werden, sowie für die Arbeit mit unstrukturierten Textdaten erweitert. SPSS Modeler Premium umfasst die folgenden Komponenten.

IBM SPSS Modeler Entity Analytics fügt den IBM SPSS Modeler-Vorhersageanalysen eine weitere Dimension hinzu. Während bei Vorhersageanalysen versucht wird, zukünftiges Verhalten aus früheren Daten vorherzusagen, liegt der Schwerpunkt bei der Entitätsanalyse auf der Verbesserung von Kohärenz und Konsistenz der aktuellen Daten, indem Identitätskonflikte innerhalb der Datensätze selbst aufgelöst werden. Bei der Identität kann es sich um die Identität einer Person, einer Organisation, eines Objekts oder einer anderen Entität handeln, bei der Unklarheiten bestehen könnten. Die Identitätsauflösung kann in einer Reihe von Bereichen entscheidend sein, darunter Customer Relationship Management, Betrugserkennung, Bekämpfung der Geldwäsche sowie nationale und internationale Sicherheit.

IBM SPSS Modeler Social Network Analysis transforms information about relationships into fields that characterize the social behavior of individuals and groups. Using data describing the relationships underlying social networks, IBM SPSS Modeler Social Network Analysis identifies social leaders who influence the behavior of others in the network. In addition, you can determine which people are most affected by other network participants. By combining these results with other measures, you can create comprehensive profiles of individuals on which to base your predictive models. Models that include this social information will perform better than models that do not.

IBM SPSS Modeler Text Analytics uses advanced linguistic technologies and Natural Language Processing (NLP) to rapidly process a large variety of unstructured text data, extract and organize the key concepts, and group these concepts into categories. Extracted concepts and categories can be combined with existing structured data, such as demographics, and applied to modeling using the full suite of IBM SPSS Modeler data mining tools to yield better and more focused decisions.

IBM SPSS Modeler-Dokumentation

Eine Dokumentation im Onlinehilfeformat finden Sie im Hilfemenü von SPSS Modeler. Diese umfasst die Dokumentation für SPSS Modeler, SPSS Modeler Server und SPSS Modeler Solution Publisher sowie das Anwendungshandbuch und weiteres Material zur Unterstützung.

Die vollständige Dokumentation für die einzelnen Produkte (einschließlich Installationsanweisungen) steht im PDF-Format im Ordner `\Documentation` auf der jeweiligen Produkt-DVD zur Verfügung. Installationsdokumente können auch aus dem Internet unter <http://www-01.ibm.com/support/docview.wss?uid=swg27038316> heruntergeladen werden.

Dokumentation in beiden Formaten steht auch im SPSS Modeler Information Center unter <http://publib.boulder.ibm.com/infocenter/spssmodl/v16r0m0/> zur Verfügung.

SPSS Modeler Professional-Dokumentation

Die SPSS Modeler Professional-Dokumentationssuite (ohne Installationsanweisungen) umfasst folgende Dokumente:

- **IBM SPSS Modeler Benutzerhandbuch.** Allgemeine Einführung in die Verwendung von SPSS Modeler, in der u. a. die Erstellung von Datenstreams, der Umgang mit fehlenden Werten, die Erstellung von CLEM-Ausdrücken, die Arbeit mit Projekten und Berichten sowie das Packen von Streams für die Bereitstellung in IBM SPSS Collaboration and Deployment Services, Predictive Applications (Vorhersageanwendungen) oder IBM SPSS Modeler Advantage beschrieben werden.
- **IBM SPSS Modeler Quellen-, Prozess- und Ausgabeknoten.** Beschreibung aller Knoten, die zum Lesen, zum Verarbeiten und zur Ausgabe von Daten in verschiedenen Formaten verwendet werden. Im Grunde sind sie alle Knoten, mit Ausnahme der Modellierungsknoten.

- **IBM SPSS Modeler Modellierungsknoten.** Beschreibungen sämtlicher für die Erstellung von Data-Mining-Modellen verwendeter Knoten. IBM SPSS Modeler bietet eine Vielzahl von Modellbildungsmethoden, die aus dem maschinellen Lernen, der künstlichen Intelligenz und der Statistik stammen.
- **IBM SPSS Modeler Algorithms Guide.** Beschreibung der mathematischen Grundlagen der in IBM SPSS Modeler verwendeten Modellierungsmethoden. Dieses Handbuch steht nur im PDF-Format zur Verfügung.
- **IBM SPSS Modeler Anwendungshandbuch.** Die Beispiele in diesem Handbuch bieten eine kurze, gezielte Einführung in bestimmte Modellierungsmethoden und -verfahren. Eine Online-Version dieses Handbuchs kann auch über das Hilfenü aufgerufen werden. Weitere Informationen finden Sie unter im Thema „Anwendungsbeispiele“ auf Seite 5.
- **IBM SPSS Modeler Handbuch für Scripterstellung und Automatisierung.** Informationen zur Automatisierung des Systems über Scripterstellung, einschließlich der Eigenschaften, die zur Bearbeitung von Knoten und Streams verwendet werden können.
- **IBM SPSS Modeler Bereitstellungshandbuch.** Informationen zum Ausführen von IBM SPSS Modeler-Streams und -Szenarios als Schritte bei der Verarbeitung von Jobs im IBM SPSS Collaboration and Deployment Services Deployment Manager.
- **IBM SPSS Modeler CLEF-Entwicklerhandbuch.** CLEF bietet die Möglichkeit, Drittanbieterprogramme, wie Datenverarbeitungsroutinen oder Modellierungsalgorithmen, als Knoten in IBM SPSS Modeler zu integrieren.
- **IBM SPSS Modeler Datenbankinternes Mining.** Informationen darüber, wie Sie Ihre Datenbank dazu einsetzen, die Leistung zu verbessern, und wie Sie die Palette der Analysefunktionen über Drittanbieteralgorithmen erweitern.
- **IBM SPSS Modeler Server Verwaltungs- und Leistungshandbuch.** Informationen zur Konfiguration und Verwaltung von IBM SPSS Modeler Server.
- **IBM SPSS Modeler Administration Console Benutzerhandbuch.** Informationen zur Installation und Nutzung der Konsolenbenutzerschnittstelle zur Überwachung und Konfiguration von IBM SPSS Modeler Server. Die Konsole ist als Plug-in für die Deployment Manager-Anwendung implementiert.
- **IBM SPSS Modeler CRISP-DM Handbuch.** Schritt-für-Schritt-Anleitung für das Data Mining mit SPSS Modeler unter Verwendung der CRISP-DM-Methode.
- **IBM SPSS Modeler Batch Benutzerhandbuch.** Vollständiges Handbuch für die Verwendung von IBM SPSS Modeler im Stapelmodus, einschließlich Details zur Ausführung des Stapelmodus und zu Befehlszeilenargumenten. Dieses Handbuch steht nur im PDF-Format zur Verfügung.

SPSS Modeler Premium Documentation

The SPSS Modeler Premium documentation suite (excluding installation instructions) is as follows.

- **IBM SPSS Modeler Entity Analytics User Guide.** Information on using entity analytics with SPSS Modeler, covering repository installation and configuration, entity analytics nodes, and administrative tasks.
- **IBM SPSS Modeler Social Network Analysis User Guide.** A guide to performing social network analysis with SPSS Modeler, including group analysis and diffusion analysis.
- **SPSS Modeler Text Analytics User's Guide.** Information on using text analytics with SPSS Modeler, covering the text mining nodes, interactive workbench, templates, and other resources.
- **IBM SPSS Modeler Text Analytics Administration Console User Guide.** Information on installing and using the console user interface for monitoring and configuring IBM SPSS Modeler Server for use with SPSS Modeler Text Analytics . The console is implemented as a plug-in to the Deployment Manager application.

Anwendungsbeispiele

Mit den Data-Mining-Tools in SPSS Modeler kann eine große Bandbreite an geschäfts- und unternehmensbezogenen Problemen gelöst werden; die Anwendungsbeispiele dagegen bieten jeweils eine kurze, gezielte Einführung in spezielle Modellierungsmethoden und -verfahren. Die hier verwendeten Datasets sind viel kleiner als die großen Datenbestände, die von einigen Data-Mining-Experten verwaltet werden müssen, die zugrunde liegenden Konzepte und Methoden sollten sich jedoch auch auf reale Anwendungen übertragen lassen.

Sie können auf die Beispiele zugreifen, indem Sie im Menü "Hilfe" in SPSS Modeler auf die Option **Anwendungsbeispiele** klicken. Die Datendateien und Beispielstreams wurden im Ordner *Demos*, einem Unterordner des Produktinstallationsverzeichnis, installiert. Weitere Informationen finden Sie unter im Thema „Ordner "Demos"“.

Beispiele für die Datenbankmodellierung. Die Beispiele finden Sie im IBM SPSS Modeler-Handbuch zum datenbankinternen Mining.

Scriptbeispiele. Die Beispiele finden Sie im *IBM SPSS Modeler Handbuch für Scripterstellung und Automatisierung*.

Ordner "Demos"

Die in den Anwendungsbeispielen verwendeten Datendateien und Beispielstreams wurden im Ordner *Demos*, einem Unterordner des Produktinstallationsverzeichnis, installiert. Auf diesen Ordner können Sie auch über die Programmgruppe IBM SPSS Modeler im Windows-Startmenü oder durch Klicken auf *Demos* in der Liste der zuletzt angezeigten Verzeichnisse im Dialogfeld "Datei öffnen" zugreifen.

Kapitel 2. Neue Funktionen in IBM SPSS Modeler 16

In diesem Release wurden IBM SPSS Modeler die folgenden Funktionen hinzugefügt.

Monte-Carlo-Simulation. Ein neuer Simulationsquellenknoten bietet eine einfache Methode, synthetische Daten unter Verwendung einer breiten Auswahl statistischer Verteilungen völlig neu zu generieren. Alternativ kann der neue Anpassungsknoten automatisch einen vorkonfigurierten Quellenknoten erstellen, der die Verteilungen von historischen Variablen und die Beziehungen zwischen diesen historischen Variablen wiedergibt. Der Simulationsevaluierungsknoten ist ein Endknoten zur Auswertung von Feldern, die aus einem simulierten Analysestream resultieren, und stellt hilfreiche Verteilungs- und Korrelationsdiagramme bereit.

Python-Scripting. Das Scripting in IBM SPSS Modeler, das zur Automatisierung von Prozessen in der Benutzerschnittstelle verwendet wird, kann nun die Python-Sprache verwenden, unterstützt aber auch weiterhin den traditionellen Scripting-Modus. Python ist eine bekannte und gängige Sprache, die eine Vielfalt von Funktionen bereitstellt, darunter eine umfangreiche und präzise Syntax, Fehlerbehandlung und leistungsfähige Zusatzmodule. **Hinweis:** Die Registerkarte **Script** unter **Tools > Streameigenschaften** heißt nun **Ausführung**.

Verwendung von Schleifen und Bedingte Ausführung. Diese neuen Optionen ermöglichen eine einfache Verwendung von Schleifen und eine bedingte Ausführung von Streams, ohne dass Scripts codiert werden müssen. Sie finden diese neuen Optionen unter **Tools > Streameigenschaften > Ausführung** oder indem Sie mit der rechten Maustaste auf einen Knoten innerhalb eines Streams klicken und die Ausführungsoption **Verwendung von Schleifen/Bedingte Ausführung** auswählen.

STB-Knoten. Erstellen Sie Klassen von Positions- und Zeitmarkendaten, um ausgereifere Analysen zu unterstützen. Im Aufenthaltsmodus gibt dieser Knoten auch die Zeiten und Orte an, an denen sich die Entitäten befinden. Zusätzliche Expression Builder-Funktionen unterstützen die Extraktion von STB-Zentroiden sowie Geohashing.

Verbesserungen in Entity Analytics. Neben der Auflösung einzelner Entitäten können nun n-Grad-Beziehungen zwischen Entitäten ermittelt werden. Des Weiteren werden die Echtzeitaktualisierung über den Streaming-Knoten, die Glättung aufgelöster Entitäten (Duplikatknoten) und die Anonymisierung von Daten unterstützt, die in ein Entitätsrepository eingefügt werden. **Hinweis:** Die vorherige lokale SolidDB-Datenbank wurde durch DB2 ersetzt.

Neuer ROC-Evaluierungsknotendiagrammtyp und AUC- und Gini-Metriken im Analyseknoten. Unterstützt binäre Ziele.

Neue Duplikatknotenoption zur Erstellung eines zusammengesetzten Datensatzes. Ermöglicht Ihnen die Angabe einer Aggregationsmethode für jedes Feld, das gruppiert wird, (erster Wert, letzter Wert, verkettete Werte usw.) anstelle des Löschens doppelter Datensätze. Während der Aggregatknoten in der Regel zum Zusammenfassen von Daten zu einer höheren Ebene verwendet wird, wird diese neue Option zur Glättung von Duplikaten verwendet; z. B. der Duplikate, die durch die Entitätsauflösung erkannt werden.

TM1-Quellen- und Exportknoten. Ermöglicht Ihnen den Zugriff auf TM1-Cube-Ansichten über den TM1-Quellenknoten und das Zurückscornen von Daten an einen vorhandenen TM1-Cube über den TM1-Exportknoten.

Aggregatausdrücke und Fensteraggregatfunktionen. Sie können benutzerdefinierte Aggregatausdrücke im Aggregatknoten erstellen und dabei integrierte Aggregatfunktionen (MEAN, SUM usw.) und/oder benutzerdefinierte Datenbankaggregatfunktionen übernehmen. In abgeleiteten Ausdrücken können Sie Fel-

der ableiten, für die Aggregatfunktionen mit Fensterfunktion erforderlich sind (wie z. B. gleitender Durchschnitt). Es sind integrierte und von Datenbanken bereitgestellte Fensteraggregatfunktionen verfügbar.

Verbesserungen des datenbankinternen Minings von IBM Netezza Analytics. Neuer Netezza-TwoStep-Algorithmus, zusätzliche Hilfsanwendungsoption zur Verwaltung (Löschen, Umbenennen usw.) von Netezza-Analysemodellen, Unterstützung für Modellviewer für Regressionsbäume, Entscheidungsbäume, K-Means und TwoStep. **Hinweis:** Für die neuen Funktionen ist INZA 3.0 erforderlich.

Neue R-Knoten und benutzerdefiniertes Dialogfelderstellungsprogramm für R. Neben dem R-Modellerstellungsknoten und dem Modellnugget, der bzw. das in SPSS Modeler 15 Fixpack 2 eingeführt wurde, werden in diesem Release zwei neue Knoten, der R-Prozessknoten und der R-Ausgabeknoten, hinzugefügt. Mit dem R-Prozessknoten können Sie Daten aus einem SPSS Modeler-Stream nehmen und Transformationen auf die Daten mithilfe des R-Scriptings anwenden. Mit dem R-Ausgabeknoten können Sie eigene benutzerdefinierte R-Scripts verwenden, um eine Datenanalyse durchzuführen und die Ergebnisse des Modellscorings zusammenzufassen.

Sie können eine Textausgabe oder eine grafische Ausgabe Ihrer Analyse erzeugen. Diese Ausgabe kann an eine Datei übertragen werden oder im Ausgabebrowser des R-Ausgabeknotens angezeigt werden. Das benutzerdefinierte Dialogerstellungsprogramm ermöglicht Ihnen die Erstellung von benutzerdefinierten Modellerstellungs-, Prozess- und Ausgabeknotentypen und Modellnuggets, einschließlich Feldauswahl, Text-, Zahlen- und Optionsfelder und Unterdialogsteuerelementen, um die Abstraktion und Parametrisierung von R-Programmen zu ermöglichen. Sie können den Knotentyp, die Zielpalette und das Knotensymbol auswählen, bevor Sie den Knoten installieren oder den Knoten für die Verwendung durch andere SPSS Modeler-Benutzer freigeben. Das benutzerdefinierte Dialogerstellungsprogramm wird aus dem Menü "Tools" gestartet. **Hinweis:** Für die Verwendung dieser Funktion müssen Sie SPSS Modeler - Essentials for R installiert haben.

R in Datenbank. Unterstützung von SQL-Pushback für R-Knoten; für Netezza, SAP Hana und Oracle durch Verwendung ihrer R-Unterstützung. **Hinweis:** Für Datenbanken müssen die entsprechenden vom Hersteller bereitgestellten R-Erweiterungen installiert sein.

Streaming-Zeitreihenprozessknoten. Sie können in einem Schritt Zeitreihenmodelle erstellen und für diese ein Scoring durchführen, um eine Bereitstellung in Echtzeit über IBM InfoSphere Streams, den Scoring-Service von IBM SPSS Collaboration and Deployment Services oder über IBM SPSS Modeler Solution Publisher zu ermöglichen.

Vorschau-schaltfläche. In Verbindung mit einer Datenbankquelle ermöglicht die Vorschau eine SQL-Pushback-Darstellung. In Verbindung mit einer Analytic Server-Datenquelle verwenden Sie die Vorschau-schaltfläche, um Informationen zu möglichen Übertragungen großer Datenmengen zu erhalten.

Neue Analytic Server-Optionen in den Knoten "Autom. Klassifikationsmerkmal", "Autonumerisch" und "Autom. Cluster". Wenn Sie einen Stream mit IBM SPSS Analytic Server ausführen, können Sie auswählen, ob Sie die Ausführung mit aktivierter Option "Aufteilungen" vornehmen (wenn Sie die Funktion für aufgeteilte Modelle verwenden, verwenden Sie diese Option) oder ob Sie die Ausführung mit den Optionen "Sehr große Datasets" vornehmen (Aufteilungen werden ignoriert und die Modellierungsziele werden für große Datenmengen und -vielfalt (Big Data) gesetzt). Scoring wird nun für die Modelle "Autom. Klassifikationsmerkmal", "Autonumerisch" und "Autom. Cluster" unterstützt, die in SPSS Modeler Server erstellt werden können.

Erweiterte Unterstützung von Scoring-Adaptern. Unterstützung für das Scoring von Textmining-Modellen durch Datenbankscoreing-Adapter. Darüber hinaus gibt es neue Datenbankscoreing-Adapter für DB2 LUW und Oracle.

Datenansichtsquellenknoten. Mithilfe des Datenansichtsknotens können Sie Daten aus einer Analysedatenansicht in einen SPSS Modeler-Stream einlesen. Die Analysedatenansicht ist eine neue Möglichkeit zur Erstellung einer einheitlichen Datenansicht in IBM SPSS Collaboration and Deployment Services 6.

Weitere Verbesserungen.

- Einteilung von Teradata-Abfragen in Datenbankverbindungsvoreinstellungen.
- Aktualisierte GLMM-Knotenoptionen.
- Unterstützung für verschlüsselte und komprimierte .sav-Dateien.
- Unterstützung für Single Sign-on (SSO) in SPSS Modeler Server, ohne dass IBM SPSS Collaboration and Deployment Services erforderlich ist.
- Datenbankschichtunterstützung für Single Sign-On.
- SQL-Pushback für den Beispielknoten für zDB2.
- R-Knoten werden nun im Rahmen der Basisinstallation von SPSS Modeler installiert.
- Modeler-Adapter für IBM SPSS Collaboration and Deployment Services werden über IBM Installation Manager installiert.
- Aktualisierung von aufgeteilten Modellen.
- Unterstützung für die Lokalisierung in Russisch.

Kapitel 3. IBM SPSS Modeler - Übersicht

Erste Schritte

Als Data-Mining-Anwendung bietet IBM SPSS Modeler eine strategische Methode zum Auffinden nützlicher Beziehungen in großen Datasets. Im Unterschied zu herkömmlicheren statistischen Methoden müssen Sie zu Beginn nicht unbedingt wissen, wonach Sie suchen. Sie können Ihre Daten durch Anpassen verschiedener Modelle und Überprüfen unterschiedlicher Beziehungen so lange untersuchen, bis Sie auf nützliche Informationen stoßen.

Starten von IBM SPSS Modeler

Klicken Sie zum Starten der Anwendung auf:

Start > [Alle] Programme > IBM SPSS Modeler 16 > IBM SPSS Modeler 16

Nach einigen Sekunden wird das Hauptfenster angezeigt.

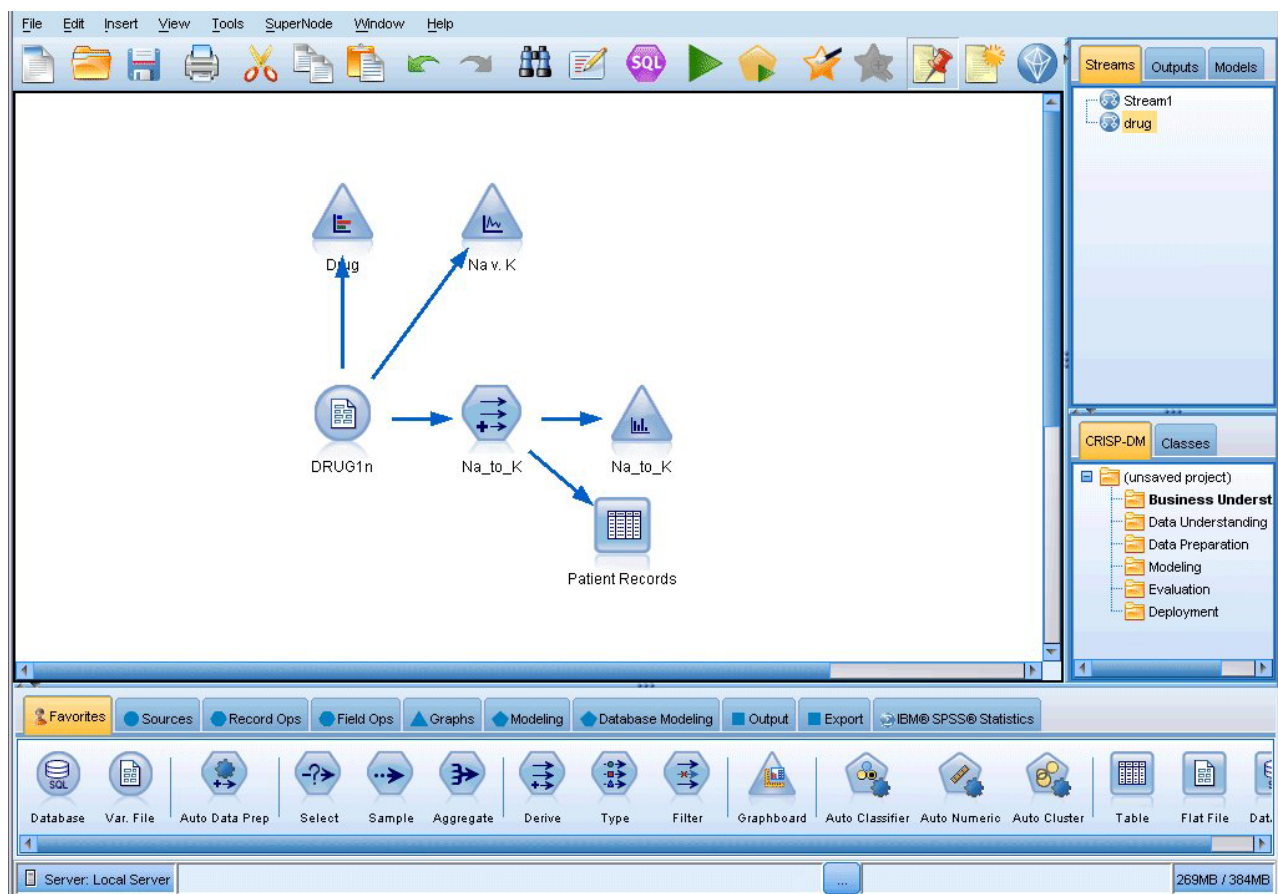


Abbildung 1. Hauptanwendungsfenster von IBM SPSS Modeler

Starten über die Befehlszeile

You can use the command line of your operating system to launch IBM SPSS Modeler as follows:

1. On a computer where IBM SPSS Modeler is installed, open a DOS, or command-prompt, window.

2. To launch the IBM SPSS Modeler interface in interactive mode, type the `modelerclient` command followed by the required arguments; for example:

```
modelerclient -stream report.str -execute
```

The available arguments (flags) allow you to connect to a server, load streams, run scripts, or specify other parameters as needed.

Verbinden mit IBM SPSS Modeler Server

IBM SPSS Modeler kann als eigenständige Anwendung oder als Client ausgeführt werden, der direkt mit IBM SPSS Modeler Server oder über das Plug-in Coordinator of Processes von IBM SPSS Collaboration and Deployment Services mit einer IBM SPSS Modeler Server-Instanz oder einem Server-Cluster verbunden ist. Der aktuelle Verbindungsstatus wird unten links im IBM SPSS Modeler-Fenster angezeigt.

Wenn Sie eine Verbindung zu einem Server herstellen möchten, können Sie den Namen des Servers, mit dem eine Verbindung hergestellt werden soll, manuell eingeben oder einen zuvor definierten Namen auswählen. Wenn Sie IBM SPSS Collaboration and Deployment Services verwenden, können Sie im Dialogfeld für die Anmeldung beim Server eine Liste von Servern bzw. Server-Clustern durchsuchen. Die Möglichkeit, die auf einem Netz ausgeführten Statistics-Dienste zu durchsuchen, wird über den Coordinator of Processes bereitgestellt.

So stellen Sie eine Verbindung mit einem Server her:

1. Klicken Sie im Menü "Extras" auf die Option **Anmelden beim Server**. Das Dialogfeld "Anmelden beim Server" wird geöffnet. Alternativ können Sie auf den Bereich des Verbindungsstatus im IBM SPSS Modeler-Fenster doppelklicken.
2. Legen Sie in diesem Dialogfeld die Optionen zum Verbinden mit dem lokalen Server-Computer fest oder wählen Sie eine Verbindung in der Tabelle aus.
 - Klicken Sie auf **Hinzufügen** bzw. **Bearbeiten**, um eine Verbindung hinzuzufügen bzw. zu bearbeiten. Weitere Informationen finden Sie unter im Thema „Hinzufügen und Bearbeiten der IBM SPSS Modeler Server-Verbindung“ auf Seite 13.
 - Klicken Sie auf **Suche**, um auf einen Server bzw. Server-Cluster in Coordinator of Processes zuzugreifen. Weitere Informationen finden Sie unter im Thema „Suchen nach Servern in IBM SPSS Collaboration and Deployment Services“ auf Seite 13.

Servertabelle. Diese Tabelle enthält die Menge der definierten Serververbindungen. In der Tabelle werden die Standardverbindung, der Servername sowie die Beschreibung und Portnummer angegeben. Sie können manuell eine neue Verbindung hinzufügen sowie eine bestehende Verbindung auswählen bzw. danach suchen. Um einen bestimmten Server als Standardverbindung einzurichten, aktivieren Sie in der Tabelle für die Verbindung das Kontrollkästchen in der Spalte "Standard".

Standarddatenpfad. Geben Sie einen Pfad an, der für Daten auf dem Server-Computer verwendet wird. Mit der Auslassungsschaltfläche (...) wechseln Sie zum gewünschten Verzeichnis.

Berechtigungs-nachweise einstellen. Lassen Sie dieses Kontrollkästchen unausgewählt, um die Funktion für **Single Sign-on** zu aktivieren. Diese versucht, Sie mithilfe Ihres lokalen Benutzernamens und Kennworts beim Server anzumelden. Falls ein Single Sign-on nicht möglich ist oder Sie das Kontrollkästchen zur Inaktivierung des Single Sign-on aktivieren (z. B. zur Anmeldung an einem Administratorkonto), wird ein weiteres Fenster angezeigt, in dem Sie aufgefordert werden, Ihre Berechtigungs-nachweise einzugeben.

Benutzer-ID. Geben Sie den Benutzernamen ein, mit dem die Anmeldung beim Server erfolgen soll.

Kennwort. Geben Sie das Kennwort ein, das zum angegebenen Benutzernamen gehört.

Domäne. Geben Sie die Domäne an, mit der die Anmeldung beim Server erfolgen soll. Ein Domänenname ist nur dann erforderlich, wenn sich der Server-Computer in einer anderen Windows-Domäne befindet als der Client-Computer.

3. Klicken Sie auf **OK**, um die Verbindung herzustellen.

So trennen Sie eine Verbindung mit einem Server:

1. Klicken Sie im Menü "Extras" auf die Option **Anmelden beim Server**. Das Dialogfeld "Anmelden beim Server" wird geöffnet. Alternativ können Sie auf den Bereich des Verbindungsstatus im IBM SPSS Modeler-Fenster doppelklicken.
2. Wählen Sie im Dialogfeld den lokalen Server aus und klicken Sie auf **OK**.

Hinzufügen und Bearbeiten der IBM SPSS Modeler Server-Verbindung

Serververbindungen können manuell im Kontrollkästchen "Anmelden beim Server" bearbeitet bzw. hinzugefügt werden. Durch Klicken auf "Hinzufügen" können Sie auf ein leeres Dialogfeld vom Typ "Server hinzufügen/bearbeiten" zugreifen, in dem Sie Details zur Serververbindung eingeben können. Durch Auswahl einer bestehenden Verbindung und Klicken auf "Bearbeiten" im Dialogfeld "Anmelden beim Server" wird das Dialogfeld "Server hinzufügen/bearbeiten" mit den Details für die betreffende Verbindung geöffnet, sodass Sie etwaige Änderungen vornehmen können.

Hinweis: Serververbindungen, die über IBM SPSS Collaboration and Deployment Services hinzugefügt wurden, können nicht bearbeitet werden, da Name, Port und sonstige Details in IBM SPSS Collaboration and Deployment Services definiert sind.

So fügen Sie Serververbindungen hinzu:

1. Klicken Sie im Menü "Extras" auf die Option **Anmelden beim Server**. Das Dialogfeld "Anmelden beim Server" wird geöffnet.
 2. Klicken Sie in diesem Dialogfeld auf **Hinzufügen**. Das Dialogfeld "Anmeldung beim Server: Server hinzufügen/bearbeiten" wird angezeigt.
 3. Geben Sie die Details für die Serververbindung ein und klicken Sie auf **OK**, um die Verbindung zu speichern und zum Dialogfeld "Anmeldung beim Server" zurückzukehren.
- **Server.** Geben Sie einen verfügbaren Server an oder wählen Sie einen aus der Liste aus. Der Server-Computer lässt sich anhand eines alphanumerischen Namens (z. B. *meinserver*) oder der dem Server-Computer zugewiesenen IP-Adresse (z. B. 202.123.456.78) identifizieren.
 - **Port.** Geben Sie die Portnummer an, die der Server überwacht. Wenn der Standardwert nicht funktioniert, fragen Sie Ihren Systemadministrator nach der richtigen Portnummer.
 - **Beschreibung.** Geben Sie eine optionale Beschreibung für diese Serververbindung ein.
 - **Verbindung verschlüsseln (mit SSL).** Legt fest, ob eine SSL-Verbindung (**Secure Sockets Layer**) verwendet werden soll. SSL ist ein weit verbreitetes Protokoll zum Schutz der über ein Netz versendeten Daten. Um diese Funktion verwenden zu können, muss SSL auf dem Server, auf dem sich IBM SPSS Modeler Server befindet, aktiviert sein. Wenden Sie sich gegebenenfalls an den lokalen Administrator, wenn Sie weitere Informationen benötigen.

So bearbeiten Sie Serververbindungen:

1. Klicken Sie im Menü "Extras" auf die Option **Anmelden beim Server**. Das Dialogfeld "Anmelden beim Server" wird geöffnet.
2. Wählen Sie in diesem Dialogfeld die zu bearbeitende Verbindung aus und klicken Sie dann auf **Bearbeiten**. Das Dialogfeld "Anmeldung beim Server: Server hinzufügen/bearbeiten" wird angezeigt.
3. Ändern Sie die Details für die Serververbindung und klicken Sie auf **OK**, um die Änderungen zu speichern und zum Dialogfeld "Anmeldung beim Server" zurückzukehren.

Suchen nach Servern in IBM SPSS Collaboration and Deployment Services

Anstatt eine Serververbindung manuell einzugeben, können Sie einen im Netz verfügbaren Server oder Server-Cluster über Coordinator of Processes auswählen. Diese Funktion ist in IBM SPSS Collaboration and Deployment Services verfügbar. Ein Server-Cluster ist eine Gruppe von Servern, aus denen Coordinator of Processes den Server ermittelt, der am besten für die Beantwortung einer Verarbeitungsanforderung geeignet ist.

Sie können zwar auch manuell Server im Dialogfeld "Anmelden beim Server" hinzufügen, durch die Suche nach verfügbaren Servern können Sie jedoch eine Verbindung zu Servern herstellen, ohne den richtigen Servernamen und die Portnummer kennen zu müssen. Diese Informationen werden automatisch bereitgestellt. Allerdings benötigen Sie auch bei dieser Variante die richtigen Anmeldeinformationen, wie Benutzername, Domäne und Kennwort.

Hinweis: Wenn Sie keinen Zugriff auf die Funktion Coordinator of Processes haben, können Sie dennoch den Namen des Servers, mit dem eine Verbindung hergestellt werden soll, manuell eingeben oder einen zuvor definierten Namen auswählen. Weitere Informationen finden Sie unter im Thema „Hinzufügen und Bearbeiten der IBM SPSS Modeler Server-Verbindung“ auf Seite 13.

So suchen Sie nach Servern und Clustern:

1. Klicken Sie im Menü "Extras" auf die Option **Anmelden beim Server**. Das Dialogfeld "Anmelden beim Server" wird geöffnet.
2. Klicken Sie in diesem Dialogfeld auf **Suche**, um das Dialogfeld "Nach Servern suchen" zu öffnen. Wenn Sie versuchen, Coordinator of Processes zu durchsuchen, ohne bei IBM SPSS Collaboration and Deployment Services angemeldet zu sein, werden Sie zur Anmeldung aufgefordert.
3. Wählen Sie den Server bzw. Server-Cluster in der Liste aus.
4. Klicken Sie auf **OK**, um das Dialogfeld zu schließen und diese Verbindung der Tabelle im Dialogfeld "Anmelden beim Server" hinzuzufügen.

Ändern des temporären Verzeichnisses

Für einige von IBM SPSS Modeler Server durchgeführte Operationen müssen möglicherweise temporäre Dateien erstellt werden. IBM SPSS Modeler verwendet standardmäßig das temporäre Systemverzeichnis zum Erstellen von temporären Dateien. Sie können das temporäre Verzeichnis wie folgt ändern:

1. Erstellen Sie ein neues Verzeichnis namens *spss* und ein Unterverzeichnis namens *servoertemp*.
2. Bearbeiten Sie die Datei *options.cfg*, die sich im Unterverzeichnis */config* Ihres IBM SPSS Modeler-Installationsverzeichnisses befindet. Bearbeiten Sie den Parameter *temp_directory* in diesem Feld, damit er wie folgt lautet: *temp_directory, "C:/spss/servoertemp"*.
3. Anschließend müssen Sie den IBM SPSS Modeler Server-Dienst neu starten. Dies ist möglich, wenn Sie auf die Registerkarte **Dienste** Ihrer Windows-Systemsteuerung klicken. Halten Sie den Dienst einfach an und starten Sie ihn erneut, um die von Ihnen durchgeführten Änderungen zu aktivieren. Durch das Neustarten des Systems wird auch der Dienst neu gestartet.

Alle temporären Dateien werden nun in dieses neue Verzeichnis geschrieben.

Hinweis: Der häufigste Fehler bei diesem Vorgang besteht in der Verwendung des falschen Schrägstrichs. Zu verwenden ist der normale Schrägstrich.

Starten mehrerer IBM SPSS Modeler-Sitzungen

Wenn Sie mehr als eine IBM SPSS Modeler-Sitzung gleichzeitig starten müssen, müssen Sie einige Änderungen an Ihren IBM SPSS Modeler- und Windows-Einstellungen vornehmen. Möglicherweise müssen Sie dies z. B. dann tun, wenn Sie über zwei separate Serverlizenzen verfügen und zwei Streams mit zwei verschiedenen Servern von demselben Client-Computer ausführen wollen.

So ermöglichen Sie mehrere IBM SPSS Modeler-Sitzungen:

1. Klicken Sie auf Folgendes:
Start > [Alle] Programme > IBM SPSS Modeler 16
2. Klicken Sie mit der rechten Maustaste auf die Verknüpfung IBM SPSS Modeler 16 (die mit dem Symbol) und wählen Sie **Eigenschaften** aus.
3. Fügen Sie in dem Textfeld **Ziel** an das Ende der Zeichenfolge *-noshare* hinzu.

4. Wählen Sie in Windows Explorer Folgendes aus:
Extras > Ordneroptionen...
5. Wählen Sie auf der Registerkarte "Dateitypen" die Option "IBM SPSS Modeler-Stream" aus und klicken Sie auf **Erweitert**.
6. Wählen Sie im Dialogfeld "Dateityp bearbeiten" die Option "Mit IBM SPSS Modeler öffnen" aus und klicken Sie auf **Bearbeiten**.
7. Fügen Sie im Textfeld **Anwendung für diesen Vorgang** vor dem Argument **-stream** die Angabe **-noshare** hinzu.

Schnittstelle von IBM SPSS Modeler auf einen Blick

An jedem Punkt des Data-Mining-Prozesses können Sie über die benutzerfreundliche Benutzerschnittstelle von IBM SPSS Modeler Ihr spezielles Fachwissen einbringen. Modellierungsalgorithmen, wie Vorhersage, Klassifizierung, Segmentierung und Assoziationserkennung, gewährleisten leistungsstarke und genaue Modelle. Die Modellergebnisse können problemlos angewendet und in Datenbanken, in IBM SPSS Statistics und in einer Vielzahl anderer Anwendungen eingelesen werden.

Die Datenarbeit mit IBM SPSS Modeler besteht aus drei Schritten.

- Zunächst lesen Sie Daten in IBM SPSS Modeler ein.
- Anschließend unterziehen Sie die Daten einer Reihe von Bearbeitungen.
- Schließlich senden Sie die Daten an ein Ziel.

Diese Reihenfolge wird als **Datenstream** bezeichnet, da die Daten Datensatz für Datensatz von der Quelle durch jeden Bearbeitungsschritt zum Ziel fließen, was entweder zu einer Datenausgabe vom Typ "Modell" oder "Typ" führt.

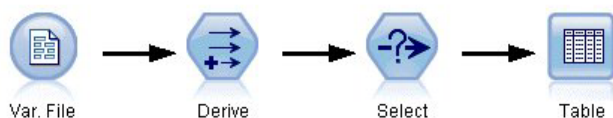


Abbildung 2. Ein einfacher Stream

Streamerstellungsbereich von IBM SPSS Modeler

Der Streamerstellungsbereich ist der größte Bereich des IBM SPSS Modeler-Fensters. Hier erstellen und bearbeiten Sie Datenstreams.

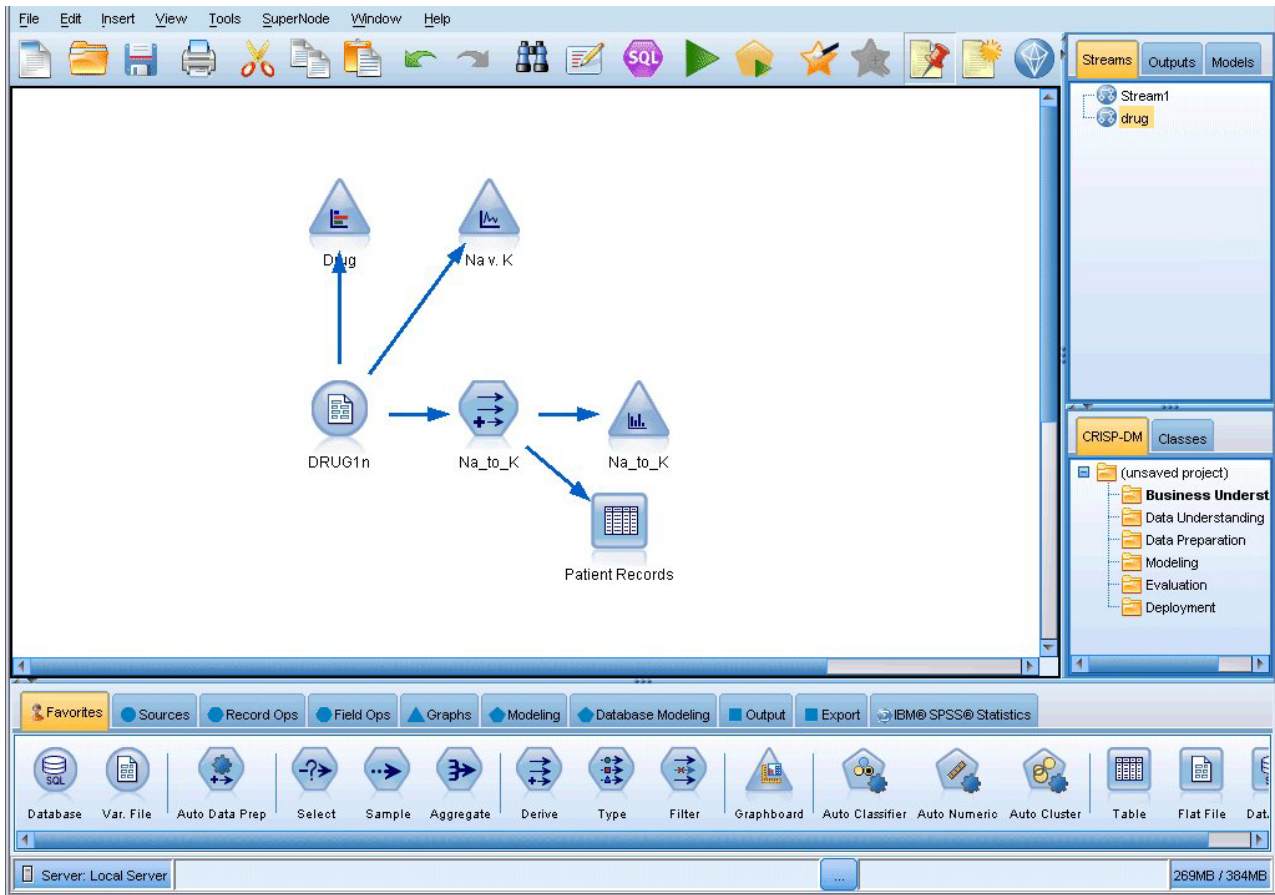


Abbildung 3. IBM SPSS Modeler-Arbeitsbereich (Standardansicht)

Die Erstellung von Streams erfolgt, indem im Haupterstellungsbereich der Benutzerschnittstelle Diagramme gezeichnet werden, in denen die für Ihren Geschäftsbetrieb relevanten Datenoperationen enthalten sind. Jede Operation wird durch ein Symbol oder einen **Knoten** dargestellt und die Knoten sind in einem **Stream** miteinander verbunden, der den Datenfluss durch jede Operation darstellt.

Sie können in IBM SPSS Modeler mit mehreren Streams gleichzeitig arbeiten, entweder in demselben Streamerstellungsbereich oder durch Öffnen eines neuen Streamerstellungsbereichs. Während einer Sitzung werden die Streams im Stream-Manager (rechts oben im IBM SPSS Modeler-Fenster) gespeichert.

Knotenpalette

Die meisten Daten und Modellierungstools in IBM SPSS Modeler befinden sich in der **Knotenpalette** (unten im Fenster) unterhalb des Streamerstellungsbereichs.

Die Palettenregisterkarte "Datensatzoperationen" beispielsweise enthält Knoten, mit denen Sie Operationen auf die **Datensätze** anwenden können, wie beispielsweise Auswählen, Zusammenführen (Mergen) und Anhängen.

Um Knoten zum Erstellungsbereich hinzuzufügen, doppelklicken Sie in den Knotenpaletten auf die entsprechenden Symbole oder ziehen Sie sie auf den Erstellungsbereich. Anschließend verbinden Sie sie, um einen **Stream** zu erstellen, der den Datenfluss darstellt.

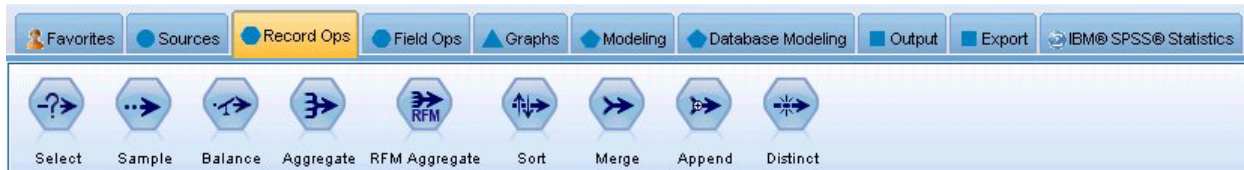


Abbildung 4. Registerkarte "Datensatzoperationen" in der Knotenpalette

Jede Palettenregisterkarte enthält eine Sammlung verwandter Knoten, die für verschiedene Phasen der Streamoperationen verwendet werden, wie:

- **Quellen.** Diese Knoten lesen Daten in IBM SPSS Modeler ein.
- **Datensatzoperationen.** Diese Knoten führen Operationen an **Datensätzen** durch, wie beispielsweise Auswählen, Zusammenführen (Mergen) und Anhängen.
- **Feldoperationen.** Diese Knoten führen Operationen an **Datenfeldern** durch, wie beispielsweise Filtern, Ableiten neuer Felder und Festlegen des Messniveaus für bestimmte Felder.
- **Diagramme.** Diese Knoten bieten eine grafische Darstellung der Daten vor und nach der Modellierung. Diagramme umfassen Plots, Histogramme, Netzdiagrammknoten und Evaluierungsdiagramme.
- **Modellierung.** Diese Knoten verwenden die in IBM SPSS Modeler verfügbaren Modellierungsalgorithmen, wie neuronale Netze, Entscheidungsbäume, Clusteralgorithmen und Datensequenzierung.
- **Datenbankmodellierung.** Diese Knoten verwenden die Modellierungsalgorithmen, die in Microsoft SQL Server-, IBM DB2-, Oracle- und Netezza-Datenbanken verfügbar sind.
- **Ausgabe.** Diese Knoten erzeugen verschiedenste Ausgabetypen für Daten, Diagramme und Modellergebnisse, die in IBM SPSS Modeler angezeigt werden können.
- **Export.** Diese Knoten erzeugen verschiedenste Ausgabetypen, die in externen Anwendungen angezeigt werden können, z. B. in IBM SPSS Data Collection oder Excel.
- **IBM SPSS Statistics.** Diese Knoten importieren Daten aus IBM SPSS Statistics oder exportieren sie nach IBM SPSS Statistics und führen IBM SPSS Statistics-Prozeduren aus.

Je vertrauter Sie im Umgang mit IBM SPSS Modeler werden, desto besser können Sie den Paletteninhalt für Ihre eigene Verwendung anpassen.

Der Berichtsbereich, der sich unter der Knotenpalette befindet, bietet Feedback zum Fortschritt verschiedener Operationen, z. B. wann die Daten in den Datenstream eingelesen werden. Außerdem bietet der Statusbereich, der sich ebenfalls unterhalb der Knotenpalette befindet, Informationen über die aktuelle Aktivität der Anwendung sowie Anweisungen, wann ein Benutzerfeedback erforderlich ist.

IBM SPSS Modeler-Manager

Oben rechts im Fenster befindet sich der Managerbereich. Dieser enthält drei Registerkarten, die zum Verwalten von Streams, Ausgaben und Modellen verwendet werden.

Sie können die Registerkarte "Streams" verwenden, um die in einer Sitzung erstellten Streams zu öffnen, umzubenennen, zu speichern und zu löschen.

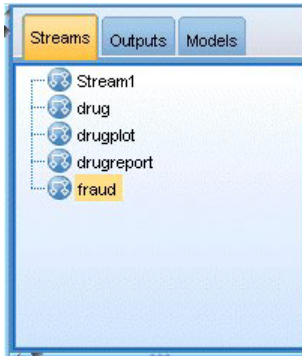


Abbildung 5. Registerkarte "Streams"



Abbildung 6. Registerkarte "Ausgaben"

Die Registerkarte "Ausgaben" enthält eine Reihe von Dateien, wie beispielsweise Diagramme und Tabellen, die von den Streamoperationen in IBM SPSS Modeler erstellt wurden. Sie können die auf dieser Registerkarte aufgeführten Tabellen, Diagramme und Berichte anzeigen, speichern, umbenennen und schließen.

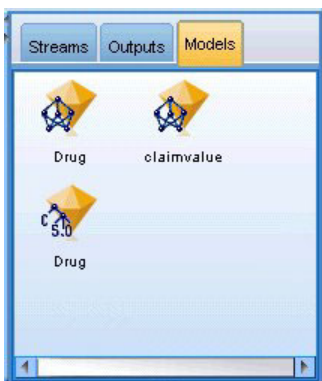


Abbildung 7. Registerkarte "Modelle" mit Modellnuggets

Die Registerkarte "Modelle" ist die leistungsstärkste der Manager-Registerkarten. Diese Registerkarte enthält sämtliche **Modellnuggets**, die die in IBM SPSS Modeler erstellten Modelle enthalten, für die aktuelle Sitzung. Auf der Registerkarte "Modelle" können diese Modelle direkt durchsucht oder dem Stream im Erstellungsbereich hinzugefügt werden.

IBM SPSS Modeler-Projekte

Unten rechts im Fenster befindet sich der Projektbereich, der zum Erstellen und Verwalten von Data-Mining-Projekten verwendet wird (Gruppen von Dateien, die in Bezug zu einer Data-Mining-Aufgabe stehen). Es gibt zwei Methoden zur Ansicht von in IBM SPSS Modeler erstellten Projekten - die Ansicht "Klassen" und die Ansicht "CRISP-DM".

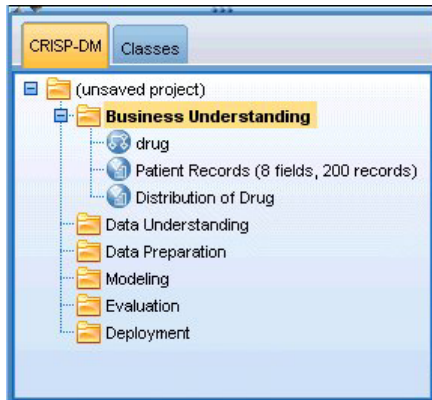


Abbildung 8. Ansicht "CRISP-DM"

Die Registerkarte "CRISP-DM" bietet ein Verfahren zur Organisation von Projekten gemäß dem Cross-Industry Standard Process for Data Mining, einer in der Branche bewährten, nicht eigentumsrechtlich geschützten Methode. Sowohl erfahrene Daten-Mining-Experten als auch Neulinge können mit dem CRISP-DM-Tool ihre Arbeit besser organisieren und an andere weitergeben.

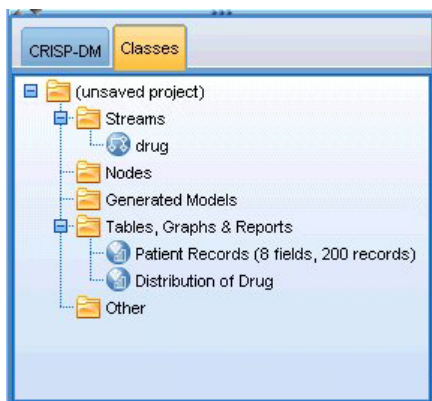


Abbildung 9. Ansicht "Klassen"

Die Registerkarte "Klassen" bietet eine Methode, mit der Sie Ihre Arbeit in IBM SPSS Modeler (nach den erstellten Objekttypen) in Kategorien organisieren können. Diese Ansicht ist hilfreich für das Inventarisieren von Daten, Streams und Modellen.

IBM SPSS Modeler-Symbolleiste






















Oben im IBM SPSS Modeler-Fenster sehen Sie eine Symbolleiste, die eine Vielzahl nützlicher Funktionen bietet. Im Folgenden werden die Schaltflächen der Symbolleiste und ihre Funktionen beschrieben.



Neuen Stream erstellen



Stream öffnen

	Stream speichern		Aktuellen Stream drucken
	Ausschneiden & in die Zwischenablage verschieben		In Zwischenablage kopieren
	Auswahl einfügen		Letzte Aktion rückgängig machen
	Wiederholen		Knoten suchen
	Streameigenschaften bearbeiten		Vorschau für SQL-Erzeugung
	Aktuellen Stream ausführen		Streamauswahl ausführen
	Stream anhalten (wird nur während der Streamausführung aktiv)		Superknoten hinzufügen
	Vergrößern (nur Superknoten)		Verkleinern (nur Superknoten)
	Kein Markup in Stream		Kommentar einfügen
	Stream-Markup ausblenden (falls vorhanden)		Ausgeblendeten Stream-Markup einblenden
	Stream in IBM SPSS Modeler Advantage öffnen		

Stream-Markup umfasst Streamkommentare, Modellverknüpfungen und die Anzeige von Scoring-Verzweigungen.

Eine Beschreibung zu Modellverknüpfungen finden Sie im Handbuch zu den *IBM SPSS-Modellierungsknoten*.

Anpassen der Symbolleiste

Sie können zahlreiche Aspekte der Symbolleiste ändern, z. B.:

- ob sie angezeigt wird
- ob für die Symbole eine QuickInfo verfügbar ist

- ob große oder kleine Symbole angezeigt werden

So schalten Sie die Anzeige der Symbolleiste ein bzw. aus:

1. Klicken Sie im Hauptmenü auf Folgendes:

Ansicht > Symbolleiste > Anzeigen

So ändern Sie die Einstellungen für QuickInfo oder Symbolgröße:

1. Klicken Sie im Hauptmenü auf Folgendes:

Ansicht > Symbolleiste > Anpassen

Klicken Sie wie gewünscht auf **QuickInfos einblenden** oder **Große Symbole**.

Anpassen des IBM SPSS Modeler-Fensters

Mit den Trennlinien zwischen den verschiedenen Bereichen der IBM SPSS Modeler-Benutzerschnittstelle können Sie je nach Bedarf die Größe von Tools ändern oder Tools schließen. Wenn Sie z. B. mit einem großen Stream arbeiten, können Sie die kleinen Pfeile, die sich auf den Teilern befinden, zum Schließen der Knotenpalette, des Managerbereichs und des Projektbereichs verwenden. So wird die Größe des Streamerstellungsbereichs vergrößert und bietet somit genügend Arbeitsfläche für große Streams oder mehrere Streams.

Alternativ können Sie auch im Menü "Ansicht" auf **Knotenpalette**, **Manager** oder **Projekt** klicken, um die Anzeige dieser Elemente ein- oder auszuschalten.

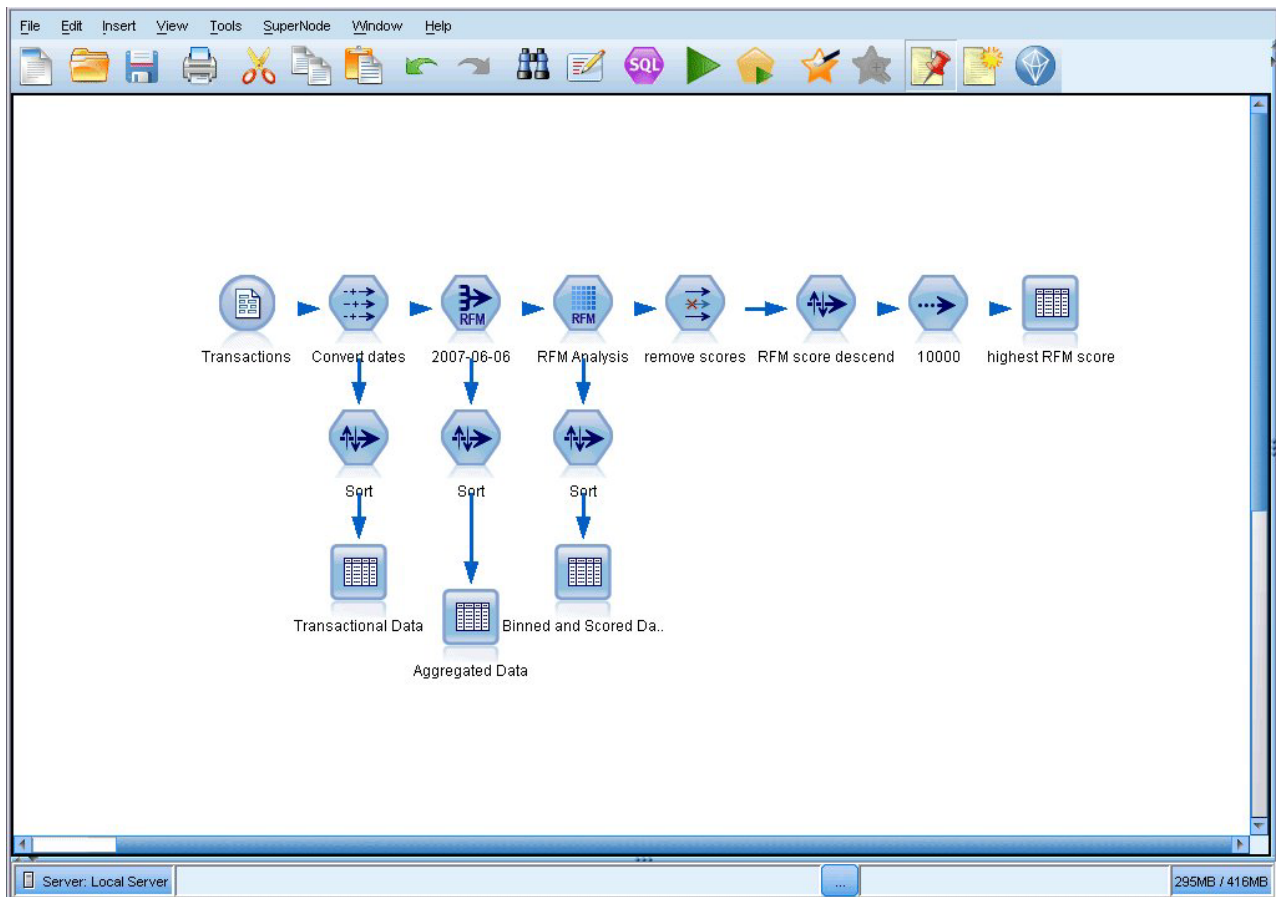


Abbildung 10. Maximierter Streamerstellungsbereich

Als Alternative zum Schließen der Knotenpalette und des Manager- und Projektbereichs können Sie den Streamerstellungsbereich als verschiebbare Seite verwenden, indem Sie die Bildlaufleisten an der Seite und unten im IBM SPSS Modeler-Fenster vertikal und horizontal bewegen.

Sie können auch die Anzeige der Anzeigemarkup steuern, die Streamkommentare, Modellverknüpfungen und die Anzeige von Scoring-Verzweigungen umfasst. Um diese Anzeige ein- oder auszublenden, klicken Sie auf:

Ansicht > Stream-Markup

Ändern der Symbolgröße für einen Stream

Sie können die Größe der Streamsymbole auf folgende Weisen ändern:

- Über eine Einstellung für Streameigenschaften
- Über ein Popup-Menü im Stream
- Über die Tastatur

Sie können die gesamte Streamansicht auf eine Anzahl von Größenwerten zwischen 8 und 200 % der Standardsymbolgröße skalieren.

Skalierung des gesamten Streams (Methode mit Streameigenschaften)

1. Wählen Sie im Hauptmenü Folgendes aus:
Tools > Streameigenschaften > Optionen > Layout.
2. Wählen Sie im Menü "Symbolgröße" die gewünschte Größe aus.
3. Klicken Sie auf **Anwenden**, um das Ergebnis anzuzeigen.
4. Klicken Sie auf **OK**, um die Änderung zu speichern.

Skalierung des gesamten Streams (Methode über Menü)

1. Klicken Sie mit der rechten Maustaste auf den Streamhintergrund im Erstellungsbereich.
2. Wählen Sie die Option **Symbolgröße** und anschließend die gewünschte Größe aus.

Skalierung des gesamten Streams (Methode über Tastatur)

1. Drücken Sie auf der Haupttastatur die Tastenkombination "Strg + [-]", um auf die nächstkleinere Größe umzuschalten.
2. Drücken Sie auf der Haupttastatur die Tastenkombination "Strg + Umschalt + [+]", um auf die nächstgrößere Größe umzuschalten.

Diese Funktion ist besonders nützlich, um einen allgemeinen Überblick über einen komplexen Stream zu erhalten. Sie können damit auch die für das Drucken eines Streams erforderliche Anzahl an Seiten minimieren.

Verwenden der Maus in IBM SPSS Modeler

Am häufigsten wird die Maus in IBM SPSS Modeler wie folgt verwendet:

- **Einfaches Klicken.** Verwenden Sie entweder die rechte oder die linke Maustaste, um Optionen aus den Menüs auszuwählen, Popup-Menüs zu öffnen und verschiedene andere Standardsteuerelemente und Optionen zu verwenden. Klicken Sie und halten Sie die Maustaste gedrückt, um Knoten zu verschieben und zu ziehen.
- **Doppelklicken.** Doppelklicken Sie mit der linken Maustaste, um Knoten auf dem Streamerstellungsbereich abzulegen und die bereits vorhandenen Knoten zu bearbeiten.
- **Klicken mit der mittleren Maustaste.** Klicken Sie auf die mittlere Maustaste und ziehen Sie den Cursor, um Knoten im Streamerstellungsbereich miteinander zu verbinden. Doppelklicken Sie auf die mitt-

lere Maustaste, um die Verbindung eines Knotens zu lösen. Wenn Sie nicht über eine Maus mit drei Tasten verfügen, können Sie diese Funktion simulieren, indem Sie die Taste Alt drücken und gleichzeitig auf die Maus klicken und ziehen.

Verwenden von Tastenkombinationen

Für viele visuelle Programmieroperationen in IBM SPSS Modeler gibt es Tastenkombinationen. Sie können z. B. einen Knoten löschen, indem Sie darauf klicken und die Löschtaste auf der Tastatur drücken. Auf ähnliche Weise können Sie einen Stream speichern, indem Sie auf die Taste "S" und gleichzeitig die Steuertaste (Taste Strg) drücken. Steuerbefehle wie dieser werden durch eine Kombination aus Strg und einer anderen Taste, z. B. Strg+S, angezeigt.

Es gibt eine Reihe von Tastenkombinationen, die für Windows-Standardoperationen verwendet werden, z. B. Strg+X zum Ausschneiden. Diese Tastenkombinationen werden in IBM SPSS Modeler zusammen mit den folgenden anwendungsspezifischen Tastenkombinationen unterstützt.

Hinweis: In manchen Fällen stehen alte, in IBM SPSS Modeler verwendete Zugriffstasten mit den Windows-Standardzugriffstasten in Konflikt. Diese alten Zugriffstasten werden durch Hinzufügen der Taste Alt unterstützt. So kann z. B. Strg+Alt+C zum Aktivieren und Inaktivieren des Cache verwendet werden.

Tabelle 1. Unterstützte Tastenkombinationen

Tastenkombination	Funktion
Strg+A	Alles markieren
Strg+X	Ausschneiden
Strg+N	Neuer Stream
Strg+O	Stream öffnen
Strg+P	Drucken
Strg+C	Kopieren
Strg+V	Einfügen
Strg+Z	Rückgängig
Strg+Q	Alle Knoten unterhalb des ausgewählten Knotens auswählen
Strg+W	Auswahl aller nachgeordneten Knoten aufheben (wechselt mit Strg+Q)
Strg+E	Vom ausgewählten Knoten ausführen
Strg+S	Aktuellen Stream speichern
Alt+Pfeiltasten	Ausgewählte Knoten im Streamerstellungsbereich in die Richtung verschieben, in die die verwendete Pfeiltaste zeigt
Umschalt+F10	Popup-Menü für den ausgewählten Knoten öffnen

Tabelle 2. Unterstützte Tastenkombinationen für alte Hot Keys

Tastenkombination	Funktion
Strg+Alt+D	Knoten duplizieren
Strg+Alt+L	Knoten laden
Strg+Alt+R	Knoten umbenennen
Strg+Alt+U	Benutzereingabeknoten erstellen
Strg+Alt+C	Cache aktivieren/inaktivieren
Strg+Alt+F	Cache löschen
Strg+Alt+X	Superknoten erweitern
Strg+Alt+Z	Vergrößern/Verkleinern

Table 2. Unterstützte Tastenkombinationen für alte Hot Keys (Forts.)

Tastenkombination	Funktion
Delete	Knoten oder Verbindung löschen

Drucken

Folgende Objekte können in IBM SPSS Modeler gedruckt werden:

- Streamdiagramme
- Grafiken
- Tabellen
- Berichte (über den Berichtsknoten und die Projektberichte)
- Scripts (über die Dialogfelder "Streameigenschaften", "Standalone-Script" oder "Superknotenscript")
- Modelle (Modellbrowser, Registerkarten des Dialogfelds mit aktuellem Fokus, Baumansichten)
- Anmerkungen (unter Verwendung der Registerkarte "Anmerkungen" für die Ausgabe)

So drucken Sie ein Objekt:

- Um ohne Vorschau zu drucken, klicken Sie auf die Schaltfläche "Drucken" in der Symbolleiste.
- Um die Seite vor dem Drucken einzurichten, wählen Sie im Menü "Datei" die Option **Seite einrichten** aus.
- Um vor dem Drucken eine Vorschau anzuzeigen, wählen Sie im Menü "Datei" die Option **Druckvorschau** aus.
- Um das Standarddialogfeld zum Drucken mit Optionen zur Auswahl von Druckern anzuzeigen und die Optionen für das Aussehen festzulegen, wählen Sie im Menü "Datei" die Option **Drucken** aus.

Automatisieren von IBM SPSS Modeler

Da es sich beim erweiterten Data-Mining um einen komplexen und manchmal langwierigen Vorgang handeln kann, bietet IBM SPSS Modeler mehrere Codierungsmöglichkeiten und Automatisierungsunterstützung.

- **Control Language for Expression Manipulation (CLEM)** ist eine Sprache zur Analyse und Bearbeitung der Daten, die IBM SPSS Modeler-Streams durchlaufen. Data-Mining-Experten verwenden CLEM in hohem Maße für Streamoperationen, um so einfache Aufgaben durchzuführen, wie das Ableiten von Profit aus Kosten- und Einkommensdaten, oder um so komplexe Aufgaben durchzuführen, wie das Umwandeln von Webprotokolldaten in ein Set von Feldern und Datensätzen mit nützlichen Informationen.
- Die **Scripterstellung** bildet ein leistungsstarkes Tool, mit dem Prozesse in der Benutzerschnittstelle automatisiert werden. Mit Scripts können dieselben Aktionen durchgeführt werden, die die Benutzer mithilfe von Maus und Tastatur durchführen. Sie können auch die Ausgabe angeben und generierte Modelle bearbeiten.

Kapitel 4. Data-Mining

Data-Mining - Übersicht

Mittels unterschiedlicher Verfahren werden beim **Data-Mining** "Informationsnuggets" im Datenhauptteil identifiziert. Beim Data-Mining werden Informationen so extrahiert, dass sie in Bereichen wie Entscheidungsunterstützung, Vorhersagen, Vorhersagen und Schätzungen verwendet werden können. Daten sind häufig sehr umfangreich, jedoch von geringem Wert und in ihrer Rohform von geringem praktischem Nutzen. Wertvoll sind die in den Daten verborgenen Informationen.

Beim Data-Mining beruht der Erfolg auf einer Kombination Ihres Wissens (oder das Ihres Experten) über die Daten und fortgeschrittener, aktiver Analyseverfahren, bei denen der Computer die zugrunde liegenden Beziehungen und Funktionen in den Daten ermittelt. Beim Data-Mining werden aus historischen Daten Modelle erstellt, die anschließend für Vorhersagen, Mustererkennung usw. verwendet werden. Das für die Erstellung dieser Modelle verwendete Verfahren wird **Maschinelernen** bzw. **Modellierung** genannt.

Modellierungsverfahren

IBM SPSS Modeler beinhaltet eine Reihe von Technologien für Maschinelernen und Modellieren, die grob anhand des Typs von Problemen unterteilt werden können, die sie lösen sollen.

- Methoden für Vorhersagemodelle enthalten Entscheidungsbäume, neuronale Netze und statistische Modelle.
- Clustermodelle konzentrieren sich auf die Ermittlung ähnlicher Datensätze und Beschriftung der Datensätze anhand der Gruppe, in die sie gehören. Zu den Clustermethoden gehören "Kohonen", "k-Means" und "TwoStep".
- Assoziationsregeln ordnen eine bestimmte Schlussfolgerung (beispielsweise den Kauf eines bestimmten Produkts) einer Menge von Bedingungen (dem Kauf mehrerer anderer Produkte) zu.
- Screening-Modelle können zum Auffinden von Feldern und Datensätzen verwendet werden, die höchstwahrscheinlich bei der Modellierung von Interesse sind oder zur Identifikation von Ausreißern, die keinem bekannten Muster entsprechen. Zu den verfügbaren Methoden gehören die Merkmalauswahl und die Anomalieerkennung.

Datenbearbeitung und -erkennung

IBM SPSS Modeler beinhaltet außerdem viele Funktionen, mit denen Sie Ihr Fachwissen auf die Daten anwenden können:

- **Datenbearbeitung.** Sie können neue Datenelemente erstellen, die aus bestehenden Daten abgeleitet sind, und die Daten in sinnvolle Subsets einteilen. Daten aus verschiedenen Quellen können zusammengeführt und gefiltert werden.
- **Durchsuchen und Visualisierung.** Sie können Aspekte der Daten mithilfe des Data Audit-Knotens anzeigen, um ein ursprüngliches Audit einschließlich Grafiken und Statistiken durchzuführen. Zur erweiterten Visualisierung gehören interaktive Grafiken, die zur Aufnahme in Projektberichte exportiert werden können.
- **Statistik.** Mit Statistiken können Sie vermutete Beziehungen zwischen den Variablen in den Daten überprüfen. Die Statistikfunktionen aus IBM SPSS Statistics können auch in IBM SPSS Modeler verwendet werden.
- **Hypothesentests.** Sie können Modelle für das Verhalten der Daten erstellen und überprüfen.

Normalerweise verwenden Sie diese Funktionen zur Ermittlung eines erfolgversprechenden Sets von Attributen in den Daten. Diese Attribute können anschließend in die Modellierungsverfahren eingespeist werden, die wiederum versuchen, etwaige zugrunde liegende Regeln und Beziehungen zu ermitteln.

Typische Anwendungen

Typische Anwendungsgebiete für Data-Mining-Verfahren sind beispielsweise:

Direktmarketing. Ermitteln Sie, welche demografischen Gruppen die höchste Rücklaufquote aufweisen. Mit diesen Informationen können Sie die Reaktionen auf zukünftige Anschreiben erhöhen.

Kreditscoring. Verwenden Sie die Kreditgeschichte einer Person als Grundlage für Kreditvergabeentscheidungen.

Humanressourcen. Gewinnen Sie einen Einblick in frühere Einstellungspraktiken und ermitteln Sie Entscheidungsrichtlinien zur Rationalisierung des Einstellungsprozesses.

Medizinische Forschung. Erstellen Sie auf medizinischen Erkenntnissen beruhende Entscheidungsregeln für die geeigneten Verfahren.

Marktanalyse. Ermitteln Sie, welche Variablen, wie Standort, Preis und Kundenmerkmale Einfluss auf den Absatz haben.

Qualitätskontrolle. Analysieren Sie Daten aus der Produktherstellung und ermitteln Sie Variablen, die entscheidenden Einfluss auf Produktmängel haben.

Grundsatzuntersuchungen. Verwenden Sie Umfragedaten zur Erarbeitung einer generellen Vorgehensweise, indem Sie anhand von Entscheidungsregeln die bedeutsamsten Variablen auswählen.

Gesundheitsfürsorge. Benutzerumfragen und klinische Daten können kombiniert werden, um Variablen zu ermitteln, die zur Gesundheit beitragen.

Terminologie

Die Ausdrücke **Attribut**, **Feld** und **Variable** beziehen sich auf ein einzelnes Datenelement, das alle betrachteten Fälle gemeinsam haben. Eine Sammlung von Attributwerten, die sich auf einen bestimmten Fall bezieht, heißt **Datensatz**, **Beispiel** oder **Fall**.

Bewerten der Daten

Data-Mining kann kaum erfolgreich sein, wenn die Daten, die Sie verwenden möchten, nicht bestimmten Kriterien entsprechen. In folgenden Abschnitten finden Sie einige Aspekte der Daten und ihrer Anwendung, die Sie beachten sollten.

Stellen Sie sicher, dass die Daten verfügbar sind.

Das mag trivial erscheinen, Sie sollten sich jedoch darüber im Klaren sein, dass möglicherweise zwar Daten vorliegen, jedoch eventuell nicht in einer problemlos verwendbaren Form. Daten können von IBM SPSS Modeler aus Datenbanken (über ODBC) oder aus Dateien importiert werden. Die Daten können jedoch in einem anderen Format auf einem System vorliegen, auf das kein direkter Zugriff möglich ist. Sie müssen heruntergeladen oder in einem geeigneten Format exportiert werden, bevor sie verwendet werden können. Möglicherweise sind sie auf verschiedene Datenbanken und Quellen verteilt und müssen zusammgebracht werden. Möglicherweise sind sie nicht einmal digital vorhanden. Wenn sie nur auf Papier vorliegen, ist eine Dateneingabe erforderlich, bevor Sie mit dem Data-Mining beginnen können.

Überprüfen Sie, ob die Daten die relevanten Attribute abdecken.

Ziel des Data-Mining ist die Ermittlung relevanter Attribute, sodass diese Überprüfung zunächst merkwürdig erscheinen mag. Es ist jedoch sinnvoll zu untersuchen, welche Daten verfügbar sind, und zu versuchen, die vermutlich relevanten Faktoren zu ermitteln, die nicht aufgezeichnet sind. Beim Versuch, den

Absatz an Speiseeis vorherzusagen, liegen Ihnen eventuell umfassende Informationen über Verkaufsstellen und frühere Absatzdaten vor, möglicherweise jedoch keine Informationen zu Wetter und Temperatur, die mit großer Wahrscheinlichkeit eine signifikante Rolle spielen. Fehlende Attribute bedeuten nicht unbedingt, dass das Data-Mining keine nützlichen Ergebnisse hervorbringt, die Genauigkeit der resultierenden Vorhersagen kann jedoch eingeschränkt sein.

Eine schnelle Möglichkeit zur Einschätzung der Situation besteht in der Durchführung eines umfassenden Audits der vorliegenden Daten. Bevor Sie fortfahren, sollten Sie der Datenquelle einen Data Audit-Knoten hinzufügen und ihn ausführen, um einen vollständigen Bericht zu generieren.

Vorsicht bei verrauschten Daten.

Daten enthalten häufig Fehler oder subjektive, und damit variable, Urteile. Diese Phänomene werden als **Rauschen** bezeichnet. Manchmal ist Rauschen in den Daten normal. Es können durchaus zugrunde liegende Regeln vorhanden sein, diese gelten jedoch eventuell nicht für 100 % der Fälle.

Üblicherweise gilt: Je mehr Rauschen in den Daten vorliegt, desto schwieriger ist es, genaue Ergebnisse zu erzielen. Die IBM SPSS Modeler-Methoden zum Maschinernen können jedoch mit verrauschten Daten umgehen und wurden erfolgreich bei Datensets eingesetzt, die annähernd 50 % Rauschen enthielten.

Stellen Sie sicher, dass ausreichend Daten vorliegen.

Beim Data-Mining ist nicht unbedingt die Größe des Datensets wichtig. Wesentlich bedeutsamer sind die *Repräsentativität* des Datensets sowie der Umfang, in dem es mögliche Ergebnisse und Variablenkombinationen abdeckt.

Üblicherweise gilt: Je mehr Attribute betrachtet werden, desto mehr Datensätze sind erforderlich, um eine repräsentative Abdeckung zu erzielen.

Wenn die Daten repräsentativ sind und allgemeine Regeln zugrunde liegen, ist es durchaus möglich, dass eine Datenstichprobe von wenigen tausend (oder sogar wenigen hundert) Datensätzen zu ebenso guten Ergebnissen führen wie eine Million Datensätze. Außerdem erhalten Sie die Ergebnisse schneller.

Kontaktieren Sie diejenigen, die sich mit den Daten auskennen.

In vielen Fällen arbeiten Sie mit Ihren eigenen Daten und ihr Inhalt und ihre Bedeutung sind Ihnen daher bestens bekannt. Wenn Sie jedoch mit Daten arbeiten, die aus einer anderen Abteilung in Ihrem Unternehmen oder von einem Kunden stammen, sollten Ihnen unbedingt Experten zur Verfügung stehen, die die Daten kennen. Sie können Sie in die Identifizierung relevanter Attribute einweisen und Ihnen helfen, die Ergebnisse des Data-Mining zu interpretieren und die echten "Informationsnuggets" von "Katzengold" bzw. Artefakten zu unterscheiden, die durch Anomalien in den Datensets verursacht wurden.

Eine Strategie für Data-Mining

Wie bei den meisten geschäftlichen Aktivitäten ist auch Data-Mining am effektivsten, wenn es auf geplante, systematische Weise durchgeführt wird. Selbst bei führenden Data-Mining-Tools wie IBM SPSS Modeler ist für den größten Teil der Arbeit beim Data-Mining ein sachkundiger Unternehmensanalyst erforderlich, damit der Prozess in zielgerichteten Bahnen abläuft. Bei Ihrer Planung sollten Sie sich folgende Fragen stellen:

- Welches inhaltliche Problem soll gelöst werden?
- Welche Datenquellen sind verfügbar und welche Teile der Daten sind für das aktuelle Problem relevant?
- Welche Art der Vorverarbeitung und Datenbereinigung ist erforderlich, bevor mit dem Data-Mining der Daten begonnen werden kann?
- Welche Data-Mining-Verfahren sollen eingesetzt werden?

- Wie sollen die Ergebnisse der Data-Mining-Analyse evaluiert werden?
- Wie können Sie die aus dem Data-Mining gewonnenen Informationen möglichst optimal nutzen?

Der übliche Data-Mining-Vorgang kann sehr schnell ziemlich kompliziert werden. Sehr viele Dinge müssen ständig bedacht werden: komplexe Geschäftsprobleme, mehrere Datenquellen, unterschiedliche Datenqualität in den einzelnen Datenquellen, diverse Data-Mining-Techniken, verschiedene Methoden zur Messung des Data-Mining-Erfolgs usw.

Um nicht den Überblick zu verlieren, sollten Sie über ein explizit definiertes Prozessmodell für Data-Mining verfügen. Das Prozessmodell hilft Ihnen dabei, die in diesem Abschnitt aufgeführten Fragen zu beantworten, und stellt sicher, dass die wichtigen Punkte beachtet werden. Es dient als grundlegender Data-Mining-Plan, der Ihnen hilft, in der Komplexität der Daten nicht das Ziel aus den Augen zu verlieren.

Für die Verwendung mit SPSS Modeler wird das Data-Mining-Prozessmodell CRISP-DM (Cross-Industry Standard Process for Data Mining) empfohlen. Wie aus dem Namen ersichtlich, wurde dieses Modell als allgemeines Modell entwickelt, das für eine Vielzahl von Branchen und Geschäftsproblemen angewendet werden kann.

Prozessmodell "CRISP-DM"

Das allgemeine Prozessmodell CRISP-DM umfasst sechs Phasen, die die Hauptprobleme beim Data-Mining behandeln. Die sechs Phasen verschmelzen zu einem zyklischen Prozess, mit dem die Einbettung von Data-Mining in Ihre umfassenderen Geschäftspraktiken möglich ist.

Bei den sechs Phasen handelt es sich um folgende:

- **Geschäftsverständnis.** Dies ist wahrscheinlich die wichtigste Phase des Data-Mining. Das Geschäftsverständnis umfasst die Ermittlung von Geschäftszielen, die Einschätzung der Situation, die Bestimmung der Data-Mining-Ziele und die Erstellung eines Projektplans.
- **Datenverständnis.** Daten sind das "Rohmaterial" beim Data-Mining. Diese Phase widmet sich der Notwendigkeit, einen Einblick in die Datenressourcen und die Eigenschaften dieser Ressourcen zu gewinnen. Dazu gehört das Sammeln der ursprünglichen Daten, das Beschreiben von Daten, das Untersuchen von Daten und das Überprüfen der Datenqualität. Der Data Audit-Knoten, der in der Palette "Ausgabeknoten" verfügbar ist, ist ein unverzichtbares Tool für das Verständnis der Daten.
- **Datenvorbereitung.** Nach der Katalogisierung der Datenressourcen müssen Sie die Daten für das Data-Mining vorbereiten. Zu den Vorbereitungen gehören die Auswahl, Bereinigung, Konstruktion, Integration und Formatierung der Daten.
- **Modellierung.** Dies ist natürlich der Hauptteil des Data-Mining, in dem mit hoch entwickelten Analysemethoden Informationen aus den Daten extrahiert werden. Zu dieser Phase gehört die Auswahl von Modellierungsverfahren, das Generieren von Testdesigns und der Aufbau und die Auswertung von Modellen.
- **Evaluation.** Nach der Auswahl der Modelle können Sie evaluieren, wie die Data-Mining-Ergebnisse zum Erreichen Ihrer Geschäftsziele beitragen können. Zu den Elementen in dieser Phase gehören die Evaluation von Ergebnissen, die Überprüfung des Data-Mining-Prozesses und die Bestimmung der nächsten Schritte.
- **Bereitstellung.** Dies ist die Phase, in der Sie erstmals praktischen Nutzen aus ihren Bemühungen ziehen können. Diese Phase konzentriert sich auf die Integration des neu erworbenen Wissens in die alltäglichen Geschäftsabläufe zur Lösung des ursprünglichen Geschäftsproblems. Zu dieser Phase gehören die Planbereitstellung, Überwachung und Anpassung, das Erstellen eines endgültigen Berichts und die Überprüfung des Projekts.

Einige Punkte in diesem Prozessmodell sind besonders wichtig. Zum einen besteht zwar die allgemeine Tendenz, dass der Prozess die Schritte in der in den vorherigen Abschnitten angegebenen Reihenfolge durchläuft, aber es gibt auch eine Reihe von Stellen, an denen die Phasen sich gegenseitig auf nicht lineare Weise beeinflussen. So geht die Datenvorbereitung in der Regel der Modellierung voraus. Die während

der Modellierungsphase getroffenen Entscheidungen und gesammelten Informationen können den Benutzer häufig veranlassen, Teile der Datenvorbereitungsphase zu überdenken, was wiederum neue Modellierungsfragen aufwerfen kann. Bei den beiden Phasen findet eine gegenseitige Rückkopplung statt, bis beide Phasen adäquat gelöst wurden. Ebenso kann die Evaluierungsphase zu einer Neubewertung des ursprünglichen Geschäftsverständnisses führen und Sie könnten zu dem Schluss kommen, dass Sie sich bisher mit der falschen Frage auseinandergesetzt haben. An diesem Punkt können Sie Ihr Geschäftsverständnis überarbeiten und den Rest des Prozesses mit einem besseren Ziel im Auge erneut durchführen.

Der zweite wichtige Punkt ist die iterative Natur des Data-Mining. Es wird selten der Fall sein, dass Sie einfach ein Data-Mining-Projekt planen, ausführen und dann fertig sind und die Daten verwenden. Data-Mining für die Erfüllung der Anforderungen Ihrer Kunden ist keine einmalige Angelegenheit, sondern eine Daueraufgabe. Das aus einem Data-Mining-Zyklus gewonnene Wissen führt fast immer zu neuen Fragen, neuen Problemen und neuen Möglichkeiten zur Ermittlung und Erfüllung der Bedürfnisse Ihrer Kunden. Diese neuen Fragen, Probleme und Möglichkeiten lassen sich normalerweise durch ein erneutes Data-Mining Ihrer Daten angehen. Dieser Prozess aus Data-Mining und Ermittlung neuer Möglichkeiten sollte Ihnen in Fleisch und Blut übergehen und zu einem Grundstein Ihrer geschäftlichen Gesamtstrategie werden.

Diese Einführung kann nur einen kurzen Überblick über das Prozessmodell CRISP-DM bieten. Vollständige Einzelheiten zum Modell finden Sie in den folgenden Ressourcen:

- Handbuch *CRISP-DM Guide*, das zusammen mit anderen Dokumenten im Verzeichnis \Documentation auf dem Installationsmedium aufgerufen werden kann.
- CRISP-DM-Hilfesystem, das über das Startmenü oder durch Klicken auf **CRISP-DM-Hilfe** im IBM SPSS Modeler-Hilfemenü aufgerufen werden kann.

Typen von Modellen

IBM SPSS Modeler bietet eine Vielzahl von Modellbildungsmethoden, die aus dem Maschinenlernen, der künstlichen Intelligenz und der Statistik stammen. Mit den in der Modellierungspalette verfügbaren Methoden können Sie aus Ihren Daten neue Informationen ableiten und Vorhersagemodelle erstellen. Jede Methode besitzt ihre Stärken und eignet sich besonders für bestimmte Problemtypen.

Im *IBM SPSS Modeler-Anwendungshandbuch* finden Sie Beispiele für viele dieser Methoden sowie eine allgemeine Einführung in den Modellierungsprozess. Dieses Handbuch ist als Online-Lernprogramm und im PDF-Format verfügbar. Weitere Informationen finden Sie im Thema „Anwendungsbeispiele“ auf Seite 5.

Modellierungsmethoden werden in drei Kategorien unterteilt:

- Klassifikation
- Assoziation
- Segmentierung

Klassifizierungsmodelle

Klassifikationsmodelle verwenden den Wert mindestens eines **Eingabefeldes**, um den Wert mindestens eines Ausgabe- oder **Zielfeldes** vorherzusagen. Einige Beispiele dieser Verfahren sind: Entscheidungsbäume (C&R-Baum-, QUEST-, CHAID- und C5.0-Algorithmen), Regression (lineare, logistische, verallgemeinert lineare und Cox-Regressionsalgorithmen), neuronale Netze, Support Vector Machines und Bayes-Netze.

Klassifizierungsmodelle können Unternehmen ein bekanntes Ergebnis vorhersagen. Beispielsweise, ob ein Kunde kaufen wird oder nicht oder ob eine Transaktion mit einem bekannten Betrugsmuster übereinstimmt. Zu den Modellierungstechniken gehören Maschinenlernen, Regelinduktion, Identifikation von Untergruppen, statistische Methoden und die Erzeugung mehrerer Modelle.

Klassifikationsknoten



Mit dem Knoten "Autom. Klassifikationsmerkmal" können Sie eine Reihe verschiedener Modelle für binäre Ergebnisse ("Ja" oder "Nein", "Abwanderung" oder "Keine Abwanderung" usw.) erstellen und vergleichen, um den besten Ansatz für die jeweilige Analyse auszuwählen. Es wird eine Reihe von Modellierungsalgorithmen unterstützt, sodass Sie die gewünschten Methoden, die spezifischen Optionen für die jeweilige Methode und die Kriterien zum Vergleich der Ergebnisse auswählen können. Der Knoten generiert eine Gruppe von Modellen, die auf den angegebenen Optionen beruhen, und erstellt anhand der von Ihnen angegebenen Kriterien eine Rangordnung der besten Kandidaten.



Der Knoten "Auto-Numerisch" schätzt und vergleicht mit einer Reihe verschiedener Methoden Modelle für die Ergebnisse stetiger numerischer Bereiche. Der Knoten arbeitet auf dieselbe Weise wie der Knoten "Automatisches Klassifikationsmerkmal": Sie können die zu verwendenden Algorithmen auswählen und in einem Modellierungsdurchlauf mit mehreren Optionskombinationen experimentieren. Folgende Algorithmen werden unterstützt: C&RT-Baum, CHAID, lineare Regression, verallgemeinerte lineare Regression und Support Vector Machines (SVM). Modelle können anhand von Korrelation, relativem Fehler bzw. Anzahl der verwendeten Variablen verglichen werden.



Der Knoten für Klassifizierungs- und Regressions-Bäume (C&RT-Bäume) erstellt einen Entscheidungsbaum, mit dem Sie zukünftige Beobachtungen vorhersagen oder klassifizieren können. Bei dieser Methode wird eine rekursive Partitionierung verwendet, um die Trainingsdatensätze in Segmente aufzuteilen. Dabei wird bei jedem Schritt die Unreinheit verringert und ein Knoten im Baum wird als "rein" betrachtet, wenn 100 % der Fälle in eine bestimmte Kategorie des Zielfelds fallen. Ziel- und Eingabefelder können numerische Bereiche oder kategorial (nominal, ordinal oder Flags) sein. Alle Aufteilungen sind binär (nur zwei Untergruppen).



Der QUEST-Knoten bietet eine binäre Klassifizierungsmethode zum Erstellen von Entscheidungsbäumen, die dazu dient, die für große C&R-Baumanalysen erforderliche Verarbeitungszeit zu verkürzen. Gleichzeitig soll die in den Klassifizierungsbaummodellen festgestellte Tendenz verringert werden, die darin besteht, dass Eingaben bevorzugt werden, die mehr Aufteilungen erlauben. Eingabefelder können stetig (numerische Bereiche) sein, das Zielfeld muss aber kategorial sein. Alle Aufteilungen sind binär.



Der CHAID-Knoten erzeugt Entscheidungsbäume unter Verwendung von Chi-Quadrat-Statistiken zur Ermittlung optimaler Aufteilungen. Im Gegensatz zu den Knoten vom Typ "C&RT-Baum" und "QUEST" kann CHAID nicht binäre Bäume generieren, d. h. Bäume mit Aufteilungen mit mehr als zwei Verzweigungen. Ziel- und Eingabefelder können in einem numerischen Bereich (stetig) oder kategorial sein. Exhaustive CHAID ist eine Änderung von CHAID, die noch gründlicher vorgeht, indem sie alle möglichen Aufteilungen untersucht, allerdings mehr Rechenzeit beansprucht.



Der C5.0-Knoten erstellt entweder einen Entscheidungsbaum oder ein Regelset. Das Modell teilt die Stichprobe auf der Basis des Felds auf, das auf der jeweiligen Ebene den maximalen Informationsgewinn liefert. Das Zielfeld muss kategorial sein. Es sind mehrere Aufteilungen in mehr als zwei Untergruppen zulässig.



Der Knoten "Entscheidungsliste" kennzeichnet Untergruppen bzw. Segmente, die eine höhere oder geringere Wahrscheinlichkeit für ein bestimmtes binäres Ergebnis aufweisen als die Gesamtpopulation. Sie könnten beispielsweise nach Kunden suchen, deren Abwanderung unwahrscheinlich ist oder die mit großer Wahrscheinlichkeit positiv auf eine Kampagne reagieren. Sie können Ihr Fachwissen in das Modell integrieren, indem Sie eigene, benutzerdefinierte Segmente hinzufügen und eine Vorschau anzeigen, in der alternative Modelle nebeneinander angezeigt werden, um die Ergebnisse zu vergleichen. Entscheidungslistenmodelle bestehen aus einer Liste von Regeln, bei denen jede Regel eine Bedingung und ein Ergebnis aufweist. Regeln werden in der vorgegebenen Reihenfolge angewendet und die erste Regel, die zutrifft, bestimmt das Ergebnis.



Bei linearen Regressionsmodellen wird ein stetiges Ziel auf der Basis linearer Beziehungen zwischen dem Ziel und einem oder mehreren Prädiktoren vorhergesagt.



Der Faktor/PCA-Knoten bietet leistungsstarke Datenreduktionsverfahren zur Verringerung der Komplexität der Daten. Die Hauptkomponentenanalyse (PCA) findet lineare Kombinationen der Eingabefelder, die die Varianz im gesamten Set der Felder am besten erfassen, wenn die Komponenten orthogonal (senkrecht) zueinander sind. Mit der Faktorenanalyse wird versucht, die zugrunde liegenden Faktoren zu bestimmen, die die Korrelationsmuster innerhalb eines Sets beobachteter Felder erklären. Bei beiden Ansätzen besteht das Ziel darin, eine kleinere Zahl abgeleiteter Felder zu finden, mit denen die Informationen in der ursprünglichen Menge der Felder effektiv zusammengefasst werden können.



Der Merkmalauswahlknoten sichtet die Eingabefelder, um auf der Grundlage einer Reihe von Kriterien (z. B. dem Prozentsatz der fehlenden Werte) zu entscheiden, ob diese entfernt werden sollen. Anschließend erstellt er eine Wichtigkeitsrangfolge der verbleibenden Eingaben in Bezug auf ein angegebenes Ziel. Beispiel: Angenommen, Sie haben ein Dataset mit Hunderten potenzieller Eingaben. Welche davon sind voraussichtlich für die Modellierung von medizinischen Behandlungsergebnissen von Bedeutung?



Bei der Diskriminanzanalyse werden strengere Annahmen als bei der logistischen Regression verwendet, sie kann jedoch eine wertvolle Alternative oder Ergänzung zu einer logistischen Regressionsanalyse sein, wenn diese Annahmen erfüllt sind.



Die logistische Regression ist ein statistisches Verfahren zur Klassifizierung von Datensätzen auf der Grundlage der Werte von Eingabefeldern. Sie ist analog zur linearen Regression, außer dass statt eines numerischen Bereichs ein kategoriales Zielfeld verwendet wird.



Das verallgemeinerte lineare Modell erweitert das allgemeine lineare Modell so, dass die abhängige Variable über eine angegebene Verknüpfungsfunktion in linearem Zusammenhang zu den Faktoren und Kovariaten steht. Außerdem ist es mit diesem Modell möglich, dass die abhängige Variable eine von der Normalverteilung abweichende Verteilung aufweist. Es deckt die Funktionen einer großen Bandbreite an Statistikmodellen ab, darunter lineare Regression, logistische Regression, loglineare Modelle für Häufigkeitsdaten und Überlebensmodelle mit Intervallzensurierung.



Verallgemeinerte lineare gemischte Modelle (Generalized Linear Mixed Models; GLMM) erweitern lineare Modelle so, dass das Ziel nicht normalverteilt zu sein braucht und über eine angegebene Verknüpfungsfunktion in einer linearen Beziehung zu den Faktoren und Kovariaten steht und die Beobachtungen korreliert werden können. Verallgemeinerte lineare gemischte Modelle decken eine breite Palette verschiedener Modelle ab, von einfacher linearer Regression bis hin zu komplexen Mehrebenenmodellen für nicht normalverteilte Longitudinaldaten.



Der Knoten vom Typ "Cox-Regression" ermöglicht Ihnen auch bei zensierten Datensätzen die Erstellung eines Überlebensmodells für Daten über die Zeit bis zum Eintreten des Ereignisses. Das Modell erstellt eine Überlebensfunktion, die die Wahrscheinlichkeit vorhersagt, dass das untersuchte Ereignis für bestimmte Werte der Eingabevariablen zu einem bestimmten Zeitpunkt (t) eingetreten ist.



Der Knoten "Support Vector Machine" (SVM) ermöglicht die Klassifizierung von Daten in eine von zwei Gruppen ohne Überanpassung. SVM eignet sich gut für umfangreiche Datensets, beispielsweise solche mit einer großen Anzahl an Eingabefeldern.



Mithilfe des Bayes-Netzwerknotens können Sie ein Wahrscheinlichkeitsmodell erstellen, indem Sie beobachtete und aufgezeichnete Hinweise mit Weltwissen kombinieren, um die Wahrscheinlichkeit ihres Vorkommens zu ermitteln. Der Knoten ist speziell für Netze vom Typ "Tree Augmented Naïve Bayes" (TAN) und "Markov-Decke" gedacht, die in erster Linie zur Klassifizierung verwendet werden.



Mithilfe des Knotens für das lernfähige Antwortmodell (Self-Learning Response Model, SLRM) können Sie ein Modell erstellen, in dem das Modell anhand eines einzelnen neuen Falls oder einer kleinen Anzahl neuer Fälle neu eingeschätzt werden kann, ohne dass das Modell mit allen Daten neu trainiert werden muss.



Der Zeitreihenknoten berechnet Schätzungen für die exponentielle Glättung sowie univariate und multivariate ARIMA-Modelle (ARIMA steht für Autoregressive Integrated Moving Average (autoregressiver integrierter gleitender Durchschnitt)) für Zeitreihendaten und erstellt Vorhersagen über die zukünftige Leistung. Einem Zeitreihenknoten muss stets ein Zeitintervallknoten vorangehen.



Der Knoten "k-Nächste Nachbarn" (KNN) verknüpft einen neuen Fall mit der Kategorie oder dem Wert der k Objekte, die ihm im Prädiktorraum am nächsten liegen, wobei k eine ganze Zahl ist. Ähnliche Fälle liegen nah beieinander und Fälle mit geringer Ähnlichkeit sind weit voneinander entfernt.

Assoziationsmodelle

Assoziationsmodelle finden Muster in Ihren Daten, bei denen mindestens eine Entität (wie Ereignisse, Einkäufe oder Attribute) mindestens einer anderen Entität zugeordnet sind. Die Modelle erstellen Regelsets, die diese Beziehungen definieren. Hier können die Felder innerhalb der Daten sowohl als Eingabe- als auch als Zielfelder fungieren. Sie könnten diese Assoziationen manuell finden, doch mithilfe von Assoziationsregelalgorithmen ist die Suche wesentlich schneller und es können komplexere Muster untersucht werden. Apriori- und Carma-Modelle sind Beispiele für die Verwendung solcher Algorithmen. Ein weiterer Typ eines Assoziationsmodells ist ein Sequenzerkennungsmodell, das sequenzielle Muster in zeitstrukturierten Daten findet.

Assoziationsmodelle sind sehr nützlich für die Vorhersage mehrerer Ergebnisse - wenn beispielsweise Kunden, die Produkt X gekauft haben, auch Y und Z gekauft haben. Assoziationsmodelle verknüpfen eine bestimmte Schlussfolgerung (wie die Entscheidung, etwas zu kaufen) mit einem Set von Bedingungen. Der Vorteil von Algorithmen für Assoziationsregeln gegenüber den eher standardmäßigen Entscheidungsbaumalgorithmen (C5.0 und C&RT) ist, dass zwischen beliebigen Attributen Assoziationen bestehen können. Ein Entscheidungsbaumalgorithmus erstellt Regeln mit nur einer Schlussfolgerung, während Assoziationsalgorithmen viele Regeln zu finden versuchen, von denen jede zu einer anderen Schlussfolgerung kommen kann.

Assoziationsknoten



Der Apriori-Knoten extrahiert ein Regelset aus den Daten und daraus die Regeln mit dem höchsten Informationsgehalt. Apriori bietet fünf verschiedene Methoden zur Auswahl von Regeln und verwendet ein ausgereiftes Indizierungsschema zur effizienten Verarbeitung großer Datensätze. Bei großen Problemen ist Apriori in der Regel schneller zu trainieren, es gibt keine willkürliche Begrenzung für die Anzahl der Regeln, die beibehalten werden können, und es können Regeln mit bis zu 32 Vorbedingungen verarbeitet werden. Bei Apriori müssen alle Ein- und Ausgabefelder kategorial sein; dafür bietet es jedoch eine bessere Leistung, da es für diesen Datentyp optimiert ist.



Beim CARMA-Modell wird ein Regelset aus den Daten extrahiert, ohne dass Sie Eingabe- oder Zielfelder angeben müssen. Im Gegensatz zu Apriori bietet der CARMA-Knoten Erstellungseinstellungen für die Regelunterstützung (Unterstützung für Antezedens und Sukzedens) und nicht nur für die Antezedens-Unterstützung. Die erstellten Regeln können somit für eine größere Palette an Anwendungen verwendet werden, beispielsweise um eine Liste mit Produkten und Dienstleistungen (Antezedenzien) zu finden, deren Nachfolger (Sukzedens) das Element darstellt, das Sie in der Ferienzeit desselben Jahres bewerben möchten.



Der Sequenzknoten erkennt Assoziationsregeln in sequenziellen oder zeitorientierten Daten. Eine Sequenz ist eine Liste mit Elementsets, die in einer vorhersagbaren Reihenfolge auftreten. Beispiel: Ein Kunde, der einen Rasierer und After-Shave-Lotion kauft, kauft möglicherweise beim nächsten Einkauf Rasiercreme. Der Sequenzknoten basiert auf dem CARMA-Assoziationsregelalgorithmus, der eine effiziente bidirektionale Methode zum Suchen von Sequenzen verwendet.

Segmentierungsmodelle

Segmentierungsmodelle teilen die Daten in Segmente, oder Cluster, von Datensätzen auf, die ähnliche Muster von Eingabefeldern aufweisen. Da sie nur an den Eingabefeldern interessiert sind, verfügen Segmentierungsmodelle nicht über die Konzepte der Ausgabe- oder Zielfelder. Beispiele für Segmentierungsmodelle sind Kohonen-Netze, K-Means-Clustering, TwoStep-Clustering und Anomalieerkennung.

Segmentierungsmodelle (auch Clustermodelle genannt) empfehlen sich in Fällen, in denen das genaue Ergebnis unbekannt ist (beispielsweise, wenn neue Betrugsmuster oder bestimmte Interessengruppen im Kundenstamm ermittelt werden sollen). Clustermodelle konzentrieren sich auf die Ermittlung ähnlicher Datensätze und Beschriftung der Datensätze anhand der Gruppe, in die sie gehören. Dies erfolgt ohne den Vorteil bereits zuvor vorhandener Kenntnisse der Gruppen und der zugehörigen Merkmale. Dies unterscheidet Clustermodelle von anderen Modellierungsverfahren: Es gibt kein zuvor definiertes Ausgabe- oder Zielfeld für das vorherzusagende Modell. Für diese Modelle gibt es keine richtigen oder falschen Antworten. Ihr Wert wird durch die Möglichkeit bestimmt, interessante Gruppierungen in den Daten zu erfassen und sinnvolle Beschreibungen dieser Gruppierungen zu liefern. Clustermodelle werden häufig verwendet, um Cluster oder Segmente zu erstellen, die dann als Eingaben in nachfolgenden Analysen verwendet werden (z. B. die Segmentierung potenzieller Kunden in homogene Untergruppen).

Segmentierungsknoten



Mit dem Knoten "Autom. Cluster" können Sie Clustering-Modelle, die Gruppen und Datensätze mit ähnlichen Merkmalen identifizieren, schätzen und vergleichen. Die Funktionsweise des Knotens gleicht der von anderen Knoten für automatisierte Modellierung, und Sie können in einem einzigen Modellierungsdurchgang mit mehreren Optionskombinationen experimentieren. Modelle können mithilfe grundlegender Messwerte für Filterung und Rangfolge der Nützlichkeit von Clustermodellen verglichen werden, um ein Maß auf der Basis der Wichtigkeit von bestimmten Feldern zu liefern.



Der K-Means-Knoten teilt das Dataset in unterschiedliche Gruppen (oder Cluster) auf. Bei diesem Verfahren wird eine festgelegte Anzahl von Clustern definiert, den Clustern werden iterativ Datensätze zugewiesen und die Clusterzentren werden angepasst, bis eine weitere Verfeinerung keine wesentliche Verbesserung des Modells mehr darstellen würde. Statt zu versuchen, ein Ergebnis vorherzusagen, versucht K-Means mithilfe eines als "nicht überwachttes Lernen" bezeichneten Verfahrens Muster im Set der Eingabefelder zu entdecken.



Der Kohonen-Knoten erstellt eine Art von neuronalem Netz, das verwendet werden kann, um ein Clustering des Datensets in einzelne Gruppen vorzunehmen. Wenn das Netz voll trainiert ist, sollten ähnliche Datensätze auf der Ausgabekarte eng nebeneinander stehen, während Datensätze, die sich unterscheiden, weit voneinander entfernt sein sollten. Die Zahl der von jeder Einheit im Modellnugget erfassten Beobachtungen gibt Aufschluss über die starken Einheiten. Dadurch wird ein Eindruck von der ungefähren Zahl der Cluster vermittelt.



Der TwoStep-Knoten verwendet eine aus zwei Schritten bestehende Clusterbildungsmethode. Im ersten Schritt wird ein einzelner Durchlauf durch die Daten vorgenommen, bei dem die Eingangsrohdaten zu einem verwaltbaren Set von Subclustern komprimiert werden. Im zweiten Schritt werden die Subcluster mithilfe einer hierarchischen Methode zur Clusterbildung nach und nach in immer größere Cluster zusammengeführt. TwoStep hat den Vorteil, dass die optimale Anzahl an Clustern für die Trainingsdaten automatisch geschätzt wird. Mit dem Verfahren können gemischte Feldtypen und große Datensets effizient verarbeitet werden.



Der Anomalieerkennungsknoten ermittelt ungewöhnliche Fälle bzw. "Ausreißer", die nicht den Mustern der "normalen" Daten entsprechen. Mit diesem Knoten können Ausreißer ermittelt werden, selbst wenn sie keinem bereits bekannten Muster entsprechen und selbst wenn Sie nicht genau wissen, wonach Sie suchen.

Modelle für datenbankinternes Mining

IBM SPSS Modeler unterstützt die Integration mit Data-Mining- und Datenmodellierungstools von Datenbank Anbietern wie Oracle Data Miner, IBM DB2 InfoSphere Warehouse und Microsoft Analysis Services. Sie können Modelle in der Datenbank erstellen, bewerten und speichern, ohne die IBM SPSS Modeler-Anwendung verlassen zu müssen. Vollständige Informationen finden Sie im IBM SPSS Modeler-Handbuch zum datenbankinternen Mining, das sich auf der Produkt-DVD befindet.

Modelle für IBM SPSS Statistics

Wenn auf Ihrem Computer eine Kopie von IBM SPSS Statistics installiert und lizenziert ist, können Sie auf bestimmte IBM SPSS Statistics-Routinen in IBM SPSS Modeler zugreifen und diese ausführen, um Modelle zu erstellen und zu scoren.

Weitere Informationen

Es steht auch eine detaillierte Dokumentation zu den Modellierungsalgorithmen zur Verfügung. Weitere Informationen finden Sie im *IBM SPSS Modeler Algorithms Guide* auf der Produkt-DVD.

Data-Mining-Beispiele

Um einen Einblick in die praktische Verwendung von Data Mining zu gewinnen, beginnen Sie am besten mit einem Beispiel. Eine Reihe von Anwendungsbeispielen steht im *IBM SPSS Modeler-Anwendungshandbuch* zur Verfügung. Dieses Handbuch bietet eine kurze, gezielte Einführung in bestimmte Modellierungsmethoden und -verfahren. Weitere Informationen finden Sie im Thema „Anwendungsbeispiele“ auf Seite 5.

Kapitel 5. Erstellen von Streams

Erstellen von Streams - Übersicht

Das Data-Mining mit IBM SPSS Modeler konzentriert sich auf die Ausführung von Daten über eine Reihe von Knoten, die als **Stream** bezeichnet wird. Diese Knotenreihe stellt Operationen dar, die für die Daten durchgeführt werden, während die Verknüpfungen zwischen den Knoten die Richtung des Datenflusses angeben. In der Regel verwenden Sie einen Datenstream, um Daten in IBM SPSS Modeler einzulesen, die Daten einer Reihe von Bearbeitungen zu unterziehen und die Daten anschließend an ein Ziel, z. B. eine Tabelle oder einen Viewer, zu senden.

Nehmen wir z. B. an, dass Sie eine Datenquelle öffnen, ein neues Feld hinzufügen, Datensätze auf der Basis von Werten in dem neuen Feld auswählen und dann die Ergebnisse in einer Tabelle anzeigen wollen. In diesem Fall würde Ihr Datenstream aus vier Knoten bestehen:



Ein Knoten für variable Dateien, den Sie einrichten, um die Daten aus der Datenquelle zu lesen.



Ein Ableitungsknoten, über den Sie das neue berechnete Feld dem Dataset hinzufügen.



Ein Auswahlknoten, über den Sie Auswahlkriterien einrichten, um Datensätze aus dem Datenstream auszuschließen.



Ein Tabellenknoten, über den Sie die Ergebnisse Ihrer Bearbeitungen auf dem Bildschirm anzeigen.

Erstellen von Datenstreams

Mit der einzigartigen Benutzerschnittstelle von IBM SPSS Modeler können Sie einen visuellen Mining-Vorgang für Ihre Daten ausführen, indem Sie mit Diagrammen von Datenstreams arbeiten. Am einfachsten können Sie einen Datenstream wie folgt erstellen:

- Fügen Sie dem Streamerstellungsbereich Knoten hinzu.
- Verbinden Sie die Knoten, um einen Stream zu bilden.
- Legen Sie die Knoten- bzw. Streamoptionen fest.
- Führen Sie den Stream aus.

Dieser Abschnitt enthält detaillierte Informationen zum Arbeiten mit Knoten, um komplexere Datenstreams zu erstellen. Es werden außerdem Optionen und Einstellungen für Knoten und Streams beschrieben. Beispiele mit Einzelschrittanleitungen zum Erstellen von Streams anhand der Daten aus dem Lieferumfang von IBM SPSS Modeler (im Ordner *Demos* Ihrer Programminstallation) finden Sie hier: „Anwendungsbeispiele“ auf Seite 5.

Arbeiten mit Knoten

Knoten werden in IBM SPSS Modeler zum Untersuchen von Daten verwendet. Verschiedene Knoten im Arbeitsbereich stellen verschiedene Objekte und Aktionen dar. Die Palette unten im IBM SPSS Modeler-Fenster enthält alle Knoten, mit denen Streams erstellt werden können.

Es gibt mehrere Knotentypen. **Quellenknoten** bringen Daten in den Stream und befinden sich auf der Registerkarte "Quellen" der Knotenpalette. **Prozessknoten** führen Operationen an einzelnen Datensätzen und Feldern aus und sind auf den Registerkarten "Datensatzoperationen" und "Feldoperationen" der Palette zu finden. **Ausgabeknoten** erzeugen eine Vielzahl von Ausgaben für Daten-, Diagramm- und Modellergebnisse und befinden sich auf den Registerkarten "Diagramme", "Ausgabe" und "Export" der Knotenpalette. **Modellierungsknoten** verwenden statistische Algorithmen, um Modellnuggets zu erstellen, und befinden sich auf der Registerkarte "Modellierung" sowie der Registerkarte "Datenbankmodellierung" (falls aktiviert) der Knotenpalette. Weitere Informationen finden Sie im Thema „Knotenpalette“ auf Seite 16.

Sie verbinden die Knoten, um Streams zu bilden, mit denen Sie bei Ausführung die Beziehungen visualisieren und Schlussfolgerungen ziehen können. Streams sind wie Scripts - Sie können sie speichern und mit anderen Datendateien wieder verwenden.

Ein ausführbarer Knoten, der Streamdaten verarbeitet, wird auch **Endknoten** genannt. Ein Modellierungs- oder Ausgabeknoten ist ein Endknoten, wenn er sich am Ende eines Streams oder Streamzweigs befindet. Sie können an einen Endknoten keine weiteren Knoten anschließen.

Hinweis: Sie können die Knotenpalette anpassen. Weitere Informationen finden Sie im Thema „Anpassen der Knotenpalette“ auf Seite 162.

Hinzufügen von Knoten zu einem Stream

Es gibt mehrere Möglichkeiten zum Hinzufügen von Knoten zu einem Stream über die Knotenpalette:

- Doppelklicken Sie auf einen Knoten in der Palette. *Hinweis:* Durch das Doppelklicken auf einen Knoten wird dieser automatisch mit dem aktuellen Stream verbunden. Weitere Informationen finden Sie im Thema „Verbinden von Knoten in einem Stream“.
- Ziehen Sie einen Knoten der Palette in den Streamerstellungsbereich und legen Sie ihn dort ab.
- Klicken Sie auf einen Knoten in der Palette und dann in den Streamerstellungsbereich.
- Wählen Sie eine geeignete Option aus dem Menü "Einfügen" von IBM SPSS Modeler aus.

Sobald Sie einen Knoten zum Streamerstellungsbereich hinzugefügt haben, doppelklicken Sie auf den Knoten, um das zugehörige Dialogfeld anzuzeigen. Die verfügbaren Optionen richten sich nach dem Knotentyp, den Sie hinzufügen. Informationen zu bestimmten Steuerelementen im Dialogfeld erhalten Sie, indem Sie auf die zugehörige Schaltfläche **Hilfe** klicken.

Entfernen von Knoten

Um einen Knoten aus dem Datenstream zu entfernen, klicken Sie darauf und drücken Sie die Löschtaste oder klicken Sie mit der rechten Maustaste und wählen Sie im Menü die Option **Löschen**.

Verbinden von Knoten in einem Stream

Knoten, die dem Streamerstellungsbereich hinzugefügt wurden, bilden so lange keinen Datenstream, bis sie verbunden sind. Die Verbindungen zwischen den Knoten zeigen an, in welche Richtung die Daten von einer Operation zur nächsten fließen. Es gibt mehrere Möglichkeiten, Knoten so zu verbinden, dass sie einen Stream bilden: durch Doppelklicken, mithilfe der mittleren Maustaste oder manuell.

Hinzufügen von Knoten durch Doppelklicken und Verbinden der Knoten:

Die einfachste Methode, einen Stream zu bilden, besteht darin, auf die Knoten in der Palette doppelzuklicken. Mit dieser Methode wird der neue Knoten automatisch mit dem im Streamerstellungsbereich ausgewählten Knoten verbunden. Wenn der Erstellungsbereich z. B. einen Datenbankknoten enthält, können Sie diesen Knoten auswählen und dann auf den nächsten Knoten in der Palette doppelklicken, z. B. auf einen Ableitungsknoten. Mit dieser Aktion wird der Ableitungsknoten automatisch mit dem bereits vorhandenen Datenbankknoten verbunden. Sie können diesen Prozess so lange wiederholen, bis Sie einen

Endknoten, z. B. einen Histogramm- oder Tabellenknoten, erreicht haben. Zu diesem Zeitpunkt werden sämtliche neue Knoten mit dem letzten Knoten, der kein Endknoten ist, aufwärts verbunden.

Verbinden von Knoten mithilfe der mittleren Maustaste:

Im Streamerstellungsbereich können Sie mithilfe der mittleren Maustaste durch Klicken und Ziehen einen Knoten mit einem anderen verbinden. (Wenn Ihre Maus keine mittlere Taste aufweist, können Sie dies simulieren, indem Sie die Taste Alt drücken, während Sie die Maus von einem Knoten zu einem anderen ziehen.)

Manuelles Verbinden von Knoten:

Falls Ihre Maus keine mittlere Maustaste aufweist und Sie es bevorzugen, die Knoten manuell zu verbinden, können Sie das Popup-Menü eines Knotens verwenden, um diesen mit einem bereits im Erstellungsbereich befindlichen anderen Knoten zu verbinden.

1. Klicken Sie mit der rechten Maustaste auf den Knoten, der die Verbindung starten soll. Dadurch wird das Knotenmenü geöffnet.
2. Klicken Sie im Menü auf **Verbinden**.
3. Sowohl am Ausgangsknoten als auch am Cursor wird ein Verbindungssymbol angezeigt. Klicken Sie auf einen zweiten Knoten im Erstellungsbereich, um die beiden Knoten zu verbinden.

Beim Verbinden von Knoten müssen verschiedene Richtlinien befolgt werden. Sie erhalten eine Fehlermeldung, wenn Sie versuchen, einen der folgenden Verbindungstypen zu erstellen:

- Eine Verbindung, die zu einem Quellenknoten führt
- Eine Verbindung, die von einem Endknoten kommt
- Einen Knoten mit mehr als der maximalen Anzahl Eingabeverbindungen
- Verbindung von zwei Knoten, die bereits miteinander verbunden sind
- Umlauf (Daten kehren zu einem Knoten zurück, von dem sie bereits geflossen sind)

Umgehen von Knoten in einem Stream

Wenn Sie einen Knoten im Datenstream umgehen, werden sämtliche Ein- und Ausgabeverbindungen, die direkt von den Eingabeknoten zu den Ausgabeknoten führen, ersetzt. Wenn der Knoten nicht über Eingabe- und Ausgabeverbindungen verfügt, werden sämtliche Verbindungen gelöscht anstatt umgeleitet.

So verfügen Sie u. U. über einen Stream, der ein neues Feld ableitet, Felder filtert und dann die Ergebnisse in einem Histogramm oder in einer Tabelle untersucht. Wenn Sie dasselbe Diagramm und die Tabelle für Daten *vor* dem Filtern von Feldern anzeigen möchten, können Sie dem Stream entweder neue Histogramm- und Tabellenknoten hinzufügen oder den Filterknoten umgehen. Wenn Sie den Filterknoten umgehen, werden die Verbindungen mit dem Diagramm und der Tabelle direkt vom Ableitungsknoten übergeben. Die Verbindung des Filterknotens mit dem Stream wird aufgehoben.

So umgehen Sie einen Knoten:

1. Verwenden Sie im Streamerstellungsbereich die mittlere Maustaste, um auf den Knoten doppelzuklicken, den Sie umgehen möchten. Alternativ können Sie auch die Kombination Alt+Doppelklick verwenden.

Hinweis: Sie können diese Aktion unter Verwendung der Option **Rückgängig** im Menü "Bearbeiten" oder mit der Tastenkombination Strg+Z rückgängig machen.

Inaktivieren von Knoten in einem Stream

Prozessknoten mit einer einzelnen Eingabe innerhalb von Streams können inaktiviert werden, mit dem Ergebnis, dass der Knoten bei der Ausführung des Streams ignoriert wird. Auf diese Weise müssen Sie den Knoten nicht löschen oder umgehen und können seine Verbindung mit den restlichen Knoten beste-

hen lassen. Die Knoteneinstellungen können nach wie vor angezeigt und bearbeitet werden. Änderungen werden jedoch erst dann wirksam, wenn Sie den Knoten wieder aktivieren.

Angenommen Sie haben einen Stream, der mehrere Felder filtert und anschließend aus dem reduzierten Dataset Modelle erstellt. Wenn Sie dieselben Modelle auch *ohne* Filterung von Feldern erstellen möchten, um festzustellen, ob sie das Ergebnis verbessern, können Sie den Filterknoten inaktivieren. Wenn Sie den Filterknoten inaktivieren, werden die Verbindungen mit dem Modellierungsknoten direkt vom Ableitungsknoten an den Typknoten übergeben.

So inaktivieren Sie einen Knoten:

1. Klicken Sie im Streamerstellungsbereich mit der rechten Maustaste auf den Knoten, den Sie inaktivieren möchten.
2. Klicken Sie im Popup-Menü auf **Knoten inaktivieren**.

Alternativ können Sie auch im Menü "Bearbeiten" auf **Knoten > Knoten inaktivieren** klicken. Wenn Sie den Knoten wieder in den Stream integrieren möchten, wählen Sie auf dieselbe Weise wie oben beschriebenen **Knoten aktivieren**.

Hinweis: Sie können diese Aktion unter Verwendung der Option **Rückgängig** im Menü "Bearbeiten" oder mit der Tastenkombination Strg+Z rückgängig machen.

Hinzufügen von Knoten in bereits vorhandenen Verbindungen

Zwischen zwei verbundenen Knoten können Sie einen neuen Knoten einfügen, indem Sie den Pfeil ziehen, der die beiden Knoten verbindet.

1. Klicken Sie mit der mittleren Maustaste auf den Verbindungspfeil, in den Sie den Knoten einfügen möchten, und ziehen Sie diesen. Alternativ können Sie die Taste Alt drücken und dann mit der Maus ziehen. Damit simulieren Sie die mittlere Maustaste.
2. Ziehen Sie die Verbindung an den Knoten, den Sie einfügen möchten, und lassen Sie die Maustaste los.

Hinweis: Sie können neue Verbindungen aus dem Knoten entfernen und den ursprünglichen Knoten wiederherstellen, indem Sie den Knoten **umgehen**.

Löschen von Verbindungen zwischen Knoten

So löschen Sie die Verbindung zwischen zwei Knoten:

1. Klicken Sie mit der rechten Maustaste auf den Verbindungspfeil.
2. Klicken Sie im Menü auf **Verbindung löschen**.

Führen Sie zum Löschen aller Verbindungen zu und von einem Knoten eine der folgenden Aktionen aus:

- Wählen Sie den Knoten aus und drücken Sie F3.
- Wählen Sie einen Knoten aus und klicken Sie im Hauptmenü auf:

Bearbeiten > Knoten > Trennen

Festlegen von Optionen für Knoten

Sobald Sie Knoten erstellt und miteinander verbunden haben, gibt es verschiedene Optionen zum Anpassen der Knoten. Klicken Sie mit der rechten Maustaste auf einen Knoten und wählen Sie eine der Menüoptionen aus.

- Klicken Sie auf **Bearbeiten**, um das Dialogfeld für den ausgewählten Knoten zu öffnen.
- Klicken Sie auf **Verbinden**, um einen Knoten manuell mit einem anderen zu verbinden.
- Klicken Sie auf **Trennen**, um alle Verknüpfungen zum und vom Knoten zu löschen.
- Klicken Sie auf **Umbenennen und mit Anmerkung versehen**, um die Registerkarte "Anmerkungen" im Bearbeitungsdialogfeld aufzurufen.

- Klicken Sie auf **Neuer Kommentar**, um einem Knoten einen Kommentar hinzuzufügen. Weitere Informationen finden Sie im Thema „Hinzufügen von Kommentaren und Anmerkungen zu Knoten und Streams“ auf Seite 58.
- Klicken Sie auf **Knoten inaktivieren**, um den Knoten während der Verarbeitung "auszublenden". Um den Knoten wieder "einzublenden", klicken Sie auf **Knoten aktivieren**. Weitere Informationen finden Sie im Thema „Inaktivieren von Knoten in einem Stream“ auf Seite 39.
- Klicken Sie auf **Ausschneiden** oder **Löschen**, um die ausgewählten Knoten aus dem Streamerstellungsbereich zu entfernen. *Hinweis:* Mit der Option **Ausschneiden** können Sie Knoten einfügen, nicht jedoch mit **Löschen**.
- Klicken Sie auf **Knoten kopieren**, um eine Kopie des Knotens ohne Verbindungen zu erstellen. Diese kann einem neuen oder bereits vorhandenen Stream hinzugefügt werden.
- Klicken Sie auf **Knoten laden**, um einen zuvor gespeicherten Knoten zu öffnen und seine Optionen in den aktuell ausgewählten Knoten zu laden. *Hinweis:* Die Knoten müssen denselben Typ aufweisen.
- Klicken Sie auf **Knoten abrufen**, um einen Knoten von einem verbundenen IBM SPSS Collaboration and Deployment Services Repository abzurufen.
- Klicken Sie auf **Knoten speichern**, um die Knotendetails in einer Datei zu speichern. Sie können Knotendetails nur in einen anderen Knoten desselben Typs laden.
- Klicken Sie auf **Knoten speichern**, um den ausgewählten Knoten in einem verbundenen IBM SPSS Collaboration and Deployment Services Repository zu speichern.
- Klicken Sie auf **Cache**, um das Menü zu erweitern, das Optionen zum Zwischenspeichern des ausgewählten Knotens enthält.
- Klicken Sie auf **Datenzuordnung**, um das Menü zu erweitern, das Optionen zum Zuordnen von Daten zu einer neuen Quelle oder zum Festlegen von erforderlichen Feldern enthält.
- Klicken Sie auf **Superknoten erstellen**, um das Menü zu erweitern, das Optionen zum Erstellen eines Superknotens im aktuellen Stream enthält.
- Klicken Sie auf **Benutzereingabeknoten generieren**, um den ausgewählten Knoten zu ersetzen. Die von diesem Knoten erstellten Beispiele enthalten dieselben Felder wie der aktuelle Knoten.
- Klicken Sie auf **Ab hier ausführen**, um alle Endknoten abwärts vom ausgewählten Knoten auszuführen.

Zwischenspeichern von Optionen für Knoten

Um die Streamausführung zu optimieren, können Sie einen **Cache** für Knoten einrichten, die keine Endknoten sind. Wenn Sie einen Cache für einen Knoten einrichten, wird der Cache mit den Daten gefüllt, die bei der nächsten Ausführung des Datenstreams durch den Knoten laufen. Ab diesem Zeitpunkt werden die Daten aus dem Cache (der sich in einem temporären Verzeichnis auf dem Datenträger befindet) und nicht mehr aus der Datenquelle eingelesen.

Zwischenspeichern ist am nützlichsten im Anschluss an zeitaufwendige Operationen, wie Sortieren, Zusammenführen oder Aggregation. Angenommen, es ist beispielsweise ein Quellenknoten vorhanden, der zum Lesen von Umsatzdaten aus einer Datenbank eingerichtet ist, und ein Aggregatknoten, der den Umsatz nach Standort zusammenfasst. Sie können einen Cache an dem Aggregatknoten einrichten anstatt an dem Quellenknoten, weil Sie möchten, dass der Cache die aggregierten Daten und nicht das gesamte Datenset speichert.

Hinweis: Ein Caching bei Quellenknoten, bei denen einfach eine Kopie der ursprünglichen Dateien, die in IBM SPSS Modeler eingelesen wurden, gespeichert wird, bewirkt meist keine Leistungsverbesserung.

Knoten mit aktiviertem Caching werden mit einem kleinen Dokumentsymbol in der oberen rechten Ecke angezeigt. Wenn die Daten am Knoten zwischengespeichert werden, ist das Dokumentsymbol grün.

So aktivieren Sie einen Cache:

1. Klicken Sie im Streamerstellungsbereich mit der rechten Maustaste auf den Knoten und klicken Sie im Menü auf die Option **Cache**.

2. Klicken Sie im Untermenü auf die Option **Aktivieren**.
3. Sie können den Cache inaktivieren, indem Sie mit der rechten Maustaste auf den Knoten klicken und im Untermenü auf **Inaktivieren** klicken.

Zwischenspeichern von Knoten in einer Datenbank

Bei in einer Datenbank ausgeführten Streams muss nicht das Dateisystem als Cache verwendet werden, sondern die Daten können mitten im Stream in einer temporären Tabelle in der Datenbank zwischengespeichert werden. Bei einer Kombination mit SQL-Optimierung kann dies zu einer erheblichen Leistungssteigerung führen. So kann beispielsweise die Ausgabe aus einem Stream, der mehrere Tabellen zusammenfasst, um eine Data-Mining-Ansicht zu erstellen, im Cache gespeichert und bei Bedarf wieder verwendet werden. Durch die automatische Erzeugung von SQL für alle nachgeordneten Knoten lässt sich die Leistung weiter erhöhen.

Bei Verwendung von Datenbankcaching mit Zeichenfolgen, die mehr als 255 Zeichen umfassen, müssen Sie entweder sicherstellen, dass sich ein Typknoten im Stream oberhalb des Caching-Knotens befindet und dass die Feldwerte gelesen werden, oder Sie müssen die Zeichenfolgelänge mithilfe des Parameters `default_sql_string_length` in der Datei `options.cfg` festlegen. Dies stellt sicher, dass die entsprechende Spalte in der temporären Tabelle auf die korrekte Breite eingestellt wird, um die Zeichenfolgen anzuzeigen.

Um das Caching von Datenbanken nutzen zu können, müssen sowohl SQL-Optimierung als auch Datenbankcaching aktiviert sein. Die Optimierungseinstellungen auf dem Server haben Vorrang vor jenen auf dem Client. Weitere Informationen finden Sie im Thema „Festlegen von Optimierungsoptionen für Streams“ auf Seite 47.

Klicken Sie bei aktiviertem Datenbankcaching einfach mit der rechten Maustaste auf einen Knoten (keinen Endknoten), um die Daten an dieser Stelle im Cache zu speichern, und der Cache wird bei der nächsten Ausführung des Streams automatisch unmittelbar in der Datenbank erstellt. Wenn Datenbankcaching oder SQL-Optimierung nicht aktiviert ist, wird der Cache stattdessen in das Dateisystem geschrieben.

Hinweis: Die folgenden Datenbanken unterstützen temporäre Tabellen zu Caching-Zwecken: DB2, Netezza, Oracle, SQL Server und Teradata. Andere Datenbanken verwenden für das Datenbankcaching eine normale Tabelle. Der SQL-Code kann für bestimmte Datenbanken angepasst werden. Wenden Sie sich an den Technical Support, um Unterstützung anzufordern.

So leeren Sie einen Cache:

Ein weißes Dokumentsymbol an einem Knoten weist darauf hin, dass sein Cache leer ist. Wenn der Cache voll ist, wird das Dokumentsymbol grün. Wenn Sie den Inhalt des Caches ersetzen möchten, müssen Sie den Cache zunächst leeren und den Datenstream dann erneut ausführen, um ihn erneut zu füllen.

1. Klicken Sie im Streamerstellungsbereich mit der rechten Maustaste auf den Knoten und klicken Sie im Menü auf die Option **Cache**.
2. Wählen Sie im Untermenü die Option **Leeren** aus.

So speichern Sie einen Cache:

Sie können den Inhalt eines Cache als Datendatei von IBM SPSS Statistics (*.sav) speichern. Sie können die Datei entweder als Cache erneut laden oder Sie können einen Knoten einrichten, der die Cachedatei als Datenquelle verwendet. Sie können auch einen Cache laden, den Sie aus einem anderen Projekt gespeichert haben.

1. Klicken Sie im Streamerstellungsbereich mit der rechten Maustaste auf den Knoten und klicken Sie im Menü auf die Option **Cache**.
2. Wählen Sie im Untermenü die Option **Cache speichern** aus.

3. Blättern Sie im Dialogfeld "Cache speichern" an den Speicherort, an dem die Cachedatei gespeichert werden soll.
4. Geben Sie im Textfeld "Dateiname" einen Namen ein.
5. Stellen Sie sicher, dass in der Liste "Dateityp" die Option ***.sav** ausgewählt ist, und klicken Sie auf **Speichern**.

So laden Sie einen Cache:

Wenn Sie eine Cachedatei gespeichert haben, bevor Sie diese aus dem Knoten entfernen, können Sie sie erneut laden.

1. Klicken Sie im Streamerstellungsbereich mit der rechten Maustaste auf den Knoten und klicken Sie im Menü auf die Option **Cache**.
2. Wählen Sie im Untermenü die Option **Cache laden** aus.
3. Blättern Sie im Dialogfeld "Cache laden" an den Speicherort der Cachedatei, wählen Sie diese aus und klicken Sie auf **Laden**.

Vorschau von Daten in Knoten

Damit sichergestellt ist, dass Daten beim Erstellen eines Streams wie erwartet geändert werden, können Sie alle Ihre Daten bei jedem wesentlichen Schritt durch einen Tabellenknoten laufen lassen. Um sich das zu ersparen, können Sie von jedem Knoten eine Vorschau mit einem Muster der Daten erzeugen, die erstellt werden, und damit den Zeitaufwand für die Erstellung jedes Knotens verringern.

Für Knoten oberhalb eines Modellnuggets zeigt die Vorschau die Eingabefelder an. Für ein Modellnugget oder Knoten unterhalb des Nuggets (ausgenommen Endknoten) zeigt die Vorschau die Eingabefelder und die generierten Felder an.

Die Standardanzahl der angezeigten Zeilen ist 10. Sie können diese Einstellung jedoch in den Streameigenschaften ändern. Weitere Informationen finden Sie im Thema „Festlegen allgemeiner Optionen für Streams“ auf Seite 44.

Mithilfe des Menüs "Generieren" können Sie mehrere Arten von Knoten erzeugen.

Sperrern von Knoten

Um andere Benutzer daran zu hindern, die Einstellungen eines oder mehrerer Knoten in einem Stream zu ändern, können Sie den oder die Knoten im besonderen Knotentyp namens Superknoten verkapseln und dann den Superknoten durch Zuweisen eines Kennwortschutzes sperren.

Arbeiten mit Streams

Sobald Sie die Quell-, Prozess- und Endknoten im Streamerstellungsbereich verbunden haben, haben Sie einen Stream erstellt. Als Knotensammlung können Streams gespeichert, mit Anmerkungen versehen und zu Projekten hinzugefügt werden. Sie können für Streams auch zahlreiche Optionen wie Optimierung, Datums- und Uhrzeiteinstellungen, Parameter und Scripts festlegen. Diese Eigenschaften werden in den nachfolgenden Themen beschrieben.

In IBM SPSS Modeler können Sie innerhalb derselben IBM SPSS Modeler-Sitzung mehrere Datenstreams verwenden und ändern. Die rechte Seite des Hauptfensters enthält den Managerbereich, mit dem Sie durch die aktuell geöffneten Streams, Ausgaben und Modelle navigieren können. Wenn der Managerbereich nicht sichtbar ist, wählen Sie im Ansichtsmenü die Option **Manager** aus und klicken Sie dann auf die Registerkarte **Streams**.

Auf dieser Registerkarte können Sie:

- Auf Streams zugreifen.
- Streams speichern.
- Streams im aktuellen Projekt speichern.

- Streams schließen.
- Neue Streams öffnen.
- Streams aus einem Repository von IBM SPSS Collaboration and Deployment Services (falls an Ihrem Standort verfügbar) speichern und abrufen. Weitere Informationen finden Sie im Thema „Informationen zu IBM SPSS Collaboration and Deployment Services Repository“ auf Seite 125.

Klicken Sie mit der rechten Maustaste auf die Registerkarte "Streams", um diese Optionen aufzurufen.

Festlegen von Optionen für Streams

Sie können eine Reihe von Optionen angeben, die auf den aktuellen Stream angewendet werden sollen. Sie können diese Optionen auch als Standard zur Anwendung auf alle Ihre Streams speichern. Es handelt sich hier um die folgenden Optionen:

- **Allgemein.** Verschiedene Optionen, wie beispielsweise die im Stream zu verwendenden Symbole und die zu verwendende Textcodierung. Weitere Informationen finden Sie im Thema „Festlegen allgemeiner Optionen für Streams“.
- **Datum/Uhrzeit.** Optionen, die sich auf das Format von Datums- und Zeitausdrücken beziehen. Weitere Informationen finden Sie im Thema „Festlegen von Datums- und Zeitoptionen für Streams“ auf Seite 46.
- **Zahlenformate.** Optionen, die das Format numerischer Ausdrücke festlegen. Weitere Informationen finden Sie im Thema „Festlegen von Zahlenformatoptionen für Streams“ auf Seite 46.
- **Optimierung.** Optionen zur Optimierung der Streamleistung. Weitere Informationen finden Sie im Thema „Festlegen von Optimierungsoptionen für Streams“ auf Seite 47.
- **Protokollierung und Status.** Optionen zur Festlegung von SQL-Protokollierung und Datensatzstatus. Weitere Informationen finden Sie im Thema „Festlegen von Optionen für SQL-Protokollierung und Datensatzstatus für Streams“ auf Seite 48.
- **Layout.** Optionen, die sich auf das Layout des Streams im Erstellungsbereich beziehen. Weitere Informationen finden Sie im Thema „Festlegen von Layoutoptionen für Streams“ auf Seite 49.

So legen Sie Streamoptionen fest:

1. Klicken Sie im Menü "Datei" auf die Option **Streameigenschaften** (oder wählen Sie den Stream auf der Registerkarte "Streams" im Managerbereich aus, klicken Sie mit der rechten Maustaste und wählen Sie dann im Pop-up-Menü die Option **Streameigenschaften** aus).
2. Klicken Sie auf die Registerkarte **Optionen**.

Klicken Sie als Alternative im Menü "Extras" auf:

Streameigenschaften > Optionen

Festlegen allgemeiner Optionen für Streams: Die allgemeinen Optionen sind ein Set verschiedener Optionen, die für verschiedene Aspekte des aktuellen Streams gelten.

Dezimalzeichen. Wählen Sie entweder ein Komma (,) oder einen Punkt (.) als Dezimaltrennzeichen aus.

Symbol für Zifferngruppierung. Als Zahlenanzeigeformat wählen Sie das Symbol aus, das zur Gruppierung von Werten verwendet wird (z. B. der Punkt in 3.000,00). Zu den Optionen gehören "Keine", "Punkt", "Komma", "Leerzeichen" und "Durch Ländereinstellung definiert" (in diesem Fall wird der Standardwert für die aktuelle Ländereinstellung verwendet).

Codierung. Geben Sie die Streamstandardmethode für die Textcodierung an. (*Hinweis:* Gilt nur für den Quellenknoten "Variable Datei" und den Exportknoten "Flatfile". Andere Knoten verwenden diese Einstellung nicht; die meisten Datendateien verfügen über eingebettete Codierungsinformationen.) Sie können entweder den Systemstandardwert oder "UTF-8" auswählen. Der Systemstandardwert wird in der Win-

dows-Systemsteuerung oder bei einer Ausführung im verteilten Modus auf dem Server-Computer angegeben. Weitere Informationen finden Sie unter im Thema „Unicode-Unterstützung in IBM SPSS Modeler“ auf Seite 185.

Regelsetauswertung. Legt fest, wie Regelsetmodelle ausgewertet werden. Standardmäßig verwenden Regelsets die Option **Voting**, um Vorhersagen aus einzelnen Regeln zu kombinieren und die endgültige Vorhersage zu bestimmen. Um sicherzustellen, dass Regelsets standardmäßig die Regel "Erster Treffer" verwenden, wählen Sie **Erster Treffer** aus. Beachten Sie, dass diese Option nicht für Entscheidungslistenmodelle gilt, die die Regel "Erster Treffer" wie im Algorithmus definiert immer verwenden.

Maximale Anzahl an Zeilen in Datenvorschau. Geben Sie die Anzahl der Zeilen an, die angezeigt werden sollen, wenn eine Vorschau der Daten für einen Knoten angefordert wird. Weitere Informationen finden Sie unter im Thema „Vorschau von Daten in Knoten“ auf Seite 43.

Maximale Mitglieder für nominale Felder. Wählen Sie diese Option aus, um eine maximale Anzahl an Mitgliedern für nominale (festgelegte) Felder anzugeben, nach der sich der Datentyp des Feldes in **Ohne Typ** ändert. Diese Option ist hilfreich, wenn Sie mit großen nominalen Feldern arbeiten. *Hinweis:* Wenn das Messniveau eines Feldes auf **Ohne Typ** gesetzt wird, wird seine Rolle automatisch auf **Keine** gesetzt. Dies bedeutet, dass die Felder nicht zur Modellierung verfügbar sind.

Satzgröße für Kohonen- und K-Means-Modellierung begrenzen. Wählen Sie diese Option aus, um eine maximale Anzahl an Mitgliedern für nominale Felder anzugeben, die bei der Kohonen-Netz- und K-Means-Modellierung verwendet werden. Die Standardsatzgröße beträgt 20. Danach wird das Feld ignoriert und es wird eine Warnung angezeigt, in der Informationen zu dem betreffenden Feld angegeben werden.

Beachten Sie, dass aus Kompatibilitätsgründen diese Option auch für den alten Knoten "Neuronales Netz" gilt, der in IBM SPSS Modeler Version 14 ersetzt wurde; einige alte Streams verfügen möglicherweise immer noch über diesen Knoten.

Quellenknoten bei Ausführung aktualisieren. Wählen Sie diese Option aus, um bei der Ausführung des aktuellen Streams alle Quellenknoten automatisch zu aktualisieren. Diese Aktion entspricht dem Klicken auf die Schaltfläche **Aktualisieren** auf einem Quellenknoten, mit der Ausnahme, dass mit dieser Option automatisch alle Quellenknoten (außer die Benutzereingabeknoten) für den aktuellen Stream aktualisiert werden.

Hinweis: Wenn Sie diese Option auswählen, werden die Caches von nachgeordneten Knoten geleert, auch wenn sich die Daten nicht geändert haben. Die Caches werden jedoch nur einmal pro Streamausführung geleert, sodass Sie nachgeordnete Caches weiterhin als temporären Speicher für eine einzelne Ausführung verwenden können. Nehmen wir z. B. an, dass Sie einen Cache in die Mitte des Streams nach einer komplexen Ableitungsoperation gesetzt haben und Sie mehrere Grafiken und Berichte auf diesen Ableitungsknoten folgend angefügt haben. Wenn Sie den Stream ausführen, wird der Cache am Ableitungsknoten geleert und neu gefüllt, jedoch nur für die erste Grafik oder den ersten Bericht. Nachfolgende Endknoten lesen die Daten aus dem Cache des Ableitungsknotens.

Feld- und Wertbeschriftungen in Ausgabe anzeigen. Zeigt Feld- und Wertbeschriftungen in Tabellen, Diagrammen und anderen Ausgabeformen an. Wenn keine Beschriftungen vorhanden sind, werden stattdessen die Feldnamen und Datenwerte angezeigt. Die Beschriftungen sind standardmäßig ausgeschaltet; Sie können die Beschriftungen jedoch individuell an anderer Stelle in IBM SPSS Modeler ein- und ausschalten. Sie können mit einer Schaltfläche zum Ein-/Ausschalten in der Symbolleiste auch auswählen, dass die Beschriftungen im Ausgabefenster angezeigt werden.



Abbildung 11. Symbolleiste zum Ein-/Ausschalten von Feld- und Wertbeschriftungen

Ausführungszeiten anzeigen. Zeigt nach der Ausführung des Streams einzelne Ausführungszeiten für Streamknoten auf der Registerkarte "Ausführungszeiten" an. Weitere Informationen finden Sie unter im Thema „Anzeigen der Knotenausführungszeiten“ auf Seite 50.

Als Standard speichern. Die angegebenen Optionen gelten nur für den aktuellen Stream. Klicken Sie auf diese Schaltfläche, um diese Optionen als Standardwert für alle Streams festzulegen.

Festlegen von Datums- und Zeitoptionen für Streams: Diese Optionen geben das für verschiedene Datums- und Zeitausdrücke im aktuellen Stream zu verwendende Format an.

Datum/Uhrzeit importieren als. Wählen Sie, ob der Speichertyp "Datum/Uhrzeit" für Datums-/Uhrzeitfelder verwendet werden soll oder ob sie als Zeichenfolgevariablen importiert werden sollen.

Datumsformat. Wählen Sie ein Datumsformat aus, das für die Datumsspeicherfelder verwendet werden soll oder wenn Zeichenfolgen von den CLEM-Datumsfunktionen als Datumsangaben interpretiert werden.

Zeitformat. Wählen Sie ein Zeitformat aus, das für die Zeitspeicherfelder verwendet werden soll oder wenn Zeichenfolgen von den CLEM-Zeitfunktionen als Zeitangaben interpretiert werden.

Tage/Minuten übertragen. Wählen Sie für Zeitformate aus, ob negative Zeitunterschiede so interpretiert werden sollen, dass sie sich auf den vorherigen Tag bzw. die vorherige Stunde beziehen.

Datumsbasis (1. Jan.). Wählen Sie die Basisjahre (immer 1. Jan.), die von den CLEM-Datumsfunktionen verwendet werden müssen, die mit einem einzelnen Datum arbeiten.

Zweistellige Datumsangaben ab. Geben Sie das Schwellenjahr an, ab dem Jahrhundertziffern für Jahre hinzugefügt werden, die mit lediglich zwei Ziffern angegeben sind. Wenn Sie beispielsweise 1930 als Schwellenjahr festlegen, wird angenommen, dass das Datum 05/11/02 im Jahr 2002 liegt. Für Daten nach 30 wird das 20. Jahrhundert verwendet, d. h., bei 05/11/73 wird angenommen, dass es sich um ein Datum aus dem Jahr 1973 handelt.

Als Standard speichern. Die angegebenen Optionen gelten lediglich für den aktuellen Stream. Klicken Sie auf diese Schaltfläche, um diese Optionen als Standard für alle Streams festzulegen.

Festlegen von Zahlenformatoptionen für Streams: Diese Optionen geben das für verschiedene numerische Ausdrücke im aktuellen Stream zu verwendende Format an.

Zahlenanzeigeformat. Sie können aus den Anzeigeformaten Standard (####.###), Wissenschaftlich (#.###E+###) und Währung (\$###.##) wählen.

Dezimalstellen (Standard, Wissenschaftlich, Währung). Gibt bei Zahlenanzeigeformaten an, wie viele Dezimalstellen bei der Anzeige oder beim Drucken reeller Zahlen verwendet werden sollen. Diese Option wird getrennt für jedes Anzeigeformat angegeben.

Berechnungen in. Wählen Sie **Radiant** oder **Grad** als Maßeinheit für trigonometrische CLEM-Ausdrücke. Weitere Informationen finden Sie im Thema „Trigonometrische Funktionen“ auf Seite 102.

Als Standard speichern. Die angegebenen Optionen gelten lediglich für den aktuellen Stream. Klicken Sie auf diese Schaltfläche, um diese Optionen als Standard für alle Streams festzulegen.

Festlegen von Optimierungsoptionen für Streams: Sie können die Optimierungseinstellungen verwenden, um die Streamleistung zu optimieren. Beachten Sie, dass die Leistungs- und Optimierungseinstellungen auf IBM SPSS Modeler Server (sofern verwendet) alle entsprechenden Einstellungen auf dem Client außer Kraft setzen.

Hinweis: Für Datenbankmodellierung und SQL-Optimierung muss auf dem IBM SPSS Modeler-Computer die IBM SPSS Modeler Server-Konnektivität aktiviert sein. Wenn diese Einstellung aktiviert ist, können Sie auf Datenbankalgorithmen zugreifen, SQL direkt aus IBM SPSS Modeler per Pushback übertragen und auf IBM SPSS Modeler Server zugreifen. Wählen Sie zur Überprüfung des aktuellen Lizenzstatus die folgenden Optionen aus dem IBM SPSS Modeler-Menü aus.

Hilfe > Info > Weitere Details

Wenn Konnektivität aktiviert ist, wird auf der Registerkarte "Lizenzstatus" die Option **Serveraktivierung** angezeigt.

Weitere Informationen finden Sie im Thema „Verbinden mit IBM SPSS Modeler Server“ auf Seite 12.

Hinweis: Ob SQL-Pushback und Optimierung unterstützt werden, hängt vom verwendeten Datenbanktyp ab. Neueste Informationen zu Datenbanken und ODBC-Treibern, die für die Verwendung mit IBM SPSS Modeler 16 unterstützt werden und getestet wurden, finden Sie auf der unternehmensweiten Support-Site unter <http://www.ibm.com/support>.

Streamneuerstellung aktivieren. Wählen Sie diese Option aus, um die Streamneuerstellung in IBM SPSS Modeler zu aktivieren. Es stehen vier Neuerstellungsarten zur Verfügung, von denen Sie eine oder mehrere auswählen können. Die Streamneuerstellung ordnet die Knoten in einem Stream für eine effizientere Operation im Hintergrund neu an, ohne die Streamsemantik zu ändern.

- **SQL-Erzeugung optimieren.** Mit dieser Option können die Knoten im Stream so neu angeordnet werden, dass ein Pushback weiterer Operationen mithilfe der SQL-Generierung für die Ausführung in der Datenbank möglich ist. Wenn die Optimierungsfunktion einen Knoten findet, der nicht in SQL wiedergegeben werden kann, sucht sie nachgeordnete Knoten, die in SQL wiedergegeben und ohne Beeinträchtigung der Streamsemantik vor den problematischen Knoten verschoben werden können. Nicht nur, dass die Datenbank die Operationen effizienter als IBM SPSS Modeler durchführen kann, solche Pushbacks dienen auch dazu, die Größe des Datensets, das an IBM SPSS Modeler zur Verarbeitung zurückgegeben wird, zu reduzieren. So wird wiederum der Netzverkehr reduziert und die Geschwindigkeit der Streamoperationen erhöht. Das Kontrollkästchen **SQL generieren** muss aktiviert sein, damit die SQL-Optimierung funktionieren kann.
- **CLEM-Ausdruck optimieren.** Mit dieser Option kann das Optimierungsprogramm nach CLEM-Ausdrücken suchen, die vor der Ausführung des Streams vorverarbeitet werden können, damit die Verarbeitungsgeschwindigkeit erhöht wird. Ein einfaches Beispiel dafür ist, dass das Optimierungsprogramm bei einem Ausdruck wie $\log(\text{Gehalt})$ den tatsächlichen Gehaltswert berechnen und diesen dann für die Verarbeitung übergeben würde. Dies kann die Leistung von SQL-Pushback und IBM SPSS Modeler Server erhöhen.
- **Syntaxausführung optimieren.** Diese Methode der Streamneuerstellung erhöht die Effizienz von Operationen, die mehr als einen Knoten mit IBM SPSS Statistics-Syntax umfassen. Die Optimierung wird durch die Zusammenfassung der Syntaxbefehle in einer einzigen Operation erreicht, anstatt jeden einzelnen Syntaxbefehl als separate Operation auszuführen.
- **Andere Ausführung optimieren.** Diese Methode der Streamneuerstellung erhöht die Effizienz von Operationen, die nicht an die Datenbank delegiert werden können. Eine Optimierung wird dadurch erreicht, dass die Datenmenge im Stream so früh wie möglich reduziert wird. Während die Datenintegrität gewahrt bleibt, wird der Stream neu erstellt, um die Operationen näher zur Datenquelle zu bringen. So werden die Daten abwärts für kostenintensive Operationen, z. B. Joins, reduziert.

Parallele Verarbeitung aktivieren. Bei Ausführung auf einem Computer mit mehreren Prozessoren ermöglicht diese Option dem System, die Last auf diese Prozessoren zu verteilen, was zu einer schnelleren

Leistung führen kann. Durch die Verwendung mehrerer Knoten oder der folgenden Einzelknoten können Sie von der Parallelverarbeitung profitieren: C5.0, Zusammenführen (nach Schlüssel), Sortieren, Klasse (Rangbildungsmethode und N-Perzentil-Methode) sowie Aggregieren (unter Verwendung mindestens eines Schlüsselfelds).

SQL generieren. Wählen Sie diese Option zur Aktivierung der SQL-Generierung. Damit wird ermöglicht, dass Streamoperationen per Pushback an die Datenbank zurückübertragen werden, indem SQL-Code zur Generierung der Ausführungsprozesse verwendet wird, was zu einer Leistungsverbesserung führen kann. Zur weiteren Steigerung der Leistung kann außerdem **SQL-Erzeugung optimieren** ausgewählt werden. Dadurch wird die Anzahl der Operationen, die per Pushback an die Datenbank zurückübertragen werden, maximiert. Wenn Operationen für einen Knoten per Pushback an die Datenbank zurückübertragen wurden, wird der Knoten bei der Ausführung des Streams violett gekennzeichnet.

- **Datenbankcaching.** Bei Streams, die in der Datenbank auszuführende SQL generieren, muss nicht das Dateisystem als Cache verwendet werden, sondern die Daten können mitten im Stream in einer temporären Tabelle in der Datenbank zwischengespeichert werden. Bei einer Kombination mit SQL-Optimierung kann dies zu einer erheblichen Leistungssteigerung führen. So kann beispielsweise die Ausgabe aus einem Stream, der mehrere Tabellen zusammenfasst, um eine Data-Mining-Ansicht zu erstellen, im Cache gespeichert und bei Bedarf wieder verwendet werden. Klicken Sie bei aktiviertem Datenbankcaching einfach mit der rechten Maustaste auf einen Knoten (keinen Endknoten), um die Daten an dieser Stelle im Cache zu speichern, und der Cache wird bei der nächsten Ausführung des Streams automatisch unmittelbar in der Datenbank erstellt. Dadurch kann SQL für nachgeordnete Knoten erzeugt werden, was zu weiteren Leistungsverbesserungen führt. (Alternativ kann diese Option bei Bedarf inaktiviert werden, beispielsweise wenn es aufgrund von Richtlinien oder Berechtigungen nicht möglich ist, Daten in die Datenbank zu schreiben. Wenn Datenbankcaching oder SQL-Optimierung nicht aktiviert ist, wird der Cache stattdessen in das Dateisystem geschrieben. Weitere Informationen finden Sie im Thema „Zwischenspeichern von Optionen für Knoten“ auf Seite 41.
- **Unproblematische Umwandlung verwenden.** Mit dieser Option können Daten von Zeichenfolgen in Zahlen oder von Zahlen in Zeichenfolgen konvertiert werden, sofern sie in einem geeigneten Format gespeichert sind. Wenn die Daten beispielsweise als Zeichenfolge in der Datenbank gespeichert sind, tatsächlich jedoch eine sinnvolle Zahl enthalten, können die Daten beim Eintreten des Pushbacks zur Verwendung konvertiert werden.

Hinweis: Aufgrund kleinerer Unterschiede in der SQL-Implementierung weichen die Ergebnisse der in einer Datenbank ausgeführten Streams möglicherweise leicht von den Ergebnissen bei Ausführung in IBM SPSS Modeler ab. Aus ähnlichen Gründen können diese Unterschiede auch je nach Datenbankanbieter variieren.

Als Standard speichern. Die angegebenen Optionen gelten lediglich für den aktuellen Stream. Klicken Sie auf diese Schaltfläche, um diese Optionen als Standard für alle Streams festzulegen.

Festlegen von Optionen für SQL-Protokollierung und Datensatzstatus für Streams: Zu diesen Einstellungen gehören verschiedene Optionen zur Festlegung der Anzeige der vom Stream generierten SQL-Anweisungen sowie der Anzeige für die Anzahl der vom Stream verarbeiteten Datensätze.

SQL wird während der Streamausführung im Nachrichtenprotokoll angezeigt. Gibt an, ob die während der Ausführung des Streams erzeugte SQL an das Nachrichtenprotokoll übergeben wird.

Einzelheiten zur SQL-Erzeugung während der Streamvorbereitung im Nachrichtenprotokoll anzeigen. Legt während der Streamvorschau fest, ob eine Vorschau der SQL, die generiert werden würde, an das Nachrichtenprotokoll übergeben wird.

SQL anzeigen. Gibt an, ob ein im Protokoll angezeigtes SQL systemeigene SQL-Funktionen oder Standard-ODBC-Funktionen in der Form {fn FUNC(...)}, wie von IBM SPSS Modeler erzeugt, enthalten soll. Ersteres richtet sich nach den ODBC-Treiberfunktionen, die u. U. nicht implementiert werden. Dieses Steuerelement hätte z. B. für SQL Server keinerlei Auswirkungen.

SQL zur Verbesserung der Lesbarkeit neu formatieren. Gibt an, ob im Protokoll angezeigtes SQL zur Verbesserung der Lesbarkeit formatiert werden soll.

Status für Datensätze anzeigen. Gibt an, ob Datensätze gemeldet werden sollen, wenn sie an Endknoten ankommen. Geben Sie eine Zahl an, um den Status alle N Datensätze zu aktualisieren.

Als Standard speichern. Die angegebenen Optionen gelten lediglich für den aktuellen Stream. Klicken Sie auf diese Schaltfläche, um diese Optionen als Standard für alle Streams festzulegen.

Festlegen von Layoutoptionen für Streams: Diese Einstellungen bieten eine Reihe von Optionen, die sich auf Anzeige und Verwendung des Streamerstellungsbereichs beziehen.

Minimale Breite des Streamerstellungsbereichs. Geben Sie die minimale Breite des Streamerstellungsbereichs in Pixeln an.

Minimale Höhe des Streamerstellungsbereichs. Geben Sie die minimale Höhe des Streamerstellungsbereichs in Pixeln an.

Streamblättersgeschwindigkeit. Geben Sie die Blättersgeschwindigkeit für den Streamerstellungsbereich an, um die Blättersgeschwindigkeit des Streamerstellungsbereichs zu steuern, wenn ein Knoten von einer Position im Erstellungsbereich an eine andere gezogen wird. Je höher die Zahl, desto höher die Blättersgeschwindigkeit.

Maximum für Symbolnamen. Geben Sie eine Begrenzung in der Anzahl Zeichen für die Namen von Knoten im Streamerstellungsbereich ein.

Symbolgröße. Wählen Sie eine Option aus, um die gesamte Streamansicht auf eine Anzahl an Größwerten zwischen 8 und 200 % der Standardsymbolgröße zu skalieren.

Rasterzellengröße. Wählen Sie eine Rasterzellengröße in der Liste aus. Diese Zahl wird für die Ausrichtung von Knoten im Streamerstellungsbereich mithilfe eines unsichtbaren Rasters verwendet. Die Standardrasterzellengröße lautet 0,25.

Am Raster ausrichten. Wählen Sie diese Option aus, um Symbole an einem unsichtbaren Rastermuster auszurichten (Standardauswahl).

Generierte Symbolplatzierung. Wählen Sie im Erstellungsbereich die gewünschte Position für Symbole der Knoten, die aus Modellnuggets erzeugt wurden. Standard ist oben links.

Als Standard speichern. Die angegebenen Optionen gelten lediglich für den aktuellen Stream. Klicken Sie auf diese Schaltfläche, um diese Optionen als Standard für alle Streams festzulegen.

Analytic Server-Streameigenschaften: Diese Einstellungen bieten eine Reihe von Optionen für die Arbeit mit Analytic Server.

Maximale Anzahl der Datensätze, die außerhalb von Analytic Server zu verarbeiten sind

Geben Sie die maximale Anzahl der Datensätze an, die aus einer Analytic Server-Datenquelle auf den SPSS Modeler-Server importiert werden sollen.

Nachricht, wenn ein Knoten nicht in Analytic Server verarbeitet werden kann

Diese Einstellung bestimmt, was geschieht, wenn ein Stream, der an Analytic Server übergeben würde, einen Knoten enthält, der in Analytic Server nicht verarbeitet werden kann. Geben Sie an, ob eine Warnung abgesetzt und die Verarbeitung des Streams fortgesetzt werden soll oder ob ein Fehler ausgelöst und die Verarbeitung gestoppt werden soll.

Speichereinstellungen für aufgeteilte Modelle

Modell nach Verweis in Analytic Server speichern, wenn die Modellgröße (MB) übersteigt
Modellnuggets werden in der Regel als Teil des Streams gespeichert. Aufgeteilte Modelle mit vielen Aufteilungen können große Nuggets erzeugen und das Vor- und Zurückschieben des Nuggets zwischen dem Stream und Analytic Server kann sich auf die Leistung auswirken. Wenn ein aufgeteiltes Modell die angegebene Größe überschreitet, wird es daher in Analytic Server gespeichert und das Nugget in SPSS Modeler enthält einen Verweis auf das Modell.

Standardordner, in dem Modelle nach Verweis in Analytic Server gespeichert werden, wenn die Ausführung beendet ist

Geben Sie den Standardpfad an, unter dem Sie aufgeteilte Modelle in Analytic Server speichern wollen. Der Pfad sollte mit einem gültigen Analytic Server-Projektnamen beginnen.

Ordner zum Speichern hochgestufter Modelle

Geben Sie den Standardpfad an, unter dem Sie "hochgestufte" Modelle speichern wollen. Ein hochgestuftes Modell wird nicht bereinigt, wenn die SPSS Modeler-Sitzung vorüber ist.

Anzeigen von Streamoperationsnachrichten

Nachrichten zu Streamoperationen wie Ausführung, Optimierung und vergangene Zeit für die Modellerstellung und -evaluierung können problemlos mithilfe der Registerkarte "Nachrichten" im Dialogfeld "Streameigenschaften" angezeigt werden. In dieser Tabelle werden auch Fehlernachrichten gespeichert.

So zeigen Sie Streamnachrichten an:

1. Klicken Sie im Menü "Datei" auf die Option **Streameigenschaften** (oder wählen Sie den Stream auf der Registerkarte "Streams" im Managerbereich aus, klicken Sie mit der rechten Maustaste und wählen Sie dann im Pop-up-Menü die Option **Streameigenschaften** aus).
2. Klicken Sie auf die Registerkarte **Nachrichten**.

Klicken Sie als Alternative im Menü "Extras" auf:

Streameigenschaften > Nachrichten

Zusätzlich zu den Nachrichten zu Streamoperationen werden hier Fehlernachrichten gespeichert. Wenn die Streamausführung aufgrund eines Fehlers beendet wird, wird in diesem Dialogfeld die Registerkarte "Nachrichten" mit der sichtbaren Fehlernachricht geöffnet. Außerdem wird der Knoten mit den Fehlern im Streamerstellungsbereich rot markiert.

Falls die Optionen für die SQL-Optimierung und Protokollierung im Dialogfeld "Benutzeroptionen" aktiviert sind, werden auch Informationen über die generierte SQL angezeigt. Weitere Informationen finden Sie im Thema „Festlegen von Optimierungsoptionen für Streams“ auf Seite 47.

Sie können hier für einen Stream gemeldete Nachrichten speichern, indem Sie in der Dropdown-Liste der Schaltfläche "Speichern" (links, direkt unter dem Register "Nachrichten") auf **Nachrichten speichern** klicken. Sie können auch alle Nachrichten für einen bestimmten Stream löschen, indem Sie in dieser Liste auf **Alle Nachrichten löschen** klicken.

Anzeigen der Knotenausführungszeiten

Auf der Registerkarte **Nachrichten** können Sie auswählen, dass die Registerkarte **Ausführungszeiten** angezeigt werden soll, auf der die einzelnen Ausführungszeiten für alle Knoten im Stream angezeigt werden, die in IBM SPSS Modeler Server ausgeführt werden. Die Ausführungszeiten sind für Streams, die in anderen Bereichen, beispielsweise R oder Analytic Server, ausgeführt werden, möglicherweise nicht genau.

Hinweis: Damit diese Funktion ausgeführt werden kann, muss auf der Registerkarte **Optionen** in der Einstellung **Allgemein** das Kontrollkästchen **Ausführungszeiten anzeigen** aktiviert sein.

In der Tabelle der Knotenausführungszeiten gibt es die unten angegebenen Spalten. Klicken Sie auf eine Spaltenüberschrift, um die Einträge in aufsteigender bzw. absteigender Reihenfolge zu sortieren (z. B. um anzuzeigen, welche Knoten die längsten Ausführungszeiten haben).

Endknoten. Die ID der Verzweigung, zu der der Knoten gehört. Die ID ist der Name des Endknotens am Ende der Verzweigung.

Knotenbeschriftung. Der Name des Knotens, auf den sich die Ausführungszeit bezieht.

Knoten-ID. Die eindeutige ID des Knotens, auf den sich die Ausführungszeit bezieht. Diese ID wird bei der Knotenerstellung vom System generiert.

Ausführungszeit(en). Die Zeitdauer der Ausführung dieses Knotens in Sekunden.

Festlegen der Stream- und Sitzungsparameter

Parameter können für die Verwendung in CLEM-Ausdrücken und Scripts definiert werden. Es handelt sich dabei im Grunde um benutzerdefinierte Variablen, die mit dem aktuellen Stream, der aktuellen Sitzung bzw. dem aktuellen Superknoten gespeichert und persistent gemacht werden. Ein Zugriff auf diese Variablen ist über die Benutzerschnittstelle und über Scripts möglich. Wenn Sie also einen Stream speichern, werden alle für diesen Stream festgelegten Parameter ebenfalls gespeichert. (Dies unterscheidet sie von lokalen Scriptvariablen, die nur in dem Script verwendet werden können, in dem sie deklariert sind.) Parameter werden häufig in Scripts als Teil eines CLEM-Ausdrucks verwendet, bei dem der Parameterwert im Script angegeben wird.

Der Bereich eines Parameters hängt davon ab, wo er festgelegt wird:

- Streamparameter können in einem Stream-Script oder im Dialogfeld "Streameigenschaften" festgelegt werden und sie stehen für alle Knoten im Stream zur Verfügung. Sie werden in der Parameterliste im Expression Builder angezeigt.
- Sitzungsparameter können in einem Standalone-Script oder im Dialogfeld "Sitzungsparameter" festgelegt werden. Sie stehen allen Streams zur Verfügung, die in der aktuellen Sitzung verwendet werden (alle Streams auf der Registerkarte "Streams" im Managerbereich).

Parameter können auch für Superknoten festgelegt werden. In diesem Fall sind sie nur für Knoten sichtbar, die in dem betreffenden Superknoten gekapselt sind.

So können Sie die Stream- und Sitzungsparameter über die Benutzerschnittstelle festlegen:

1. Zum Festlegen der Streamparameter klicken Sie im Hauptmenü auf:
Tools > Streameigenschaften > Parameter
2. Zum Festlegen der Sitzungsparameter klicken Sie im Menü "Extras" auf die Option **Sitzungsparameter festlegen**.

Eingabeaufforderung? Markieren Sie dieses Kontrollkästchen, wenn der Benutzer während der Laufzeit zur Eingabe eines Werts für diesen Parameter aufgefordert werden soll.

Name. Hier werden die Parameternamen aufgelistet. Sie können einen neuen Parameter erstellen, indem Sie in diesem Feld einen Namen eingeben. Um beispielsweise einen Parameter für die Mindesttemperatur zu erstellen, könnten Sie `minvalue` eingeben. Verwenden Sie nicht das Präfix `$P-`, das Parameter in CLEM-Ausdrücken kennzeichnet. Dieser Name wird auch zur Anzeige im CLEM Expression Builder verwendet.

Langer Name. Listet den beschreibenden Namen für die einzelnen erstellten Parameter auf.

Speichertyp. Wählen Sie einen Speichertyp aus der Liste aus. Der Speichertyp gibt an, wie die Datenwerte im Parameter gespeichert werden. Wenn Sie z. B. mit Werten arbeiten, die führende Nullen enthalten und die Sie beibehalten möchten (wie 008), sollten Sie **Zeichenfolge** als Speichertyp wählen. Andernfalls

werden die Nullen vom Wert abgezogen. Verfügbare Speichertypen sind "Zeichenfolge", "Ganze Zahl", "Reelle Zahl", "Uhrzeit", "Datum" und "Zeitmarke". Beachten Sie, dass bei Datumsparametern die Werte gemäß der im nächsten Absatz erläuterten ISO-Standardnotierung eingegeben werden müssen.

Wert. Listet den aktuellen Wert für die einzelnen Parameter auf. Ändern Sie den Parameter wie gewünscht. Datumsparameter müssen in ISO-Standardnotierung angegeben werden (d. h. in der Form JJJJ-MM-TT). Datumsangaben in anderen Formaten sind nicht zulässig.

Typ (optional). Wenn Sie den Stream für eine externe Anwendung bereitstellen möchten, wählen Sie aus der Liste ein Messniveau aus. Andernfalls sollten Sie die Spalte *Typ* so belassen, wie sie ist. Wenn Sie Wertbeschränkungen für den Parameter festlegen möchten, z. B. die Ober- und Untergrenze für einen numerischen Bereich, wählen Sie **Angeben** aus der Liste aus.

Die Optionen "Langer Name", "Speichertyp" und "Typ" können für Parameter nur über die Benutzerschnittstelle festgelegt werden. Die Festlegung dieser Optionen mithilfe von Scripts ist nicht möglich.

Klicken Sie auf die Pfeile rechts, um den ausgewählten Parameter in der Liste verfügbarer Parameter weiter nach oben oder weiter nach unten zu verschieben. Verwenden Sie die Schaltfläche zum Löschen (mit einem X markiert), um den ausgewählten Parameter zu entfernen.

Festlegen von Eingabeaufforderungen für Parameterwerte während der Laufzeit

Wenn Sie in manchen Streams zu verschiedenen Gelegenheiten für denselben Parameter unterschiedliche Werte eingeben müssen, können Sie Laufzeiteingabeaufforderungen für einen oder mehrere Stream- oder Sitzungsparameterwerte festlegen.

Parameter. (Optional) Geben Sie einen Wert für den Parameter ein oder belassen Sie den Standardwert, sofern vorhanden.

Diese Eingabeaufforderungen inaktivieren. Aktivieren Sie dieses Kontrollkästchen, wenn diese Eingabeaufforderungen beim Ausführen des Streams nicht angezeigt werden sollen. Sie können sie wieder anzeigen, indem Sie das Kontrollkästchen **Eingabeaufforderung?** im Dialogfeld für Stream- oder Sitzungseigenschaften, in dem die Parameter definiert wurden, aktivieren. Weitere Informationen finden Sie im Thema „Festlegen der Stream- und Sitzungsparameter“ auf Seite 51.

Festlegen von Wertbeschränkungen für einen Parametertyp

Sie können Wertbeschränkungen für einen Parameter während einer Streambereitstellung für eine externe Anwendung, die Datenmodellierungsstreams liest, zur Verfügung stellen. In diesem Dialogfeld können Sie die Werte angeben, die einem externen Benutzer, der den Stream ausführt, zur Verfügung stehen. Je nach Datentyp variieren die Wertbeschränkungen im Dialogfeld dynamisch. Die hier angezeigten Optionen stimmen mit den Optionen überein, die für Werte vom Typknoten zur Verfügung stehen.

Typ. Zeigt das aktuell ausgewählte Messniveau an. Sie können diesen Wert so ändern, wie der Parameter in IBM SPSS Modeler verwendet werden soll.

Speichertyp. Zeigt den Speichertyp an, sofern dieser bekannt ist. Die Speichertypen werden nicht von dem für die Arbeit in IBM SPSS Modeler ausgewählten Messniveau (stetig, nominal oder Flag) beeinflusst. Sie können den Speichertyp auf der Hauptregisterkarte "Parameter" ändern.

Die untere Hälfte des Dialogfelds ändert sich dynamisch in Abhängigkeit vom im Feld **Typ** ausgewählten Messniveau.

Stetige Messniveaus

Minimum. Geben Sie eine Untergrenze für die Parameterwerte ein.

Maximum. Geben Sie eine Obergrenze für die Parameterwerte ein.

Beschriftungen. Sie können Beschriftungen für jeden Wert eines Bereichsfelds angeben. Klicken Sie auf die Schaltfläche **Beschriftungen**, um ein gesondertes Dialogfeld zur Angabe der Wertbeschriftungen zu öffnen.

Nominale Messniveaus

Werte. Mit dieser Option können Sie Werte für einen Parameter angeben, der als nominales Feld verwendet wird. Die Werte werden im IBM SPSS Modeler-Stream nicht erzwungen, werden jedoch in einer Dropdown-Liste für externe Bereitstellungsanwendungen verwendet. Mit den Pfeilschaltflächen und der Löschschriftfläche können Sie bestehende Werte bearbeiten sowie Werte neu sortieren und löschen.

Flagmessniveaus

Wahr. Geben Sie einen Flagwert für den Parameter an, wenn die Bedingung erfüllt ist.

Falsch. Geben Sie einen Flagwert für den Parameter an, wenn die Bedingung nicht erfüllt ist.

Beschriftungen. Sie können Beschriftungen für die Werte eines Flagfelds angeben.

Bereitstellungsoptionen für den Stream

Auf der Registerkarte "Bereitstellung" im Dialogfeld für die Streameigenschaften können Sie Optionen zum Bereitstellen des Streams als Szenario in IBM SPSS Collaboration and Deployment Services zur Modellaktualisierung, für automatische Aufgabenplanung oder weitere Verwendung durch IBM Analytical Decision Management angeben. Alle Streams benötigen eine spezielle Scoring-Verzweigung, bevor sie bereitgestellt werden können; weitere Voraussetzungen und Optionen hängen vom Bereitstellungstyp ab. Weitere Informationen finden Sie im Thema „Speichern und Bereitstellen von Repository-Objekten“ auf Seite 126.

Ausführung unter Verwendung von Schleifen für Streams

Auf der Registerkarte "Ausführung" im Dialogfeld "Streameigenschaften" können Sie Bedingungen für die Verwendung von Schleifen festlegen, um sich wiederholende Aufgaben für den aktuellen Stream zu automatisieren.

Nachdem Sie diese Bedingungen festgelegt haben, können Sie diese Registerkarte als Einführung in das Scripting verwenden, da das Scriptfenster mit grundlegenden Scripting-Angaben für Ihren Stream ausgefüllt wird, die Sie dann ändern können, um sie vielleicht als Basis für die Erstellung besserer Scripts zu verwenden. Weitere Informationen finden Sie im Thema „Globale Funktionen“ auf Seite 120.

So legen Sie die Verwendung von Schleifen für einen Stream fest:

1. Klicken Sie im Menü "Datei" auf die Option **Streameigenschaften** (oder wählen Sie den Stream auf der Registerkarte "Streams" im Managerbereich aus, klicken Sie mit der rechten Maustaste und wählen Sie dann im Popup-Menü die Option **Streameigenschaften** aus).
2. Klicken Sie auf die Registerkarte **Ausführung**.
3. Wählen Sie den Ausführungsmodus **Verwendung von Schleifen/Bedingte Ausführung** aus.
4. Klicken Sie auf die Registerkarte **Verwendung von Schleifen**.

Klicken Sie als Alternative im Menü "Extras" auf:

Streameigenschaften > Ausführung

Als weitere Alternative klicken Sie mit der rechten Maustaste auf den Knoten und klicken Sie im Kontextmenü auf Folgendes:

Verwendung von Schleifen/Bedingte Ausführung > Einstellung für die Verwendung von Schleifen bearbeiten

Iteration. Sie können diesen Zeilennummernwert nicht bearbeiten, Sie können eine Iteration jedoch mit den Schaltflächen auf der rechten Seite der Tabelle hinzufügen, löschen oder nach oben oder unten verschieben.

Tabellenkopfzeilen. Diese geben den Iterationsschlüssel und alle Iterationsvariablen an, sofern vorhanden, die Sie beim Definieren der Schleife erstellt haben.

Anzeigen von globalen Werten für Streams

Mit der Registerkarte "Globalwerte" im Dialogfeld "Streameigenschaften" können Sie die globalen, für den aktuellen Stream festgelegten Werte anzeigen. Globalwerte werden mit einem Knoten vom Typ "Globalwerte" erstellt, um Statistiken zu ermitteln, wie Mittelwert, Summe oder Standardabweichung für ausgewählte Felder.

Sobald der Knoten "Globalwerte" ausgeführt wird, stehen diese Werte für eine Vielzahl von Anwendungen in Streamoperationen zur Verfügung. Weitere Informationen finden Sie im Thema „Globale Funktionen“ auf Seite 120.

So zeigen Sie Globalwerte für einen Stream an:

1. Klicken Sie im Menü "Datei" auf die Option **Streameigenschaften** (oder wählen Sie den Stream auf der Registerkarte "Streams" im Managerbereich aus, klicken Sie mit der rechten Maustaste und wählen Sie dann im Popup-Menü die Option **Streameigenschaften** aus).
2. Klicken Sie auf die Registerkarte **Globalwerte**.

Klicken Sie als Alternative im Menü "Extras" auf:

Streameigenschaften > Globalwerte

Verfügbare Globalwerte. Die verfügbaren Globalwerte werden in dieser Tabelle aufgelistet. Hier können Sie keine Globalwerte bearbeiten, aber alle Globalwerte für einen Stream unter Verwendung der Schaltfläche "Alle Werte löschen" rechts in der Tabelle löschen.

Suchen nach Knoten in einem Stream

Sie können Knoten in einem Stream suchen, indem Sie eine Reihe von Suchkriterien wie Knotenname, Kategorie und ID angeben. Diese Funktion ist besonders nützlich für komplexe Streams, die eine große Anzahl an Knoten enthalten.

So suchen Sie Knoten in einem Stream:

1. Klicken Sie im Menü "Datei" auf die Option **Streameigenschaften** (oder wählen Sie den Stream auf der Registerkarte "Streams" im Managerbereich aus, klicken Sie mit der rechten Maustaste und wählen Sie dann im Popup-Menü die Option **Streameigenschaften** aus).
2. Klicken Sie auf die Registerkarte **Suchen**.

Klicken Sie als Alternative im Menü "Extras" auf:

Streameigenschaften > Suchen

Sie können mehrere Optionen zur Einschränkung der Suche angeben. Einzig die Suche nach Knoten-ID (mithilfe des Felds **ID ist gleich**) schließt die anderen Optionen aus.

Knotenbeschriftung enthält. Markieren Sie dieses Kontrollkästchen und geben Sie eine Knotenbeschriftung ganz oder teilweise für die Suche nach einem bestimmten Knoten ein. Suchvorgänge unterscheiden Groß-/Kleinschreibung nicht und mehrere Wörter werden als ein einziges Textstück betrachtet.

Knotenkategorie. Aktivieren Sie dieses Kontrollkästchen und klicken Sie in der Liste auf eine Kategorie, um nach einem bestimmten Knotentyp zu suchen. **Prozessknoten** bedeutet einen Knoten aus der Registerkarte "Datensatzoperationen" oder "Feldoperationen" der Knotenpalette; **Modellknoten anwenden** bezieht sich auf ein Modellnugget.

Schlüsselwörter enthalten. Aktivieren Sie dieses Kontrollkästchen und geben Sie eines oder mehrere vollständige Schlüsselwörter ein, um Knoten zu suchen, für die diese Schlüsselwörter auf der Registerkarte "Anmerkungen" im Knotendialogfeld eingegeben wurden. Der eingegebene Schlüsselworttext muss eine exakte Übereinstimmung sein. Trennen Sie mehrere Schlüsselwörter durch Semikolons, um nach Alternativen zu suchen (z. B. werden mit `proton;neutron` alle Knoten mit einem dieser Schlüsselwörter gefunden). Weitere Informationen finden Sie im Thema „Anmerkungen“ auf Seite 63.

Anmerkung enthält. Aktivieren Sie dieses Kontrollkästchen und geben Sie eines oder mehrere Wörter ein, um Knoten zu suchen, die diesen Text im Haupttextbereich der Registerkarte "Anmerkungen" im Knotendialogfeld enthalten. Suchvorgänge unterscheiden Groß-/Kleinschreibung nicht und mehrere Wörter werden als ein einziges Textstück betrachtet. Weitere Informationen finden Sie im Thema „Anmerkungen“ auf Seite 63.

Feld generieren mit Name. Markieren Sie dieses Kontrollkästchen und geben Sie den Namen eines generierten Felds ein (zum Beispiel `$C-Drug`). Mithilfe dieser Option können Sie Modellierungsknoten suchen, die ein bestimmtes Feld generieren. Geben Sie nur einen Feldnamen ein, der eine exakte Übereinstimmung sein muss.

ID ist gleich. Markieren Sie dieses Kontrollkästchen und geben Sie eine Knoten-ID ein, um einen bestimmten Knoten mit dieser Kennung zu finden. (Durch Auswählen dieser Option werden alle vorherigen Optionen inaktiviert.) Knoten-IDs werden beim Erstellen eines Knotens vom System zugewiesen. Sie können als Verweis auf den Knoten zur Scripterstellung oder Automatisierung verwendet werden. Geben Sie nur eine Knoten-ID ein, die eine exakte Übereinstimmung sein muss. Weitere Informationen finden Sie im Thema „Anmerkungen“ auf Seite 63.

In Superknoten suchen. Dieses Kontrollkästchen ist standardmäßig markiert, d. h., die Suche erfolgt an Knoten innerhalb und außerhalb von Superknoten. Entfernen Sie die Markierung aus dem Kontrollkästchen, wenn Sie die Suche nur an Knoten außerhalb von Superknoten auf der oberen Streamebene ausführen möchten.

Suchen. Wenn Sie alle gewünschten Optionen festgelegt haben, klicken Sie auf diese Schaltfläche, um den Suchvorgang zu starten.

Knoten, die den angegebenen Optionen entsprechen, werden im unteren Bereich des Dialogfelds aufgelistet. Wählen Sie einen Knoten in der Liste aus, um ihn im Streamerstellungsbereich hervorzuheben.

Umbenennen von Streams

Auf der Registerkarte "Anmerkungen" im Dialogfeld "Streameigenschaften" können Sie beschreibende Anmerkungen für einen Stream hinzufügen und einen benutzerdefinierten Namen für den Stream erstellen. Diese Optionen sind besonders hilfreich, wenn Sie Berichte für Streams generieren, die dem Projektbereich hinzugefügt werden. Weitere Informationen finden Sie im Thema „Anmerkungen“ auf Seite 63.

Streambeschreibungen

Für jeden erstellten Stream generiert IBM SPSS Modeler eine Streambeschreibung, die Informationen zum Inhalt des Streams enthält. Diese Beschreibungen sind hilfreich, wenn Sie die Funktion eines Streams interessiert, Sie jedoch IBM SPSS Modeler nicht installiert haben (z. B. wenn Sie über IBM SPSS Collaboration and Deployment Services auf einen Stream zugreifen).

Die Streambeschreibung wird als HTML-Dokument mit verschiedenen Abschnitten angezeigt.

Allgemeine Streaminformationen

Dieser Abschnitt enthält den Streamnamen sowie Details über seine Erstellung und die letzte Speicherung.

Beschreibung und Kommentare

Dieser Abschnitt enthält Folgendes:

- Streamanmerkungen (siehe „Anmerkungen“ auf Seite 63)
- Nicht mit bestimmten Knoten verknüpfte Kommentare
- Kommentare, die sowohl in der Modellerstellungs- als auch der Scoring-Verzweigung des Streams mit Knoten verknüpft sind

Scoring-Informationen

Dieser Abschnitt enthält unter mehreren Überschriften Informationen zur Scoring-Verzweigung des Streams.

- **Kommentare.** Enthält nur Kommentare zu Knoten in der Scoring-Verzweigung.
- **Eingaben.** Listet die Eingabefelder zusammen mit ihren Speichertypen auf (z. B. "Zeichenfolge", "Ganze Zahl", "Reelle Zahl" usw.).
- **Ausgaben.** Listet die Ausgabefelder inklusive der zusätzlichen, vom Modellierungsknoten erstellten Felder zusammen mit ihren Speichertypen auf.
- **Parameter.** Listet alle Parameter auf, die sich auf die Scoring-Verzweigung des Streams beziehen und die bei jedem Modellscoring angezeigt oder bearbeitet werden können. Diese Parameter werden ermittelt, wenn Sie auf die Schaltfläche **Scoring-Parameter** auf der Registerkarte **Bereitstellung** des Dialogfelds für die Streameigenschaften klicken.
- **Modellknoten.** Zeigt den Namen und den Typ des Modells an (z. B. "Netz", "C&RT-Baum" usw.). Dieses Modellnugget ist für das Feld **Modellknoten** auf der Registerkarte **Bereitstellung** des Dialogfelds für die Streameigenschaften ausgewählt.
- **Modelldetails.** Zeigt Details zum unter der vorherigen Überschrift identifizierten Modellnugget an. Wenn möglich werden für das Modell auch die Bedeutsamkeit der Prädiktoren und Evaluierungsdiagramme angegeben.

Modellinformationen

Enthält Informationen zur Modellierungsverzweigung des Streams.

- **Kommentare.** Listet alle Kommentare oder Anmerkungen zu den Knoten in der Modellierungsverzweigung auf.
- **Eingaben.** Listet die Eingabefelder inklusive ihrer Rolle in der Modellierungsverzweigung auf (in Form des Werts für die Feldrolle, z. B. "Eingabe", "Ziel", "Aufteilung" usw.).
- **Parameter.** Listet alle Parameter auf, die sich auf die Modellierungsverzweigung des Streams beziehen und die bei jeder Aktualisierung des Modells angezeigt oder bearbeitet werden können. Diese Parameter werden ermittelt, wenn Sie auf die Schaltfläche **Modellerstellungsparameter** auf der Registerkarte **Bereitstellung** des Dialogfelds für die Streameigenschaften klicken.
- **Modellierungsknoten.** Zeigt Name und Typ des Modellknotens an, der zur Erstellung oder Aktualisierung des Modells verwendet wird.

Vorschau der Streambeschreibungen

Sie können den Inhalt einer Streambeschreibung in einem Web-Browser anzeigen, indem Sie im Dialogfeld für die Streameigenschaften auf eine Option klicken. Der Inhalt der Beschreibung hängt von den auf der Registerkarte "Bereitstellung" des Dialogfelds angegebenen Optionen ab. Weitere Informationen finden Sie im Thema „Bereitstellungsoptionen für den Stream“ auf Seite 139.

So zeigen Sie eine Streambeschreibung an:

1. Klicken Sie im IBM SPSS Modeler-Hauptmenü auf:

Tools > Streameigenschaften > Bereitstellung

2. Geben Sie Bereitstellungstyp, den zugehörigen Scoring-Knoten sowie eventuelle Scoring-Parameter an.
3. Für den Bereitstellungstyp "Modellaktualisierung" können Sie optional Folgendes auswählen:
 - Modellierungsknoten und eventuelle Modellerstellungsparameter
 - Modellnugget in der Scoring-Verzweigung des Streams
4. Klicken Sie auf die Schaltfläche **Vorschau der Streambeschreibung anzeigen**.

Exportieren von Streambeschreibungen

Sie können den Inhalt der Streambeschreibung in eine HTML-Datei exportieren.

So exportieren Sie eine Streambeschreibung:

1. Klicken Sie im Hauptmenü auf Folgendes:
Datei > Streambeschreibung exportieren
2. Geben Sie einen Namen für die HTML-Datei ein und klicken Sie auf **Speichern**.

Ausführen von Streams

Sobald Sie die gewünschten Optionen für Streams angegeben und die gewünschten Knoten verbunden haben, können Sie den Stream ausführen, indem Sie die Daten durch die Knoten im Stream laufen lassen. Es gibt verschiedene Möglichkeiten, einen Stream in IBM SPSS Modeler auszuführen. Sie verfügen über folgende Möglichkeiten:

- Klicken Sie im Menü "Extras" auf **Ausführen**.
- Klicken Sie auf eine der Schaltflächen **Ausführen...** in der Symbolleiste. Mit diesen Schaltflächen können Sie den gesamten Stream ausführen oder lediglich den ausgewählten Endknoten. Weitere Informationen finden Sie im Thema „IBM SPSS Modeler-Symbolleiste“ auf Seite 19.
- Um einen einzelnen Datenstream auszuführen, klicken Sie mit der rechten Maustaste auf einen Endknoten und dann im Popup-Menü auf **Ausführen**.
- Um einen Teil eines Datenstreams auszuführen, klicken Sie mit der rechten Maustaste auf einen Knoten, der kein Endknoten ist, und klicken Sie dann im Popup-Menü auf **Ab hier ausführen**. Dadurch werden nur die Operationen nach dem ausgewählten Knoten ausgeführt.

Um die Ausführung eines gerade laufenden Streams anzuhalten, können Sie auf die rote Schaltfläche "Stopp" in der Symbolleiste oder im Menü "Extras" auf **Ausführung anhalten** klicken.

Wenn ein Stream mehr als drei Sekunden für die Ausführung benötigt, wird das Dialogfeld "Ausführungs-Feedback" eingeblendet, in dem der Fortschritt angezeigt wird.

Bei einigen Knoten gibt es weitere Anzeigen, die zusätzliche Informationen zur Streamausführung bieten. Diese werden durch Auswahl der entsprechenden Zeile im Dialogfeld angezeigt. Die erste Zeile wird automatisch ausgewählt.

Arbeiten mit Modellen

Wenn ein Stream einen Modellierungsknoten enthält (d. h. einen Knoten aus der Registerkarte "Modellierung" oder "Datenbankmodellierung" der Knotenpalette), wird bei Ausführung des Streams ein **Modellnugget** erstellt. Ein Modellnugget ist ein Behälter für ein **Modell**, d. h. das Set von Regeln, Formeln und Gleichungen, das Ihnen Vorhersagen zu Ihren Quelldaten ermöglicht und das den Kern von Vorhersageanalysen ausmacht.



Abbildung 12. Modellnugget

Wenn Sie einen Modellierungsknoten erfolgreich ausführen, wird ein entsprechendes Modellnugget auf den Streamerstellungsbereich platziert, wo er durch ein goldfarbenes Rautensymbol repräsentiert wird (daher der Name "Nugget"). Sie können das Nugget öffnen und durch seinen Inhalt navigieren, um Details über das Modell zu sehen. Zur Anzeige der Vorhersagen verbinden Sie einen oder mehrere Modellknoten und führen diese aus. Die Ausgabe davon präsentiert die Vorhersagen in einer lesbaren Form.

Ein typischer Modellierungsstream besteht aus zwei Verzweigungen. Die **Modellierungsverzweigung** enthält den Modellierungsknoten zusammen mit der Quelle und vorangehenden Verarbeitungsknoten. Die **Modellierungsverzweigung** wird beim Ausführen des Modellierungsknotens erstellt und enthält das Modellnugget sowie den oder die Endknoten, die Sie zur Anzeige der Vorhersagen verwenden.

Weitere Informationen finden Sie im Handbuch *IBM SPSS Modeler Modellierungsknoten*.

Hinzufügen von Kommentaren und Anmerkungen zu Knoten und Streams

Sie müssen einen Stream eventuell für andere Personen in Ihrer Organisation beschreiben. Als Hilfe können Sie Streams, Knoten und Modellnuggets erklärende Kommentare hinzufügen.

Andere können diese Kommentare dann am Bildschirm ansehen oder Sie können eine Abbildung des Streams samt Kommentaren ausdrucken.

Sie können alle Kommentare für einen Stream oder Superknoten auflisten, die Reihenfolge der Kommentare in der Liste ändern, den Kommentartext bearbeiten und die Vorder- oder Hintergrundfarbe eines Kommentars ändern. Weitere Informationen finden Sie im Thema „Auflisten von Streamkommentaren“ auf Seite 62.

Sie können auch Knoten in der Form von Textanmerkungen zu Streams, Knoten und Nuggets mithilfe der Registerkarte "Anmerkungen" eines Dialogfelds mit Streameigenschaften, eines Knotendialogfelds oder eines Modellnuggetfensters hinzufügen. Diese Anmerkungen sind nur sichtbar, wenn die Registerkarte "Anmerkungen" geöffnet ist, aber die Streamanmerkungen lassen sich auch als Bildschirmkommentare anzeigen. Weitere Informationen finden Sie im Thema „Anmerkungen“ auf Seite 63.





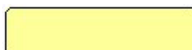

Kommentare

Kommentare können die Form von Textfeldern haben, in die Sie eine beliebige Textmenge eingeben können, und Sie können beliebig viele Kommentare hinzufügen. Ein Kommentar kann freistehend (mit keinerlei Streamobjekten verbunden) oder mit einem oder mehreren Knoten oder Modellnuggets im Stream verbunden sein. Freistehende Kommentare werden in der Regel verwendet, um den Gesamtzweck des Streams zu beschreiben; verbundene Kommentare beschreiben den Knoten oder das Nugget, mit dem sie verbunden sind. Mit Knoten und Nuggets können mehrere Kommentare verbunden sein, und der Stream kann eine beliebige Anzahl freistehender Kommentare aufweisen.

Hinweis: Sie können Streamanmerkungen auch als Bildschirmkommentare anzeigen, allerdings lassen sich diese nicht an Knoten oder Nuggets anhängen. Weitere Informationen finden Sie im Thema „Umwandeln von Anmerkungen in Kommentare“ auf Seite 62.

Das Aussehen des Textfelds ändert sich entsprechend dem aktuellen Modus des Kommentars (oder der als Kommentar angezeigten Anmerkung). Siehe dazu die folgende Tabelle.

Tabelle 3. Modi des Kommentar- und Anmerkungstextfelds

Kommentartextfeld	Anmerkungstextfeld	Modus	Bedeutung	Erzielt durch...
		Bearbeiten	Kommentar ist zur Bearbeitung geöffnet.	Erstellen eines neuen Kommentars bzw. einer neuen Anmerkung oder Doppelklick auf einen bestehenden Kommentar.
		Zuletzt ausgewählt	Kommentar lässt sich verschieben oder löschen oder seine Größe kann geändert werden.	Klick auf den Streamhintergrund nach der Bearbeitung oder Klick auf einen bestehenden Kommentar bzw. eine bestehende Anmerkung.
		Ansicht	Bearbeitung ist abgeschlossen.	Klick auf einen anderen Knoten, einen anderen Kommentar oder eine andere Anmerkung nach der Bearbeitung.

Wenn Sie einen neuen freistehenden Kommentar erstellen, wird dieser anfangs in der oberen linken Ecke des Streamerstellungsbereichs angezeigt.

Wenn Sie einen Kommentar mit einem Knoten oder ein Nugget verknüpfen, wird der Kommentar anfangs über dem Streamobjekt angezeigt, mit dem er verbunden ist.

Das Textfeld ist weiß, um zu zeigen, dass sich Text eingeben lässt. Wenn Sie den Text eingegeben haben, klicken Sie außerhalb des Textfelds. Der Kommentarhintergrund wird gelb, um zu zeigen, dass der Texteintrag abgeschlossen ist. Der Kommentar bleibt ausgewählt, damit Sie ihn verschieben, löschen oder seine Größe ändern können.

Bei erneutem Klicken ändert sich die Umrandung in durchgehende Linien und zeigt damit, dass die Bearbeitung abgeschlossen ist.

Durch Doppelklicken auf einen Kommentar wechselt das Textfeld in den Bearbeitungsmodus und der Hintergrund wird weiß, d. h. der Kommentartext kann bearbeitet werden.

Außerdem können Sie Kommentare mit Superknoten verknüpfen.

Operationen mit Kommentaren: Sie können eine Vielzahl von Operationen mit Kommentaren ausführen. Sie verfügen über folgende Möglichkeiten:

- Hinzufügen eines freistehenden Kommentars
- Verknüpfen eines Kommentars mit einem Knoten oder Nugget
- Bearbeiten eines Kommentars
- Größenänderung eines Kommentars
- Verschieben eines Kommentars
- Lösen der Verbindung eines Kommentars
- Löschen eines Kommentars
- Ein-/Ausblenden aller Kommentare für einen Stream

So fügen Sie einen freistehenden Kommentar hinzu:

1. Stellen Sie sicher, dass im Stream keine Auswahl getroffen ist.
2. Führen Sie eine der folgenden Aktionen aus:

- Klicken Sie im Hauptmenü auf Folgendes:
Einfügen > Neuer Kommentar
 - Klicken Sie mit der rechten Maustaste auf den Streamhintergrund und klicken Sie im Popup-Menü auf **Neuer Kommentar**.
 - Klicken Sie in der Symbolleiste auf die Schaltfläche **Neuer Kommentar**.
3. Geben Sie den Kommentartext ein (oder fügen Sie Text aus der Zwischenablage ein).
 4. Klicken Sie auf einen Knoten im Stream, um den Kommentar zu speichern.

So verknüpfen Sie einen Kommentar mit einem Knoten oder Nugget:

1. Wählen Sie einen oder mehrere Knoten oder Nuggets im Streamerstellungsbereich aus.
2. Führen Sie eine der folgenden Aktionen aus:
 - Klicken Sie im Hauptmenü auf Folgendes:
Einfügen > Neuer Kommentar
 - Klicken Sie mit der rechten Maustaste auf den Streamhintergrund und klicken Sie im Popup-Menü auf **Neuer Kommentar**.
 - Klicken Sie in der Symbolleiste auf die Schaltfläche **Neuer Kommentar**.
3. Geben Sie den Kommentartext ein.
4. Klicken Sie auf einen anderen Knoten im Stream, um den Kommentar zu speichern.
Sie haben außerdem folgende Möglichkeiten:
5. Einfügen eines freistehenden Kommentars (siehe vorherigen Abschnitt)
6. Führen Sie eine der folgenden Aktionen aus:
 - Wählen Sie den Kommentar aus, drücken Sie F2 und wählen Sie dann den Knoten oder das Nugget aus.
 - Wählen Sie den Knoten oder das Nugget aus, drücken Sie F2 und wählen Sie dann den Kommentar aus.
 - (Nur mit 3-Tasten-Maus) Bewegen Sie den Mauszeiger über den Kommentar, halten Sie die mittlere Taste gedrückt, ziehen Sie den Mauszeiger auf den Knoten oder das Nugget und lassen Sie die Maustaste los.

So verknüpfen Sie einen Kommentar mit einem zusätzlichen Knoten oder Nugget:

Wenn ein Kommentar bereits mit einem Knoten oder Nugget verknüpft ist oder sich auf Streamebene befindet und Sie diesen mit einem weiteren Knoten oder Nugget verknüpfen möchten, führen Sie eine der folgenden Aktionen aus:

- Wählen Sie den Kommentar aus, drücken Sie F2 und wählen Sie dann den Knoten oder das Nugget aus.
- Wählen Sie den Knoten oder das Nugget aus, drücken Sie F2 und wählen Sie dann den Kommentar aus.
- (Nur mit 3-Tasten-Maus) Bewegen Sie den Mauszeiger über den Kommentar, halten Sie die mittlere Taste gedrückt, ziehen Sie den Mauszeiger auf den Knoten oder das Nugget und lassen Sie die Maustaste los.

So bearbeiten Sie einen bestehenden Kommentar:

1. Führen Sie eine der folgenden Aktionen aus:
 - Doppelklicken Sie auf das Kommentartextfeld.
 - Wählen Sie das Textfeld aus und drücken Sie die Eingabetaste.
 - Klicken Sie mit der rechten Maustaste auf das Textfeld, um sein Menü zu öffnen, und klicken Sie auf "Bearbeiten".

2. Bearbeiten Sie den Kommentartext. Sie können die Windows-Standardtastenkombinationen zur Bearbeitung verwenden, z. B. Strg+C, um Text zu kopieren. Andere Optionen während der Bearbeitung werden im Popup-Menü für den Kommentar aufgelistet.
3. Klicken Sie einmal außerhalb des Textfelds, um die Größenänderungssteuerungen anzuzeigen, und noch einmal, um den Kommentar abzuschließen.

So ändern Sie die Größe eines Kommentartextfelds:

1. Wählen Sie den Kommentar aus, um die Größenänderungssteuerungen anzuzeigen.
2. Klicken Sie auf eine Steuerung und ziehen Sie sie, um die Größe des Felds zu ändern.
3. Klicken Sie außerhalb des Textfelds, um die Änderung zu speichern.

So verschieben Sie einen bestehenden Kommentar:

Wenn Sie einen Kommentar ohne seine verknüpften Objekte (falls vorhanden) verschieben möchten, führen Sie eine der folgenden Aktionen aus:

- Bewegen Sie den Mauszeiger über den Kommentar, halten Sie die linke Maustaste gedrückt und ziehen Sie den Kommentar an die neue Position.
- Wählen Sie den Kommentar aus, halten Sie die Taste Alt gedrückt und verschieben Sie den Kommentar mithilfe der Pfeiltasten.

Wenn Sie einen Kommentar zusammen mit etwaigen verknüpften Knoten oder Nuggets verschieben möchten:

1. Wählen Sie alle Objekte aus, die Sie verschieben möchten.
2. Führen Sie eine der folgenden Aktionen aus:
 - Verschieben Sie den Mauszeiger über eines der Objekte, halten Sie die linke Maustaste gedrückt und ziehen Sie die Objekte an die neue Position.
 - Wählen Sie eines der Objekte aus, halten Sie die Taste Alt gedrückt und verschieben Sie die Objekte mithilfe der Pfeiltasten.

So lösen Sie einen Kommentar von einem Knoten oder Nugget:

1. Wählen Sie einen oder mehrere Kommentare aus, die Sie lösen möchten.
2. Führen Sie eine der folgenden Aktionen aus:
 - Drücken Sie die Taste F3.
 - Klicken Sie mit der rechten Maustaste auf einen ausgewählten Kommentar und klicken Sie in dessen Menü auf "Trennen".

So löschen Sie einen Kommentar:

1. Wählen Sie einen oder mehrere Kommentare aus, die Sie löschen möchten.
2. Führen Sie eine der folgenden Aktionen aus:
 - Drücken Sie die Lösch Taste.
 - Klicken Sie mit der rechten Maustaste auf einen ausgewählten Kommentar und klicken Sie in dessen Menü auf "Löschen".

Wenn der Kommentar mit einem Knoten oder Nugget verknüpft war, wird die Verbindungslinie ebenfalls gelöscht.

Wenn der Kommentar ursprünglich eine Stream- oder Superknotenannotation war, die in einen freistehenden Kommentar umgewandelt wurde, wird er aus dem Erstellungsbereich gelöscht, sein Text bleibt jedoch auf der Registerkarte "Anmerkungen" für den Stream bzw. Superknoten erhalten.

So blenden Sie alle Kommentare für einen Stream ein oder aus:

1. Führen Sie eine der folgenden Aktionen aus:

- Klicken Sie im Hauptmenü auf Folgendes:

Ansicht > Kommentare

- Klicken Sie in der Symbolleiste auf die Schaltfläche **Kommentare ein-/ausblenden**.

Auflisten von Streamkommentaren: Sie können eine Liste aller Kommentare anzeigen, die für einen bestimmten Stream oder Superknoten erstellt wurden.

Diese Liste bietet die folgenden Möglichkeiten:

- Ändern der Reihenfolge der Kommentare
- Bearbeiten des Kommentartexts
- Ändern der Vorder- oder Hintergrundfarbe eines Kommentars

Auflisten von Kommentaren

Führen Sie zum Auflisten der Kommentare für einen Stream eine der folgenden Aktionen aus:

- Klicken Sie im Hauptmenü auf Folgendes:

Tools > Streameigenschaften > Kommentare

- Klicken Sie mit der rechten Maustaste auf einen Stream im Managerbereich und klicken Sie auf **Streameigenschaften** und dann auf **Kommentare**.
- Klicken Sie mit der rechten Maustaste in einen Streamhintergrund im Erstellungsbereich und klicken Sie auf **Streameigenschaften** und dann auf **Kommentare**.

Text. Der Text des Kommentars. Doppelklicken Sie auf den Text, um das Feld in ein editierbares Textfeld zu verwandeln.

Verknüpfungen. Der Name des Knotens, mit dem der Kommentar verbunden ist. Wenn dieses Feld leer ist, gilt der Kommentar für den Stream.

Positionierungsschaltflächen. Diese verschieben einen ausgewählten Kommentar in der Liste nach oben oder unten.

Kommentarfarben. Um die Vorder- oder Hintergrundfarbe eines Kommentars zu ändern, wählen Sie den Kommentar aus, markieren das Kontrollkästchen **Benutzerdefinierte Farben** und wählen dann eine Farbe aus der Liste **Hintergrund** und/oder **Vordergrund**. Klicken Sie auf **Anwenden** und dann in den Streamhintergrund, um die Wirkung der Änderung zu sehen. Klicken Sie auf **OK**, um die Änderung zu speichern.

Umwandeln von Anmerkungen in Kommentare: Anmerkungen zu Streams oder Superknoten können in Kommentare umgewandelt werden.

Bei Streams wird die Anmerkung im Streamerstellungsbereich in einen freistehenden Kommentar umgewandelt (d. h. ohne Verknüpfung zu Knoten).

Wenn eine Superknotenankündigung in einen Kommentar umgewandelt wird, wird dieser im Erstellungsbereich nicht mit dem Superknoten verknüpft, aber er ist sichtbar, sobald Sie den Superknoten vergrößern.

So wandeln Sie eine Streamankündigung in einen Kommentar um:

1. Klicken Sie im Menü "Extras" auf **Streameigenschaften**. (Als Alternative können Sie mit der rechten Maustaste auf einen Stream im Managerbereich und dann auf **Streameigenschaften** klicken.)
2. Klicken Sie auf die Registerkarte **Anmerkungen**.
3. Aktivieren Sie das Kontrollkästchen **Anmerkung als Kommentar anzeigen**.

4. Klicken Sie auf **OK**.

So wandeln Sie eine Superknotenannotation in einen Kommentar um:

1. Doppelklicken Sie auf das Symbol des Superknotens im Erstellungsbereich.
2. Klicken Sie auf die Registerkarte **Anmerkungen**.
3. Aktivieren Sie das Kontrollkästchen **Anmerkung als Kommentar anzeigen**.
4. Klicken Sie auf **OK**.

Anmerkungen

Knoten, Streams und Modelle können auf verschiedene Art und Weise mit Anmerkungen versehen werden. Sie können beschreibende Anmerkungen hinzufügen und einen benutzerdefinierten Namen angeben. Diese Optionen sind besonders hilfreich, wenn Sie Berichte für Streams generieren, die dem Projektbereich hinzugefügt werden. Für Knoten und Modellnuggets können Sie auch QuickInfo-Text hinzufügen, um zwischen ähnlichen Knoten im Streamerstellungsbereich zu unterscheiden.

Hinzufügen von Anmerkungen

Wenn Sie einen Knoten oder ein Modellnugget bearbeiten, wird ein Dialogfeld mit Registerkarten geöffnet, das eine Registerkarte "Anmerkungen" enthält, mit der Sie eine Vielzahl von Anmerkungen festlegen können. Sie können die Registerkarte "Anmerkungen" auch direkt öffnen.

1. Um Anmerkungen für einen Knoten oder ein Nugget zu erstellen, klicken Sie mit der rechten Maustaste auf den Knoten bzw. das Nugget im Streamerstellungsbereich und dann auf **Umbenennen und mit Anmerkung versehen**. Das Bearbeitungsdialogfeld wird so angezeigt, dass die Registerkarte "Anmerkungen" sichtbar ist.
2. Um einen Stream mit Anmerkungen zu versehen, klicken Sie im Menü "Extras" auf **Streameigenschaften**. (Als Alternative können Sie mit der rechten Maustaste auf einen Stream im Managerbereich und dann auf **Streameigenschaften** klicken.) Klicken Sie auf die Registerkarte "Anmerkungen".

Name. Wählen Sie **Benutzerdefiniert** aus, um den automatisch generierten Namen zu ändern oder um einen eindeutigen Namen für den Knoten zu erstellen, wie er im Streamerstellungsbereich angezeigt wird.

QuickInfo-Text. (Nur für Knoten und Modellnuggets) Geben Sie Text ein, der im Streamerstellungsbereich als QuickInfo verwendet wird. Dies ist besonders hilfreich, wenn Sie mit einer größeren Anzahl ähnlicher Knoten arbeiten.

Schlüsselwörter. Geben Sie Schlüsselwörter an, die in Projektberichten und bei der Suche nach Knoten in einem Stream oder beim Verfolgen von Objekten verwendet werden sollen, die im Repository gespeichert sind (siehe „Informationen zu IBM SPSS Collaboration and Deployment Services Repository“ auf Seite 125). Mehrere Schlüsselwörter können durch Semikolons voneinander getrennt werden, z. B. Einkommen; Feldfruchttyp; Subventionshöhe. Leerzeichen am Anfang und am Ende von Schlüsselwörtern werden ausgelassen; Einkommen; Feldfruchttyp führt z. B. zu demselben Ergebnis wie Einkommen;Feldfruchttyp. (Die Leerzeichen innerhalb von Schlüsselwörtern werden jedoch nicht ausgelassen. Zum Beispiel sind Höhe der Subventionen mit einem Leerzeichen vor "der" und Höhe der Subventionen mit zwei Leerzeichen vor "der" nicht dasselbe.)

Der Haupttextbereich kann verwendet werden, um lange Anmerkungen zu den Operationen des Knotens oder Entscheidungen im Knoten einzugeben. Wenn Sie z. B. Streams gemeinsam nutzen und erneut verwenden, ist es hilfreich, Anmerkungen zu Entscheidungen zu machen, z. B. Verwerfen eines Felds mit zahlreichen Leerzeichen unter Verwendung eines Filterknotens. Wenn Sie den Knoten mit Anmerkungen versehen, werden diese Informationen zusammen mit dem Knoten gespeichert. Sie können diese Anmerkungen auch in einem Projektbericht berücksichtigen, der im Projektbereich erstellt wurde. Weitere Informationen finden Sie im Thema „Einführung in Projekte“ auf Seite 149.

Anmerkung als Kommentar anzeigen. (Nur für Stream- und Superknotenanmerkungen) Markieren Sie dieses Kontrollkästchen, um die Anmerkung in einen freistehenden Kommentar umzuwandeln, der im Streamerstellungsbereich sichtbar ist. Weitere Informationen finden Sie im Thema „Hinzufügen von Kommentaren und Anmerkungen zu Knoten und Streams“ auf Seite 58.

ID. Zeigt eine eindeutige ID an, die zur Referenzierung des Knotens für Scripterstellung oder Automatisierung verwendet werden kann. Dieser Wert wird bei der Erstellung des Knotens automatisch generiert und ändert sich nicht. Beachten Sie außerdem, dass zur Vermeidung einer Verwechslung mit dem Buchstaben "O" in Knoten-IDs keine Nullen verwendet werden. Mit der Kopierschaltfläche auf der rechten Seite können Sie die ID kopieren und nach Bedarf in Scripts oder anderswo einfügen.

Speichern von Datenstreams

Nachdem Sie einen Stream erstellt haben, können Sie ihn für eine zukünftige Wiederverwendung speichern.

So speichern Sie einen Stream:

1. Klicken Sie im Menü "Datei" auf **Stream speichern** oder **Stream speichern unter**.
2. Blättern Sie im Dialogfeld "Speichern" zu dem Ordner, in dem die Streamdatei gespeichert werden soll.
3. Geben Sie im Textfeld "Dateiname" einen Namen für den Stream ein.
4. Wählen Sie **Zu Projekt hinzufügen**, wenn Sie den gespeicherten Stream zum aktuellen Projekt hinzufügen möchten.

Wenn Sie auf **Speichern** klicken, wird der Stream mit der Erweiterung **.str* im angegebenen Verzeichnis gespeichert.

Automatische Sicherungsdateien. Jedes Mal, wenn ein Stream gespeichert wird, wird die zuvor gespeicherte Version der Datei automatisch als Sicherung aufbewahrt und ein Bindestrich zum Dateinamen hinzugefügt (z. B. *meinStream.str-*). Um die gesicherte Version wiederherzustellen, löschen Sie einfach den Bindestrich und öffnen Sie die Datei erneut.

Speichern von Statusangaben

Zusätzlich zu Streams können Sie **Statusangaben** speichern, u. a. das aktuell angezeigte Streamdiagramm und alle Modellnuggets, die Sie erstellt haben (auf der Registerkarte "Modelle" im Managerbereich aufgelistet).

So speichern Sie einen Status:

1. Klicken Sie im Menü "Datei" auf:
Status > Status speichern oder **Status speichern unter**
2. Blättern Sie im Dialogfeld "Speichern" zu dem Ordner, in dem die Statusdatei gespeichert werden soll.

Wenn Sie auf **Speichern** klicken, wird der Status mit der Erweiterung **.cst* im angegebenen Verzeichnis gespeichert.

Speichern von Knoten

Sie können auch einen einzelnen Knoten speichern, indem Sie mit der rechten Maustaste auf den Knoten im Streamerstellungsbereich klicken und im Popup-Menü auf **Knoten speichern** klicken. Verwenden Sie die Dateinamenerweiterung **.nod*.

Speichern mehrerer Streamobjekte

Wenn Sie IBM SPSS Modeler verlassen, obwohl mehrere Objekte, wie Streams, Projekte oder Modellnuggets, nicht gespeichert sind, werden Sie zum Speichern aufgefordert, bevor Sie die Software vollständig beenden. Wenn Sie Elemente speichern möchten, wird ein Dialogfeld mit Optionen zum Speichern der einzelnen Objekte angezeigt.

1. Aktivieren Sie einfach die Kontrollkästchen für die Objekte, die Sie speichern möchten.
2. Klicken Sie auf **OK**, um die Objekte am gewünschten Speicherort zu speichern.

Es wird dann für jedes Objekt ein Standarddialogfeld zum Speichern aufgerufen. Wenn Sie alle Objekte gespeichert haben, wird die Anwendung wie ursprünglich angekündigt geschlossen.

Speichern der Ausgabe

Tabellen, Diagramme und Berichte, die aus IBM SPSS Modeler-Ausgabeknoten erstellt wurden, können nun im Ausgabeobjektformat (*.cou) gespeichert werden.

1. Um die zu speichernde Ausgabe anzuzeigen, klicken Sie in den Menüs des Ausgabefensters auf:
Datei > Speichern
2. Geben Sie einen Namen und einen Speicherort für die Ausgabedatei an.
3. Wählen Sie optional **Datei zu Projekt hinzufügen** im Dialogfeld "Speichern", um die Datei in das aktuelle Projekt aufzunehmen. Weitere Informationen finden Sie im Thema „Einführung in Projekte“ auf Seite 149.

Als Alternative können Sie mit der rechten Maustaste auf ein beliebiges Ausgabeobjekt im Managerbereich klicken und im Popup-Menü die Option **Speichern** auswählen.

Verschlüsseln und Entschlüsseln von Informationen

Beim Speichern von Streams, Knoten, Projekten, Ausgabedateien oder Modellnuggets können Sie das betreffende Element verschlüsseln, um es vor Verwendung durch Unbefugte zu schützen. Dazu wählen Sie beim Speichern eine zusätzliche Option aus und legen ein Kennwort für das zu speichernde Element fest. Diese Verschlüsselung kann für alle Objekte festgelegt werden, die Sie speichern, und bietet Ihnen zusätzliche Sicherheit; es ist nicht dasselbe wie die SSL-Verschlüsselung, die beim Austausch von Dateien zwischen IBM SPSS Modeler und IBM SPSS Modeler Server verwendet wird.

Wenn Sie versuchen, ein verschlüsseltes Element zu öffnen, werden Sie zur Eingabe des Kennworts aufgefordert. Nach Eingabe des richtigen Kennworts wird das Element automatisch entschlüsselt und wie gewohnt geöffnet.

So verschlüsseln Sie ein Objekt:

1. Klicken Sie im Dialogfeld "Speichern" für das zu verschlüsselnde Element auf **Optionen**. Das Dialogfeld "Verschlüsselungsoptionen" wird geöffnet.
2. Wählen Sie die Option **Diese Datei verschlüsseln** aus.
3. Für noch größere Sicherheit können Sie optional **Kennwort maskieren** auswählen. Dadurch werden alle Ihre Eingaben nur als eine Reihe von Punkten angezeigt.
4. Geben Sie das Kennwort ein. *Warnung:* Wenn Sie das Kennwort vergessen, kann die Datei bzw. das Modell nicht mehr geöffnet werden.
5. Wenn Sie **Kennwort maskieren** ausgewählt haben, müssen Sie das Kennwort erneut eingeben, um sicherzustellen, dass die erste Eingabe korrekt war.
6. Klicken Sie auf **OK**, um zum Dialogfeld "Speichern" zurückzukehren.

Hinweis: Wenn Sie eine Kopie eines durch Verschlüsselung geschützten Elements speichern, wird das neue Element automatisch in verschlüsselter Form unter Verwendung des ursprünglichen Kennworts gespeichert, es sei denn, Sie ändern die Einstellungen im Dialogfeld "Verschlüsselungsoptionen".

Laden von Dateien

Sie können eine Vielzahl gespeicherter Objekte in IBM SPSS Modeler erneut laden:

- Streams (.str)
- Statusangaben (.cst)

- Modelle (.gm)
- Modellpalette (.gen)
- Knoten (.nod)
- Ausgabe (.cou)
- Projekte (.cpj)

Öffnen neuer Dateien

Streams können direkt im Menü "Datei" geladen werden.

1. Klicken Sie im Menü "Datei" auf **Stream öffnen**.

Alle anderen Dateitypen können mithilfe der Untermenüoptionen im Menü "Datei" geöffnet werden. Um z. B. ein Modell zu laden, klicken Sie im Menü "Datei" auf:

Modelle > Modell öffnen oder Modellpalette laden

Öffnen der zuletzt verwendeten Dateien

Um die zuletzt verwendeten Dateien schnell zu laden, können Sie die Optionen unten im Menü "Datei" verwenden.

Wählen Sie **Zuletzt verwendete Streams**, **Zuletzt verwendete Projekte** oder **Zuletzt verwendete Statusangaben**, um eine Liste mit den zuletzt verwendeten Dateien zu öffnen.

Zuordnen von Datenstreams

Mit dem Zuordnungstool können Sie eine neue Datenquelle mit einem bereits vorhandenen Stream verbinden. Mit dem Zuordnungstool können Sie nicht nur die Verbindung einrichten, sondern auch festlegen, wie Felder in der neuen Quelle die im bereits vorhandenen Stream ersetzen. Anstatt einen kompletten Datenstream für eine neue Datenquelle erneut zu erstellen, können Sie einfach eine Verbindung mit einem bereits vorhandenen Stream herstellen.

Mit dem Datenzuordnungstool können Sie zwei Streamfragmente zusammenfügen und sicherstellen, dass alle (wichtigen) Feldnamen zugeordnet werden. Das Zuordnen von Daten führt hauptsächlich zur Erstellung eines neuen Filterknotens, der die entsprechenden Felder zuordnet, indem er sie umbenennt.

Es gibt zwei gleichwertige Möglichkeiten zum Zuordnen von Daten:

Auswählen des Ersatzknotens. Diese Methode beginnt mit dem zu ersetzenden Knoten. Klicken Sie zuerst mit der rechten Maustaste auf den zu ersetzenden Knoten. Wählen Sie dann über die Option **Datenzuordnung > Ersatzknoten auswählen** im Popup-Menü den Knoten aus, der den ersten Knoten ersetzen soll.

Zuordnen zu. Diese Methode beginnt mit dem Knoten, der im Stream eingefügt werden soll. Klicken Sie zuerst mit der rechten Maustaste auf den einzufügenden Knoten. Wählen Sie dann über die Option **Datenzuordnung > Zuordnen zu** im Popup-Menü den Knoten aus, dem der erste Knoten hinzugefügt werden soll. Diese Methode ist besonders geeignet für die Zuordnung zu einem Endknoten. *Hinweis:* Für Zusammenführungs- oder Anhangknoten sind keine Zuordnungen möglich. Stattdessen sollte der Stream auf die normale Art und Weise mit dem Zusammenführungsknoten verbunden werden.

Datenzuordnung ist eng in die Streamerstellung integriert. Wenn Sie versuchen, einen Knoten zu verbinden, der bereits über eine Verbindung verfügt, erhalten Sie die Option die Verbindung oder Zuordnung dieses Knotens zu ersetzen.

Zuordnen von Daten zu einer Vorlage

Um die Datenquelle für einen Vorlagenstream durch einen neuen Quellenknoten zu ersetzen, der Ihre eigenen Daten in IBM SPSS Modeler einbringt, sollten Sie aus dem Popup-Menü "Datenzuordnung" die Option **Ersatzknoten auswählen** verwenden. Diese Option steht für alle Knoten mit Ausnahme von Merge-, Aggregat- und allen Endknoten zur Verfügung. Wenn Sie das Datenzuordnungstool zum Durchführen dieser Aktion verwenden, wird sichergestellt, dass die Felder zwischen den bereits vorhandenen Streamoperationen und der neuen Datenquelle ordnungsgemäß zugeordnet werden. Die folgenden Schritte bieten einen Überblick über den Datenzuordnungsprozess.

Schritt 1: Geben Sie die erforderlichen Felder im ursprünglichen Quellenknoten an. Damit die Streamoperationen ordnungsgemäß ausgeführt werden können, müssen die erforderlichen Felder angegeben werden. Weitere Informationen finden Sie im Thema „Festlegen von wichtigen Feldern“ auf Seite 68.

Schritt 2: Fügen Sie dem Streamerstellungsbereich eine neue Datenquelle hinzu. Bringen Sie mithilfe eines der Quellenknoten die neuen Ersatzdaten ein.

Schritt 3: Ersetzen Sie den Vorlagenquellenknoten. Wählen Sie im Popup-Menü des Vorlagenquellenknotens die Option "Datenzuordnung" aus, klicken Sie auf **Ersatzknoten auswählen** und wählen Sie dann den Quellenknoten für die Ersatzdaten aus.

Schritt 4: Überprüfen Sie die zugeordneten Felder. Prüfen Sie in dem Dialogfeld, das nun geöffnet wird, ob die Software die Felder aus der Ersatzdatenquelle dem Stream ordnungsgemäß zuordnet. Sämtliche nicht zugeordneten, aber erforderlichen Felder werden rot dargestellt. Diese Felder werden in Streamoperationen verwendet und müssen durch ein ähnliches Feld in der neuen Datenquelle ersetzt werden, damit die weiter unten im Stream liegenden Operationen ordnungsgemäß ausgeführt werden können. Weitere Informationen finden Sie im Thema „Prüfen von zugeordneten Feldern“ auf Seite 68.

Nachdem Sie das Dialogfeld verwendet haben, um sicherzustellen, dass alle erforderlichen Felder ordnungsgemäß zugeordnet sind, wird die alte Datenquelle getrennt und die neue Datenquelle mithilfe des Filterknotens namens *Zuordnen* mit dem Stream verbunden. Dieser Filterknoten leitet die tatsächliche Zuordnung der Felder im Stream. Ein Filterknoten namens *Zuordnung aufheben* ist ebenfalls im Streamerstellungsbereich vorhanden. Der Filterknoten *Zuordnung aufheben* kann dazu verwendet werden, die Feldnamenzuordnung umzukehren, indem er dem Stream hinzugefügt wird. Die Zuordnung der bereits zugeordneten Felder wird dadurch aufgehoben. Beachten Sie jedoch, dass sämtliche weiter unten liegenden Endknoten für eine erneute Auswahl der Felder und Überlagerungen bearbeitet werden müssen.

Zuordnen zwischen Streams

Ähnlich wie beim Verbinden von Knoten müssen bei dieser Datenzuordnungsmethode keine erforderlichen Felder vorab festgelegt werden. Mit dieser Methode verbinden Sie einen Stream mit einem anderen, indem Sie im Popup-Menü "Datenzuordnung" die Option **Zuordnen zu** auswählen. Diese Art der Datenzuordnung ist für das Zuordnen zu Endknoten und zum Kopieren und Einfügen zwischen Streams sinnvoll. *Hinweis:* Mit der Option **Zuordnen zu** können Sie keine Zuordnung zu Zusammenführungs-, Anhang- und anderen Quellenknoten herstellen.

So ordnen Sie Daten zwischen Streams zu:

1. Klicken Sie mit der rechten Maustaste auf den Knoten, den Sie zum Verbinden mit dem neuen Stream verwenden möchten.
2. Klicken Sie im Menü auf:
Datenzuordnung > Zuordnen zu
3. Verwenden Sie den Cursor, um einen Zielknoten im Ziel-Stream auszuwählen.
4. Vergewissern Sie sich, dass in dem Dialogfeld, das jetzt angezeigt wird, die Felder ordnungsgemäß zugeordnet sind. Klicken Sie dann auf **OK**.

Festlegen von wichtigen Feldern

Bei der Zuordnung zu einem bestehenden Stream werden wichtige Felder in der Regel vom Streamautor festgelegt. Diese wichtigen Felder zeigen an, ob ein bestimmtes Feld in weiter unten im Stream liegenden Operationen verwendet wird. Der bereits vorhandene Stream erstellt u. U. ein Modell, das ein Feld namens *Verlust* verwendet. In diesem Stream ist *Verlust* ein wichtiges Feld, da eine Modellerstellung ohne dieses Feld nicht möglich ist. Auf ähnliche Weise sind Felder, die für Bearbeitungsknoten, z. B. für einen Ableitungsknoten, verwendet werden, zum Ableiten des neuen Felds erforderlich. Wenn Sie solche Felder ausdrücklich als wichtig markieren, können Sie sicherstellen, dass ihnen im neuen Quellenknoten die richtigen Felder zugeordnet werden. Wenn keine wichtigen Felder zugeordnet sind, erhalten Sie eine Fehlermeldung. Wenn Sie feststellen, dass einige Bearbeitungs- oder Ausgabeknoten nicht erforderlich sind, können Sie die Knoten aus dem Stream löschen und die entsprechenden Felder aus der Liste der wichtigen Felder entfernen.

So legen Sie die wichtigen Felder fest:

1. Klicken Sie mit der rechten Maustaste auf den Quellenknoten des Vorlagenstreams, der ersetzt werden soll.
2. Klicken Sie im Menü auf:
Datenzuordnung > Wichtige Felder festlegen
3. Mit der Feldauswahlschaltfläche können Sie der Liste Felder hinzufügen oder entfernen. Um die Feldauswahl zu aktivieren, klicken Sie auf das Symbol rechts neben der Feldliste.

Prüfen von zugeordneten Feldern

Sobald Sie den Punkt ausgewählt haben, an dem ein Datenstream bzw. eine Datenquelle einem anderen zugeordnet wird, wird ein Dialogfeld angezeigt, in dem Sie Felder für die Zuordnung auswählen oder sicherstellen können, dass die Systemstandardzuordnung richtig ist. Wenn für den Stream oder die Datenquelle wichtige Felder festgelegt und nicht zugeordnet wurden, werden diese Felder rot dargestellt. Sämtliche nicht zugeordneten Felder aus der Datenquelle durchlaufen den Filterknoten unverändert. Beachten Sie jedoch, dass Sie auch Felder zuordnen können, die nicht "wichtig" sind.

Original. Listet alle Felder in der Vorlage oder im bereits vorhandenen Stream auf - alle Felder, die weiter unten im Stream vorhanden sind. Diesen Feldern werden Felder aus der neuen Datenquelle zugeordnet.

Zugeordnet. Listet die Felder auf, die für die Zuordnung zu Vorlagenfeldern ausgewählt wurden. Dabei handelt es sich um die Felder, deren Namen eventuell geändert werden müssen, um den in Streamoperationen verwendeten ursprünglichen Feldern zu entsprechen. Klicken Sie in die Tabellenzelle für ein Feld, um eine Liste mit verfügbaren Feldern zu aktivieren.

Wenn Sie sich nicht sicher sind, welche Felder zugeordnet werden sollen, sollten Sie die Quelldaten vor der Zuordnung detailliert untersuchen. Sie können z. B. die Registerkarte "Typen" im Quellenknoten verwenden, um eine Zusammenfassung der Quelldaten zu prüfen.

Tipps und Tastenkombinationen

Arbeiten Sie schnell und problemlos, indem Sie sich mit den folgenden Tastenkombinationen und Tipps vertraut machen:

- **Schnelles Erstellen von Streams durch Doppelklicken.** Doppelklicken Sie einfach auf einen Knoten in der Palette, um ihn zum aktuellen Stream hinzuzufügen und zu verbinden.
- **Verwenden von Tastenkombinationen zur Auswahl von nachgeordneten Knoten.** Drücken Sie Strg+Q und Strg+W, um alle weiter unten liegenden Knoten aus- bzw. abzuwählen.
- **Verwenden von Tastenkombinationen zum Verbinden und Trennen von Knoten.** Wenn ein Knoten im Erstellungsbereich ausgewählt ist, drücken Sie F2, um eine Verbindung zu beginnen, drücken Sie die Tabulatortaste, um zum gewünschten Knoten zu gelangen, und Umschalt+Leertaste, um die Verbindung fertig zu stellen. Drücken Sie F3, um alle Ein- und Ausgaben vom ausgewählten Knoten zu trennen.

- **Anpassen der Registerkarte "Knotenpalette" mit Ihren Lieblingsknoten.** Klicken Sie im Menü "Extras" auf **Paletten verwalten**, um ein Dialogfeld zu öffnen, in dem Sie die in der Knotenpalette angezeigten Knoten hinzufügen, entfernen bzw. verschieben können.
- **Umbenennen von Knoten und Hinzufügen von QuickInfo.** Jedes Knotendialogfeld enthält eine Registerkarte "Anmerkungen", auf der Sie einen benutzerdefinierten Namen für Knoten im Erstellungsbereich angeben können sowie QuickInfo hinzufügen können, wodurch Ihr Stream einfacher organisiert wird. Sie können auch lange Anmerkungen zur Fortschrittsverfolgung einfügen, Prozessdetails speichern und sämtliche Geschäftsentscheidungen notieren, die erforderlich sind oder umgesetzt wurden.
- **Automatisches Einfügen von Werten in einen CLEM-Ausdruck.** Mit dem Expression Builder, den Sie in einer Vielzahl von Dialogfeldern aufrufen können (z. B. Dialogfelder für Ableitungs- und Filterknoten), können Sie automatisch Feldwerte in einen CLEM-Ausdruck einfügen. Klicken Sie auf die Werteschaltfläche im Expression Builder, um eine Auswahl aus den bereits vorhandenen Feldwerten zu treffen.



Abbildung 13. Werteschaltfläche

- **Schnelles Suchen nach Dateien.** Verwenden Sie beim Suchen von Dateien in einem Dialogfeld "Öffnen" die Liste "Datei" (klicken Sie auf das gelbe Rautensymbol oben im Dialogfeld neben dem Feld "Suchen in"), um zuvor verwendete Verzeichnisse sowie IBM SPSS Modeler-Standardverzeichnisse aufzurufen. Verwenden Sie die Vor- und Zurückschaltflächen, um durch die Verzeichnisse zu blättern.
- **Minimieren überflüssiger Elemente im Ausgabefenster.** Sie können die Ausgabe schnell schließen und entfernen, indem Sie die Schaltfläche mit dem roten X verwenden, die Sie in allen Ausgabefenstern in der rechten oberen Ecke finden. So können Sie nur die vielversprechenden und interessanten Ergebnisse auf der Registerkarte "Ausgaben" des Managerbereichs beibehalten.

Für die Software steht eine vollständige Palette mit Tastenkombinationen zur Verfügung. Weitere Informationen finden Sie im Thema „Tastatureingabehilfen“ auf Seite 174.

Wussten Sie, dass Sie ...

- ... mithilfe der Maus eine Gruppe von Knoten in den Streamerstellungsbereich ziehen und auswählen können?
- ... Knoten von einem Stream zu einem anderen kopieren und einfügen können?
- ... in jedem Dialogfeld und Ausgabefenster die Hilfe aufrufen können?
- ... Hilfe zu CRISP-DM (Cross-Industry Standard Process for Data Mining) aufrufen können? (Klicken Sie dazu im Hilfemenü auf **CRISP-DM-Hilfe**.)

Kapitel 6. Umgang mit fehlenden Werten

Fehlende Werte - Übersicht

In der Datenvorbereitungsphase des Data-Mining kann es häufig vorkommen, dass Sie fehlende Werte in den Daten ersetzen möchten. **Fehlende Werte** sind Werte im Dataset, die unbekannt, nicht erfasst oder falsch eingegeben sind. In der Regel sind solche Werte für die entsprechenden Felder ungültig. Beispielsweise sollte das Feld *Geschlecht* die Werte *M* und *W* enthalten. Wenn Sie die Werte *Y* oder *Z* in diesem Feld vorfinden, können Sie mit Sicherheit davon ausgehen, dass sie ungültig sind und daher als Leerzeichen interpretiert werden sollten. Ebenso ist ein negativer Wert für das Feld *Alter* sinnlos und sollte als Leerzeichen interpretiert werden. Häufig werden in einem Fragebogen derartige offenkundig falschen Werte absichtlich eingegeben oder es werden Felder leer gelassen, um eine Nichtbeantwortung zu kennzeichnen. Mitunter ist es vielleicht erforderlich, diese Leerzeichen genauer zu untersuchen, um festzustellen, ob eine Nichtbeantwortung, wie eine verweigerter Altersangabe, einen Faktor für die Voraussage eines bestimmten Ergebnisses darstellt.

Manche Modellierungsverfahren gehen besser mit fehlenden Daten um als andere. C5.0 und Apriori kommen beispielsweise gut mit Werten zurecht, die in einem Typknoten ausdrücklich als "fehlend" definiert sind. Bei anderen Modellierungsverfahren treten Probleme bei der Bearbeitung fehlender Werte und eine längere Trainingszeit auf, was zu ungenaueren Modellen führt.

In Clementine gibt es mehrere Arten von fehlenden Werten, die von IBM SPSS Modeler erkannt werden:

- **Nullwerte oder systemdefiniert fehlende Werte.** Bei diesen Werten handelt es sich um Nicht-Zeichenfolgenwerte, die in der Datenbank bzw. der Quellendatei leer gelassen und nicht speziell in einem Quellen- oder Typknoten als "fehlend" definiert wurden. Systemdefiniert fehlende Werte werden als **\$null\$** angezeigt. Beachten Sie, dass leere Zeichenfolgen in IBM SPSS Modeler nicht als Nullen betrachtet werden, auch wenn sie von bestimmten Datenbanken (siehe unten) als Nullen behandelt werden können.
- **Leere Zeichenfolgen und leere Bereiche.** Leere Zeichenfolgewerte und leere Bereiche (Zeichenfolgen ohne sichtbare Zeichen) werden anders als Nullwerte behandelt. Leere Zeichenfolgen werden in den meisten Fällen als äquivalent mit leeren Bereichen (Leerzeichen) behandelt. Beispiel: Wenn Sie die Option auswählen, dass leere Bereiche in einem Quellen- oder Typknoten als Leerstellen behandelt werden sollen, gilt diese Einstellung auch für leere Zeichenfolgen.
- **Leere oder benutzerdefiniert fehlende Werte.** Es handelt sich hierbei um Werte wie unbekannt, 99 oder -1, die in einem Quellen- oder Typknoten ausdrücklich als fehlend definiert sind. Optional können Sie auch auswählen dass Nullen und leere Bereiche als Leerzeichen behandelt werden sollen. Dadurch können sie mit Flags für eine spezielle Behandlung versehen und aus den meisten Berechnungen ausgeschlossen werden. Beispielsweise können Sie die Funktion @BLANK verwenden, um diese Werte gemeinsam mit anderen Arten von fehlenden Werten als Leerstellen zu behandeln.

Einlesen gemischter Daten. Beachten Sie, dass beim Einlesen von Feldern mit numerischem Speichertyp (ganze Zahl, reelle Zahl, Zeit, Zeitmarke oder Datum) alle nicht numerischen Werte auf *null* oder auf *systemdefiniert fehlend* gesetzt werden. Dies liegt daran, dass im Gegensatz zu einigen anderen Anwendungen hier keine gemischten Speichertypen innerhalb eines Felds zulässig sind. Um dies zu vermeiden, sollten alle Felder mit gemischten Daten als Zeichenfolgen eingelesen werden, indem der Speichertyp im Quellenknoten bzw. in der externen Anwendung nach Bedarf geändert wird.

Lesen leerer Zeichenfolgen aus Oracle. Beim Lesen aus oder Schreiben in Oracle-Datenbanken sollten Sie darauf achten, dass Oracle im Gegensatz zu IBM SPSS Modeler und den meisten anderen Datenbanken leere Zeichenfolgewerte wie Nullwerte behandelt und speichert. Dies bedeutet, dass dieselben Daten sich unterschiedlich verhalten können und unterschiedliche Ergebnisse ausgeben können, je nachdem ob sie aus einer Oracle-Datenbank oder aus einer anderen Datenbank bzw. einer Datei extrahiert wurden.

Umgang mit fehlenden Werten

Entscheiden Sie vor dem Hintergrund Ihres Geschäfts- oder Domänenwissens, wie fehlende Werte behandelt werden sollen. Um die Trainingszeit zu verkürzen und die Genauigkeit zu erhöhen, können Sie beispielsweise einerseits Leerzeichen aus Ihrem Dataset entfernen. Andererseits kann das Vorhandensein leerer Werte neue Geschäftschancen eröffnen oder für zusätzliche Einblicke sorgen. Bei der Auswahl des besten Verfahrens sollten Sie folgende Aspekte Ihrer Daten berücksichtigen:

- Größe des Datensets
- Anzahl der Felder mit Leerzeichen
- Menge der fehlenden Daten

Im Allgemeinen gibt es zwei Ansätze, die Sie verfolgen können:

- Sie können Felder bzw. Datensätze mit fehlenden Werten ausschließen.
- Sie können fehlende Werte mit einer Reihe von Methoden imputieren, ersetzen oder erzwingen.

Beide Ansätze lassen sich mithilfe des Data Audit-Knotens weitgehend automatisieren. Beispielsweise können Sie einen Filterknoten generieren, der Felder ausschließt, die so viele fehlende Werte aufweisen, dass sie bei der Modellierung nicht mehr sinnvoll einsetzbar sind, und einen Superknoten generieren, der fehlende Werte für bestimmte oder alle verbleibenden Felder imputiert. Hier zeigt sich die Leistungsfähigkeit des Audits: Sie können nicht nur den aktuellen Status Ihrer Daten bewerten, sondern auch aufgrund dieser Bewertung Maßnahmen ergreifen.

Behandeln von Datensätzen mit fehlenden Werten

Wenn sich die Mehrzahl der fehlenden Werte in einer kleinen Anzahl von Datensätzen konzentriert, können Sie diese Datensätze einfach ausschließen. Eine Bank pflegt beispielsweise normalerweise detaillierte und vollständige Datensätze über ihre Darlehenskunden. Wenn die Bank bei der Gewährung von Darlehen an ihre Mitarbeiter weniger restriktiv ist, weisen die für Mitarbeiterdarlehen erhobenen Daten wahrscheinlich einige leere Felder auf. In diesem Fall stehen zwei Optionen zur Behandlung dieser fehlenden Werte zur Verfügung.

- Sie können die Personaldatensätze mithilfe eines Auswahlknotens entfernen.
- Bei einem großen Dataset können Sie alle Datensätze mit Leerstellen verwerfen.

Behandeln von Feldern mit fehlenden Werten

Wenn ein Großteil der fehlenden Werte in einer kleinen Anzahl von Feldern konzentriert ist, können Sie sie anstatt auf der Datensatzebene auf der Feldebene angehen. Mit diesem Ansatz können Sie außerdem mit der relativen Wichtigkeit bestimmter Felder experimentieren, bevor Sie sich für eine Methode zur Handhabung fehlender Werte entscheiden. Wenn ein Feld für die Modellierung unwichtig ist, lohnt es sich wahrscheinlich nicht, es beizubehalten, unabhängig davon, wie viele fehlende Werte es aufweist.

Ein Marktforschungsunternehmen erhebt beispielsweise Daten anhand eines allgemeinen Fragebogens mit 50 Fragen. Zwei Fragen betreffen das Alter und die politische Überzeugung. Diese Informationen werden von vielen Menschen ungern angegeben. In diesem Fall weisen die Felder *Alter* und *Politische_Überzeugung* viele fehlende Werte auf.

Feldmessniveau

Bei der Bestimmung der zu verwendenden Methode sollten Sie auch das Messniveau von Feldern mit fehlenden Werten in die Überlegungen einfließen lassen.

Numerische Felder. Für numerische Feldtypen wie *Stetig* sollten Sie immer nicht numerische Werte eliminieren, bevor Sie ein Modell erstellen, da viele Modelle nicht funktionieren, wenn Leerzeichen in numerischen Feldern enthalten sind.

Kategoriale Felder. Für kategoriale Felder, wie *Nominal* und *Flag*, ist das Ändern fehlender Werte zwar nicht erforderlich, steigert jedoch die Genauigkeit des Modells. Ein Modell mit dem Feld *Geschlecht* funktioniert zwar auch mit sinnlosen Werten, wie *Y* und *Z*, durch Entfernen aller Werte außer *M* und *W* wird jedoch die Genauigkeit des Modells erhöht.

Sichten bzw. Entfernen von Feldern

Zum Ausschluss von Feldern mit zu vielen fehlenden Werten stehen mehrere Optionen zur Verfügung:

- Mit dem Data Audit-Knoten können Sie Felder nach Qualität filtern.
- Mithilfe eines Merkmalauswahlknotens können Sie Felder ausschließen, die mehr als einen angegebenen Prozentsatz fehlender Werte enthalten und eine Rangordnung der Felder auf der Grundlage der Wichtigkeit in Bezug auf ein angegebenes Ziel erstellen.
- Statt die Felder zu entfernen, können Sie die Feldrolle anhand eines Typknotens auf **Keine** festlegen. Dadurch werden die Felder im Dataset belassen, aber aus den Modellierungsprozessen ausgeschlossen.

Imputieren oder Auffüllen fehlender Werte

Wenn nur wenige fehlende Werte vorliegen, kann es sinnvoll sein, die Leerstellen durch eingefügte Werte zu ersetzen. Dies kann über den Data Audit-Bericht erfolgen, mit dem Sie nach Bedarf Optionen für bestimmte Felder angeben und dann einen Superknoten erstellen können, der Werte anhand einer Reihe von Methoden imputiert. Dies ist die flexibelste Methode, die außerdem die Bearbeitung einer großen Anzahl an Feldern in einem einzelnen Knoten gestattet.

Folgende Methoden stehen zum Imputieren fehlender Werte zur Verfügung:

Fest. Ersetzt einen festen Wert (Feldmittelwert, Mittelpunkt des Bereichs oder eine von Ihnen angegebene Konstante).

Zufällig. Ersetzt einen Zufallswert auf der Grundlage einer Normal- oder Gleichverteilung.

Ausdruck. Ermöglicht die Angabe eines benutzerdefinierten Ausdrucks. Beispielsweise könnten Sie Werte durch eine globale Variable ersetzen, die vom Globalwerteknoten erstellt wurde.

Algorithmus. Ersetzt einen von einem Modell vorhergesagten Wert auf der Grundlage eines C&RT-Algorithmus. Für jedes Feld, das mit dieser Methode imputiert wurde, gibt es ein separates C&RT-Modell sowie einen Füllerknoten, der Leerstellen und Nullen durch den vom Modell vorhergesagten Wert ersetzt. Anschließend werden die vom Modell generierten Vorhersagefelder mithilfe eines Filterknotens entfernt.

Um alternativ Werte für bestimmte Felder zu erzwingen, können Sie mithilfe eines Typknotens sicherstellen, dass die Feldtypen nur zulässige Werte abdecken, und dann die Spalte *Überprüfen* für die Felder, deren leere Werte ersetzt werden müssen, auf **Erzwingen** setzen.

CLEM-Funktionen für fehlende Werte

Zur Behandlung fehlender Werte stehen mehrere Funktionen zur Verfügung. Die folgenden Funktionen werden häufig in Auswahl- und Füllerknoten zum Verwerfen oder Auffüllen fehlender Werte verwendet.

- `count_nulls(LISTE)`
- `@BLANK(FELD)`
- `@NULL(FELD)`
- `undef`

Die @-Funktionen können in Verbindung mit der Funktion @FIELD verwendet werden, um vorhandene Leerzeichen oder Nullwerte in einem oder mehreren Feldern zu ermitteln. Die Felder können bei vorhandenen Leerzeichen oder Nullwerten einfach markiert werden oder sie können mit Ersatzwerten ausgefüllt oder in zahlreichen verschiedenen Operationen verwendet werden.

Sie können Nullen folgendermaßen in einer Liste von Feldern zählen:

```
count_nulls(['cardtenure' 'card2tenure' 'card3tenure'])
```

Wenn Sie eine der Funktionen verwenden, die eine Liste von Feldern als Eingabe akzeptieren, können die Sonderfunktionen @FIELDS_BETWEEN und @FIELDS_MATCHING verwendet werden, wie im folgenden Beispiel gezeigt:

```
count_nulls(@FIELDS_MATCHING('card*'))
```

Mit der Funktion undef können Sie Felder mit dem systemdefiniert fehlenden Wert, der als \$null\$ angezeigt wird, ausfüllen. Um beispielsweise einen beliebigen numerischen Wert zu ersetzen, können Sie eine bedingte Anweisung verwenden. Beispiel:

```
if not(Alter > 17) or not(Alter < 66) then undef else Alter endif
```

Mit dieser Anweisung werden alle Werte ersetzt, die nicht mit einem systemdefiniert fehlenden Wert, der als \$null\$ angezeigt wird, in einem Bereich liegen. Mit der Funktion not() können Sie alle anderen numerischen Werte, einschließlich negativer Werte, erfassen. Weitere Informationen finden Sie im Thema „Funktionen zum Umgang mit Leerstellen und Nullwerten“ auf Seite 121.

Hinweis zum Verwerfen von Datensätzen

Beachten Sie beim Verwenden eines Auswahlknotens zum Verwerfen von Datensätzen, dass die Syntax eine dreiwertige Logik verwendet und automatisch Nullwerte in SELECT-Anweisungen integriert. Um Nullwerte (systemdefiniert fehlende Werte) aus einem SELECT-Ausdruck auszuschließen, müssen Sie dies ausdrücklich angeben, indem Sie and not im Ausdruck verwenden. Um beispielsweise alle Datensätze, deren Medikamenttyp Medikament C ist, auszuwählen und einzuschließen, wird folgende SELECT-Anweisung verwendet:

```
Drug = 'drugC' and not(@NULL(Drug))
```

In früheren Versionen wurden Nullwerte in solchen Fällen ausgeschlossen.

Kapitel 7. Erstellen von CLEM-Ausdrücken

Informationen zu CLEM

CLEM (Control Language for Expression Manipulation) ist eine leistungsfähige Sprache zur Analyse und Bearbeitung von Daten, die durch IBM SPSS Modeler-Streams fließen. Data-Mining-Experten verwenden CLEM in hohem Maße für Streamoperationen, um so einfache Aufgaben durchzuführen, wie das Ableiten von Profit aus Kosten- und Einkommensdaten, oder um so komplexe Aufgaben durchzuführen, wie das Umwandeln von Webprotokolldaten in ein Set von Feldern und Datensätzen mit nützlichen Informationen.

CLEM wird innerhalb von IBM SPSS Modeler für folgende Zwecke eingesetzt:

- Bedingungen in Datensatzfeldern vergleichen und evaluieren
- Werte für neue Felder ableiten
- Neue Werte für vorhandene Felder ableiten
- Reihenfolge der Datensätze erarbeiten
- Daten aus Datensätzen in Berichte einfügen

Scripting. Ein Subset der CLEM-Sprache kann außerdem in Scripts in der Benutzerschnittstelle verwendet werden. Dadurch können Sie viele gleichartige Datenbearbeitungen automatisieren.

CLEM-Ausdrücke sind für die Datenvorbereitung in IBM SPSS Modeler unverzichtbar und können in einer Vielzahl von Knoten verwendet werden - von Datensatz- und Feldoperationen (Auswahl, Balancierung, Füller) bis hin zu Plots und Ausgaben (Analyse, Bericht, Tabelle). Sie können beispielsweise einen CLEM-Ausdruck in einem Ableitungsknoten verwenden, um ein neues Feld anhand einer Formel, wie das Verhältnis, zu erstellen.

CLEM-Ausdrücke können auch für globales Suchen und Ersetzen verwendet werden. Der Ausdruck @NULL(@FIELD) beispielsweise kann in einem Füllerknoten verwendet werden, um **systemdefiniert fehlende Werte** durch den ganzzahligen Wert 0 zu ersetzen. (Zum Ersetzen von **benutzerdefiniert fehlenden Werten**, auch als "fehlende Werte" bezeichnet, verwenden Sie die Funktion @BLANK.)

Es können auch komplexere CLEM-Ausdrücke erstellt werden. Beispielsweise können Sie neue Felder auf der Grundlage eines bedingten Regelsets ableiten, etwa eine neue Wertekategorie, die mithilfe folgender Ausdrücke erstellt wird: `If: CardID = @OFFSET(CardID,1), Then: @OFFSET(ValueCategory,1), Else: 'exclude'`.

In diesem Beispiel wird die Funktion @OFFSET für folgende Anweisung verwendet: Wenn der Wert des Felds *CardID* für einen bestimmten Datensatz mit dem für den vorherigen Datensatz identisch ist, soll der Wert des Felds *ValueCategory* für den vorherigen Datensatz zurückgegeben werden. Andernfalls soll die Zeichenfolge "ausschließen" zugewiesen werden. Mit anderen Worten: Wenn die Werte für *CardID* für angrenzende Datensätze identisch sind, sollte ihnen dieselbe Wertekategorie zugewiesen werden. (Datensätze mit der Zeichenfolge "ausschließen" können später mithilfe eines Auswahlknotens entfernt werden.)

CLEM - Beispiele

Zur Veranschaulichung der korrekten Syntax sowie der mit CLEM möglichen Ausdrücke werden im Folgenden Beispielausdrücke genannt.

Einfache Ausdrücke

Formeln können sehr einfach sein. Die folgende Formel leitet beispielsweise ein neues Feld basierend auf den Werten der Felder *Nach* und *Vor* ab:

```
(Nach - Vor) / Vor * 100.0
```

Beachten Sie, dass Feldnamen nicht in Anführungszeichen gesetzt werden, wenn sie sich auf die Feldwerte beziehen.

Ähnlich gibt der folgende Ausdruck einfach den Logarithmus der einzelnen Werte des Felds *Gehalt* zurück.

```
log(Gehalt)
```

Komplexe Ausdrücke

Ausdrücke können auch relativ lang und komplex sein. Der folgende Ausdruck gibt den Wert *true* zurück, wenn der Wert zweier Felder (*\$KX-Kohonen* und *\$KY-Kohonen*) innerhalb der angegebenen Bereiche liegt. Beachten Sie, dass die Feldnamen in diesem Fall in einfache Anführungszeichen gesetzt werden, da die Feldnamen Sonderzeichen enthalten.

```
(' $KX-Kohonen' >= -0.2635771036148072 and ' $KX-Kohonen' <= 0.3146203637123107  
and ' $KY-Kohonen' >= -0.18975617885589602 and  
' $KY-Kohonen' <= 0.17674794197082522) -> T
```

Einige Funktionen, wie Zeichenfolgenfunktionen, erfordern die Eingabe mehrerer Parameter in der korrekten Syntax. Im folgenden Beispiel wird die Funktion *subscr* verwendet, um das erste Zeichen eines Felds *produce_ID* zurückzugeben. Dabei wird angegeben, ob ein Element organisch, genetisch verändert oder herkömmlich ist. Die Ergebnisse eines Ausdrucks werden durch *-> 'Ergebnis'* beschrieben.

```
subscr(1,produce_ID) -> 'c'
```

Ebenso lautet der folgende Ausdruck:

```
stripchar('3','123') -> '12'
```

Beachten Sie unbedingt, dass Zeichen immer in einfache umgekehrte Anführungszeichen gesetzt werden müssen.

Kombinieren von Funktionen in einem Ausdruck

CLEM-Ausdrücke bestehen häufig aus einer Kombination von Funktionen. Die folgende Funktion kombiniert *subscr* und *lowertoupper*, um das erste Zeichen von *produce_ID* zurückzugeben und es in einen Großbuchstaben zu konvertieren.

```
lowertoupper(subscr(1,produce_ID)) -> 'C'
```

Der gleiche Ausdruck kann in Kurzform wie folgt geschrieben werden:

```
lowertoupper(produce_ID(1)) -> 'C'
```

Eine andere häufig verwendete Kombination von Funktionen lautet:

```
locchar_back('n', (length(web_page)), web_page)
```

Mit diesem Ausdruck wird das Zeichen 'n' in den Werten des Felds *web_page* gesucht, wobei vom letzten Zeichen des Feldwerts an rückwärts gelesen wird. Durch Integrieren der Funktion *length* berechnet der Ausdruck dynamisch die Länge des aktuellen Werts, anstatt eine statische Zahl wie 7 zu verwenden. Diese Zahl ist ungültig für Werte, die weniger als sieben Zeichen enthalten.

Sonderfunktionen

Es sind zahlreiche Sonderfunktionen (die mit dem Symbol @ beginnen) verfügbar. Häufig verwendete Funktionen sind u. a.:

```
@BLANK('referrer ID') -> T
```

Häufig werden Sonderfunktionen kombiniert verwendet. Diese Methode wird im Allgemeinen dazu verwendet, Leerstellen in mehr als einem Feld gleichzeitig zu kennzeichnen.

```
@BLANK(@FIELD)-> T
```

Weitere Beispiele werden in der CLEM-Dokumentation erörtert. Weitere Informationen finden Sie im Thema „CLEM-Referenz - Überblick“ auf Seite 89.

Werte und Datentypen

CLEM-Ausdrücke ähneln Formeln, die aus Werten, Feldnamen, Operatoren und Funktionen gebildet werden. Der einfachste gültige CLEM-Ausdruck besteht aus einem Wert oder einem Feldnamen. Gültige Werte sind zum Beispiel:

```
3
1.79
'banana'
```

Feldnamen sind zum Beispiel:

```
Product_ID
'$P-NextField'
```

Dabei ist *Product* der Name eines Feldes aus einem Warenkorbdataset, '*\$P-NextField*' der Name eines Parameters und der Wert des Ausdrucks ist der Wert des benannten Feldes. In der Regel beginnen Feldnamen mit einem Buchstaben und können auch Ziffern und Unterstriche (_) enthalten. Wenn Sie den Namen in Anführungszeichen setzen, können Sie Namen verwenden, die diese Regeln nicht einhalten.

CLEM-Werte können folgende sein:

- Zeichenfolgen - z. B. "c1", "Typ 2", "freier Text"
- Ganze Zahlen - z. B. 12, 0, -189
- Reelle Zahlen - z. B. 12,34, 0,0, -0,0045
- Datums-/Zeitfelder - z. B. 05/12/2002, 12/05/2002, 12/05/02

Es können auch folgende Elemente verwendet werden:

- Zeichencodes - z. B. `a` oder 3
- Elementlisten - z. B. [1 2 3], ['Typ 1' 'Typ 2']

Zeichencodes und -listen erscheinen normalerweise nicht als Feldwerte. In der Regel werden sie als Argumente für CLEM-Funktionen verwendet.

Regeln für Anführungszeichen

Die Software ist zwar flexibel in Bezug auf die Bestimmung der in einem CLEM-Ausdruck verwendeten Felder, Werte, Parameter und Zeichenfolgen. Die folgenden allgemeinen Regeln dienen jedoch als Liste der "besten Verfahrensweisen" für das Erstellen von Ausdrücken.

- **Zeichenfolgen** - Setzen Sie Zeichenfolgen immer in doppelte Anführungszeichen ("Typ 2" oder "Wert"). Bei einfachen Anführungszeichen besteht die Gefahr, dass sie mit in Anführungszeichen gesetzten Feldern verwechselt werden.
- **Zeichen** - Verwenden Sie immer einfache umgekehrte Anführungszeichen, nämlich ` . Beachten Sie beispielsweise das Zeichen d in der Funktion stripchar(`d`, "drugA"). Die einzige Ausnahme zu dieser Regel ist die Verwendung einer Ganzzahl für den Verweis auf ein bestimmtes Zeichen in einer Zeichenfolge. Beachten Sie beispielsweise das Zeichen 5 in der Funktion lowertoupper("drugA"(5)) -> "A".

Hinweis: Auf einer englischen und US-amerikanischen Standardtastatur befindet sich die Taste für das umgekehrte Anführungszeichen (Accent grave, Unicode 0060) unterhalb der Escapetaste, auf europäischen Standardtastaturen unterhalb der Taste F10.

- **Felder** - Felder werden in der Regel in CLEM-Ausdrücken nicht in Anführungszeichen gesetzt (`subscr(2,arrayID)` → CHAR). Wenn Leerzeichen oder andere Sonderzeichen eingeschlossen werden müssen, verwenden Sie einfache Anführungszeichen ('Auftragsnummer'). Felder, die in Anführungszeichen gesetzt sind, im Dataset jedoch nicht definiert sind, werden irrtümlich als Zeichenfolgen interpretiert.
- **Parameter** - Verwenden Sie immer einfache Anführungszeichen ('\$P-threshold').

Ausdrücke und Bedingungen

CLEM-Ausdrücke können ein Ergebnis zurückgeben (beim Ableiten neuer Werte). Beispiel:

```
Weight * 2.2
Age + 1
sqrt(Signal-Echo)
```

Oder sie können *wahr* oder *falsch* überprüfen (beim Auswählen einer Bedingung). Beispiel:

```
Drug = "drugA"
Age < 16
not(PowerFlux) and Power > 2000
```

In CLEM-Ausdrücken können Operatoren und Funktionen willkürlich miteinander kombiniert werden. Beispiel:

```
sqrt(abs(Signal)) * max(T1, T2) + Baseline
```

Die Reihenfolge, in der der Ausdruck berechnet wird, wird durch Klammern und Operatorvorrang bestimmt. Im genannten Beispiel lautet die Berechnungsreihenfolge wie folgt:

- `abs(Signal)` wird berechnet und `sqrt` wird auf das Ergebnis angewendet.
- `max(T1, T2)` wird berechnet.
- Die beiden Ergebnisse werden multipliziert: `x` hat Vorrang vor `+`.
- Schließlich wird `Baseline` zum Ergebnis addiert.

Die absteigende Vorrangfolge (d. h. Rechenoperationen, die zuerst ausgeführt werden, bis hin zu Operationen, die zuletzt ausgeführt werden) lautet wie folgt:

- Funktionsargumente
- Funktionsaufrufe
- `xx`
- `x / mod div rem`
- `+ -`
- `> < >= <= /== == = /=`

Wenn Sie den Vorrang umgehen möchten oder in Bezug auf die Berechnungsreihenfolge unsicher sind, können Sie zur Verdeutlichung Klammern verwenden. Beispiel:

```
sqrt(abs(Signal)) * (max(T1, T2) + Baseline)
```

Stream-, Sitzungs- und Superknotenparameter

Parameter können für die Verwendung in CLEM-Ausdrücken und Scripts definiert werden. Es handelt sich dabei im Grunde um benutzerdefinierte Variablen, die mit dem aktuellen Stream, der aktuellen Sitzung bzw. dem aktuellen Superknoten gespeichert und persistent gemacht werden. Ein Zugriff auf diese Variablen ist über die Benutzerschnittstelle und über Scripts möglich. Wenn Sie also einen Stream spei-

chern, werden alle für diesen Stream festgelegten Parameter ebenfalls gespeichert. (Dies unterscheidet sie von lokalen Scriptvariablen, die nur in dem Script verwendet werden können, in dem sie deklariert sind.) Parameter werden häufig in Scripts als Teil eines CLEM-Ausdrucks verwendet, bei dem der Parameterwert im Script angegeben wird.

Der Bereich eines Parameters hängt davon ab, wo er festgelegt wird:

- Streamparameter können in einem Stream-Script oder im Dialogfeld "Streameigenschaften" festgelegt werden und sie stehen für alle Knoten im Stream zur Verfügung. Sie werden in der Parameterliste im Expression Builder angezeigt.
- Sitzungsparameter können in einem Standalone-Script oder im Dialogfeld "Sitzungsparameter" festgelegt werden. Sie stehen allen Streams zur Verfügung, die in der aktuellen Sitzung verwendet werden (alle Streams auf der Registerkarte "Streams" im Managerbereich).

Parameter können auch für Superknoten festgelegt werden. In diesem Fall sind sie nur für Knoten sichtbar, die in dem betreffenden Superknoten gekapselt sind.

Verwenden von Parametern in CLEM-Ausdrücken

Parameter werden in CLEM-Ausdrücken durch \$P-pname dargestellt. Dabei ist pname der Name des Parameters. Wenn Parameter in CLEM-Ausdrücken verwendet werden, müssen sie in einfache Anführungszeichen eingeschlossen werden. Beispiel: '\$P-scale'.

Die verfügbaren Parameter können einfach mit dem Expression Builder angezeigt werden. So zeigen Sie die aktuellen Parameter an:

1. Klicken Sie in einem beliebigen Dialogfeld, in das CLEM-Ausdrücke eingegeben werden können, auf die Schaltfläche des Expression Builder.
2. Wählen Sie in der Liste "Felder" die Option **Parameter**.

Sie können in der Liste Parameter auswählen, um sie in den CLEM-Ausdruck einzufügen. Weitere Informationen finden Sie im Thema „Auswählen von Feldern, Parametern und globalen Variablen“ auf Seite 84.

Arbeiten mit Zeichenfolgen

Für Zeichenfolgen stehen eine Reihe von Operationen zur Verfügung, einschließlich:

- Konvertieren einer Zeichenfolge in Groß- oder Kleinbuchstaben - `uppertoLower(ZEICHEN)`.
- Entfernen angegebener Zeichen, z. B. 'ID_' oder '\$', aus einer Zeichenfolgevariablen - `stripchar(ZEICHEN, ZEICHENFOLGE)`.
- Bestimmen der Länge (Zeichenanzahl) einer Zeichenfolgevariablen - `length(ZEICHENFOLGE)`.
- Überprüfen der alphabetischen Sortierung von Zeichenfolgewerten - `alphabefore(ZEICHENFOLGE1, ZEICHENFOLGE2)`.
- Entfernen führender oder folgender Leerzeichen aus Werten - `trim(ZEICHENFOLGE)`, `trim_start(ZEICHENFOLGE)` oder `trim_end(ZEICHENFOLGE)`.
- Entfernen der ersten bzw. letzten *n* Zeichen aus einer Zeichenfolge - `startstring(LÄNGE, ZEICHENFOLGE)` oder `endstring(LÄNGE, ZEICHENFOLGE)`. Angenommen, es liegt beispielsweise ein Feld mit der Bezeichnung *Element* vor, in dem ein Produktname und ein 4-stelliger ID-Code kombiniert sind (ACME CAMERA-D109). Geben Sie die folgende Formel in einem Ableitungsknoten an, um ein neues Feld zu erstellen, das nur den 4-stelligen Code enthält:
`endstring(4, Element)`
- Abgleichen eines bestimmten Musters - `ZEICHENFOLGE matches MUSTER`. Um beispielsweise Personen mit dem Wort "market" an einer beliebigen Stelle in ihrer Stellenbezeichnung auszuwählen, können Sie Folgendes in einem Auswahlknoten angeben:
`job_title matches "*market*"`

- Ersetzen aller Instanzen einer Teilzeichenfolge innerhalb einer Zeichenfolge - `replace(TEILZEICHENFOLGE, NEUETEILZEICHENFOLGE, ZEICHENFOLGE)`. Um vor dem Textmining zum Beispiel alle Instanzen eines nicht unterstützten Zeichens (etwa einen senkrechten Strich [|]) durch ein Semikolon zu ersetzen, verwenden Sie in einem Füllerknoten die Funktion `replace`. Wählen Sie unter **Felder ausfüllen** alle Felder aus, in denen das Zeichen möglicherweise vorkommt. Wählen Sie für die Bedingung **Ersetzen** die Option **Immer** aus und geben Sie unter **Ersetzen durch** folgende Bedingung an:

```
replace('|',';',@FIELD)
```

- Ableiten eines Flagfelds basierend auf Vorhandensein einer bestimmten Teilzeichenfolge. Sie können beispielsweise eine Zeichenfolgenfunktion in einem Ableitungsknoten verwenden, um ein gesondertes Flagfeld für jede Antwort mit einem Ausdruck der folgenden Art zu erstellen:

```
hassubstring(museums,"museum_of_design")
```

Weitere Informationen finden Sie im Thema „Zeichenfolgefunktionen“ auf Seite 105.

Umgang mit Leerzeichen und fehlenden Werten

Zu den gängigen Aufgaben, die Data-Mining-Experten zur Datenvorbereitung ausführen, gehört das Ersetzen von Leerzeichen oder fehlenden Werten. CLEM bietet eine Reihe von Tools zur Automatisierung der Handhabung von Leerzeichen. Mit Leerzeichen wird vor allem im Füllerknoten gearbeitet. Die folgenden Funktionen können jedoch mit jedem Knoten verwendet werden, der CLEM-Ausdrücke akzeptiert:

- Mit `@BLANK(FIELD)` können Datensätze bestimmt werden, deren Werte für ein bestimmtes Feld leer sind, z. B. *Alter*.
- Mit `@NULL(FIELD)` können Datensätze bestimmt werden, deren Werte für das bzw. die angegebenen Felder systemdefiniert fehlen. In IBM SPSS Modeler werden systemdefiniert fehlende Werte als **\$null\$**-Werte angezeigt.

Weitere Informationen finden Sie im Thema „Funktionen zum Umgang mit Leerstellen und Nullwerten“ auf Seite 121.

Arbeiten mit Zahlen

In IBM SPSS Modeler stehen zahlreiche Standardoperationen für numerische Werte zur Verfügung. Dazu gehören:

- Berechnen des Sinus des angegebenen Winkels - `sin(NUM)`
- Berechnen des natürlichen Logarithmus numerischer Felder - `log(NUM)`
- Berechnen der Summe zweier Zahlen - `NUM1 + NUM2`

Weitere Informationen finden Sie im Thema „Numerische Funktionen“ auf Seite 101.

Arbeiten mit Zeit- und Datumsangaben

Die Zeit- und Datumsformate können je nach Datenquelle und Ländereinstellungen variieren. Die Formate für Datum und Uhrzeit sind für jeden Stream spezifisch und werden im Dialogfeld "Streameigenschaften" festgelegt. Im Folgenden sind Beispiele für häufig verwendete Funktionen für die Arbeit mit Datums- und Zeitfeldern aufgeführt.

Berechnen der verstrichenen Zeit

Mit einer Gruppe von Funktionen wie der folgenden können Sie mühelos die Zeit berechnen, die seit einem Basisdatum verstrichen ist. Diese Funktion gibt die Zeit in Monaten und als reelle Zahl zurück, die vom Basisdatum bis zu dem Datum verstrichen ist, das durch die Datumszeichenfolge `DATUM` repräsentiert wird. Dabei handelt es sich um einen Schätzwert, der auf einem Monat mit 30,0 Tagen basiert.

`date_in_months(Datum)`

Vergleichen von Datums-/Uhrzeitwerten

Die Werte in Datums- und Uhrzeitfeldern können in den Datensätzen anhand von Funktionen wie der folgenden verglichen werden. Diese Funktion gibt den Wert *Wahr* zurück, wenn die Datumszeichenfolge DATUM1 ein Datum darstellt, das vor dem durch die Datumszeichenfolge DATUM2 repräsentierten Datum liegt. Andernfalls ergibt die Funktion den Wert 0.

`date_before(Datum1, Datum2)`

Berechnen von Differenzen

Sie können auch die Differenz zwischen zwei Uhrzeiten und zwei Datumsangaben berechnen. Verwenden Sie dazu Funktionen wie:

`date_weeks_difference(Datum1, Datum2)`

Diese Funktion gibt die Zeit in Wochen ab dem durch die Datumszeichenfolge DATUM1 repräsentierten Datum bis zu dem durch die Datumszeichenfolge DATUM2 angegebenen Datum als reelle Zahl zurück. Der Wert basiert auf einer Woche mit 7,0 Tagen. Wenn DATUM2 vor DATUM1 liegt, wird eine negative Zahl zurückgegeben.

Heutiges Datum

Das aktuelle Datum kann mit der Funktion @TODAY zum Dataset hinzugefügt werden. Das heutige Datum wird als Zeichenfolge zum angegebenen oder neuen Feld im Datumsformat hinzugefügt, das im Dialogfeld "Streameigenschaften" ausgewählt wurde. Weitere Informationen finden Sie im Thema „Datums- und Uhrzeitfunktionen“ auf Seite 110.

Zusammenfassen mehrerer Felder

Die CLEM-Sprache beinhaltet eine Reihe von Funktionen, mit denen Übersichtsstatistiken ausgegeben werden, die mehrere Felder umfassen. Diese Funktionen sind besonders nützlich bei der Analyse von Umfragedaten, bei denen mehrere Antworten auf eine Frage in mehreren Feldern gespeichert werden. Weitere Informationen finden Sie im Thema „Arbeiten mit Mehrfachantwortdaten“ auf Seite 82.

Vergleichsfunktionen

Mit den Funktionen `min_n` und `max_n` können Sie Werte in mehreren Feldern vergleichen. Beispiel:

```
max_n(['card1fee' 'card2fee' 'card3fee' 'card4fee'])
```

Außerdem können Sie eine Reihe von Zählfunktionen verwenden, um die Anzahl der Werte zu erhalten, die bestimmte Kriterien erfüllen, auch wenn diese Werte in mehreren Feldern gespeichert werden. So können Sie beispielsweise die Anzahl der Karten zählen, die mehr als 5 Jahre im Besitz des Kunden sind:

```
count_greater_than(5, ['cardtenure' 'card2tenure' 'card3tenure'])
```

So können Sie Nullwerte in derselben Gruppe von Feldern zählen:

```
count_nulls(['cardtenure' 'card2tenure' 'card3tenure'])
```

Beachten Sie, dass in diesem Beispiel die Anzahl der Karten im Besitz gezählt werden, nicht die Anzahl der Personen, die sie besitzen. Weitere Informationen finden Sie im Thema „Vergleichsfunktionen“ auf Seite 98.

Mit der Funktion `count_equal` können Sie zählen, wie oft ein angegebener Wert felderübergreifend vorkommt. Im folgenden Beispiel wird die Anzahl der Felder in der Liste gezählt, die den Wert Y enthalten.

```
count_equal("J",[Antwort1, Antwort2, Antwort3])
```

Wenn die folgenden Werte für die Felder in der Liste gegeben sind, gibt die Funktion die Ergebnisse für den Wert J wie angezeigt aus.

Tabelle 4. Funktionswerte.

Antwort1	Antwort2	Antwort3	Anzahl
J	N	J	Z
J	N	N	E

Numerische Funktionen

Mithilfe der Funktionen `sum_n`, `mean_n` und `sdev_n` können Sie Statistiken für mehrere Felder erhalten. Beispiele:

```
sum_n(['card1bal' 'card2bal''card3bal'])  
mean_n(['card1bal' 'card2bal''card3bal'])
```

Weitere Informationen finden Sie im Thema „Numerische Funktionen“ auf Seite 101.

Generieren von Feldlisten

Wenn Sie eine der Funktionen verwenden, die eine Liste von Feldern als Eingabe akzeptieren, können die Sonderfunktionen `@FIELDS_BETWEEN(start, ende)` und `@FIELDS_MATCHING(muster)` als Eingabe verwendet werden. Setzt man z. B. voraus, dass die Reihenfolge der Felder derjenigen im vorhergehenden Beispiel für `sum_n` entspricht, wäre Folgendes äquivalent:

```
sum_n(@FIELDS_BETWEEN(card1bal, card3bal))
```

Alternativ können Sie wie folgt die Anzahl der Nullwerte in allen Feldern zählen, die mit "card" beginnen:

```
count_nulls(@FIELDS_MATCHING('card*'))
```

Weitere Informationen finden Sie im Thema „Sonderfelder“ auf Seite 122.

Arbeiten mit Mehrfachantwortdaten

Eine Reihe von Vergleichsfunktionen kann zum Analysieren von Mehrfachantwortdaten herangezogen werden. Zu diesen zählen:

- `value_at`
- `first_index` / `last_index`
- `first_non_null` / `last_non_null`
- `first_non_null_index` / `last_non_null_index`
- `min_index` / `max_index`

Angenommen, in einer Frage mit Mehrfachantworten wird nach den drei wichtigsten Gründen bei der Entscheidung für einen bestimmten Kauf gefragt (z. B. Preis, persönliche Empfehlung, Bewertung, lokaler Lieferant, Sonstiges). In diesem Fall könnten Sie die Wichtigkeit des Preises ermitteln, indem Sie den Index des Felds ableiten, in dem dieser zum ersten Mal verwendet wurde:

```
first_index("Preis", [Grund1 Grund2 Grund3])
```

Nehmen wir analog dazu an, Sie haben Kunden gebeten, drei Autos nach ihrer Kaufwahrscheinlichkeit einzustufen, und die Antworten wie folgt in drei separate Felder codiert:

Tabelle 5. Autobewertungsbeispiele.

Kunden-ID	Auto1	Auto2	Auto3
101	E	3	Z
102	3	Z	E
103	Z	3	E

In diesem Fall könnten Sie den Index der Felder für die beliebtesten Autos (an Position 1 bzw. auf dem niedrigsten Rang eingestuft) mithilfe der `min_index`-Funktion ermitteln:

```
min_index(['Auto1' 'Auto2' 'Auto3'])
```

Weitere Informationen finden Sie im Thema „Vergleichsfunktionen“ auf Seite 98.

Referenzieren von Mehrfachantwortsets

Die spezielle `@MULTI_RESPONSE_SET`-Funktion kann dazu verwendet werden, alle Felder in einem Mehrfachantwortset zu referenzieren. Wenn die drei Felder *Auto* aus dem vorherigen Beispiel etwa in einem Mehrfachantwortset mit der Bezeichnung *Autobewertungen* enthalten sind, wird mit folgender Formel das gleiche Ergebnis erzielt:

```
max_index(@MULTI_RESPONSE_SET("Autobewertungen"))
```

Expression Builder

Sie können CLEM-Ausdrücke manuell eingeben oder Expression Builder verwenden. In dieser Funktion wird eine vollständige Liste der CLEM-Funktionen und Operatoren sowie Datenfelder aus dem aktuellen Stream angezeigt, sodass Sie schnell Ausdrücke erstellen können, ohne sich die genauen Namen der Felder bzw. Funktionen merken zu müssen. Außerdem fügt der Builder automatisch die entsprechenden Anführungszeichen für Felder und Werte hinzu, sodass es einfacher wird, syntaktisch korrekte Ausdrücke zu erstellen.

Hinweis: Expression Builder wird in Script- oder Parametereinstellungen nicht unterstützt.

Zugriff auf Expression Builder

Expression Builder ist in allen Knoten verfügbar, in denen ein CLEM-Ausdruck geschrieben werden muss, einschließlich der Auswahl-, Balancierungs-, Ableitungs-, Füller-, Analyse-, Berichts- und Tabellenknoten. Sie können ihn durch Klicken auf die Schaltfläche für den Taschenrechner direkt rechts neben dem Formelfeld öffnen.

Erstellen von Ausdrücken

Expression Builder bietet nicht nur vollständige Listen mit Feldern, Funktionen und Operatoren, sondern auch Zugriff auf Datenwerte, sofern Ihre Daten instanziiert sind.

So erstellen Sie einen Ausdruck mit Expression Builder:

1. Nehmen Sie Ihre Eingabe im Ausdrucksfeld mithilfe der Funktions- und Feldlisten vor.
ODER
2. Wählen Sie die gewünschten Felder und Funktionen aus den Listen aus.
3. Fügen Sie das Feld oder die Funktion durch Doppelklicken oder Klicken auf die gelbe Pfeilschaltfläche zum Ausdrucksfeld hinzu.
4. Fügen Sie die Operationen anhand der Operandenschaltflächen in der Mitte des Dialogfelds in den Ausdruck ein.

Auswählen von Funktionen

In der Funktionsliste werden alle verfügbaren CLEM-Funktionen und -Operatoren angezeigt. Blättern Sie, um eine Funktion aus der Liste auszuwählen, oder zeigen Sie mithilfe der Dropdown-Liste ein Subset von Funktionen oder Operatoren an, um die Suche zu erleichtern. Die verfügbaren Funktionen sind in Kategorien zusammengefasst, um die Suche zu erleichtern.

Die meisten dieser Kategorien sind im Referenzabschnitt der CLEM-Sprachbeschreibung zu finden. Weitere Informationen finden Sie im Thema „Funktionsreferenz“ auf Seite 95.

Die anderen Kategorien lauten wie folgt.

- **Allgemeine Funktionen** enthält eine Auswahl von einigen der am häufigsten verwendeten Funktionen.
- **Zuletzt verwendet** - Enthält eine Liste der CLEM-Funktionen, die in der aktuellen Sitzung verwendet wurden.
- **@ Functions** enthält eine Liste aller Sonderfunktionen, deren Namen das Zeichen "@" vorangestellt ist.
- **Datenbankfunktionen.** Wenn der Stream eine Datenbankverbindung beinhaltet (mithilfe eines Datenbankquellenknotens) listet diese Auswahl die Funktionen auf, die in dieser Datenbank zur Verfügung stehen, einschließlich benutzerdefinierter Funktionen (UDFs).
- **Datenbankaggregate.** Wenn der Stream eine Datenbankverbindung beinhaltet (mithilfe eines Datenbankquellenknotens) listet diese Auswahl die Aggregationsoptionen auf, die in dieser Datenbank zur Verfügung stehen.
- **Datenbankfensteraggregate.** Wenn der Stream eine Datenbankverbindung beinhaltet (mithilfe eines Datenbankquellenknotens) listet diese Auswahl die Fensteraggregationsoptionen auf, die Sie in dieser Datenbank verwenden können.

Obwohl diese Optionen "Aggregationsoptionen" genannt werden, sind sie nicht für die Verwendung im Aggregatknoten gedacht, sondern vielmehr für Knoten wie Ableitungs- oder Auswahlknoten anwendbar. Dies liegt daran, dass ihre Ausgabe skalar und kein echtes Aggregat ist. Das bedeutet, sie reduzieren die in der Ausgabe angezeigte Datenmenge nicht so wie der Aggregatknoten. Sie können beispielsweise diese Art der Aggregation verwenden, um einen gleitenden Durchschnitt für Datenzeilen bereitzustellen, z. B. "Durchschnitt der aktuellen Zeile und aller vorherigen Zeilen".

- **Integrierte Aggregate.** Enthält eine Liste der möglichen Aggregationsmodi, die verwendet werden können.
- **Operatoren** listet alle Operatoren auf, die Sie beim Erstellen von Ausdrücken verwenden können. Operatoren stehen auch über die Schaltflächen in der Mitte des Dialogfelds zur Verfügung.
- **Alle Funktionen** - Enthält eine vollständige Liste der verfügbaren CLEM-Funktionen.

Nachdem Sie eine Gruppe von Funktionen ausgewählt haben, fügen Sie die Funktionen durch Doppelklicken in das Ausdrucksfeld an der Stelle ein, an der sich der Cursor befindet.

Auswählen von Feldern, Parametern und globalen Variablen

In der Feldliste werden alle Felder angezeigt, die an diesem Punkt im Datenstream verfügbar sind. Blättern Sie zu einem Feld, um es aus der Liste auszuwählen. Fügen Sie durch Doppelklicken oder Klicken der gelben Pfeilschaltfläche ein Feld zum Ausdruck hinzu.

Weitere Informationen finden Sie im Thema „Stream-, Sitzungs- und Superknotenparameter“ auf Seite 78.

Neben Feldern können Sie darüber hinaus aus den folgenden Elementen auswählen:

Mehrfachantwortsets. Weitere Informationen finden Sie im *IBM SPSS Modeler-Handbuch zu Quellenknoten, Prozessknoten und Ausgabeknoten*.

Zuletzt verwendet enthält eine Liste der Felder, Mehrfachantwortsets, Parameter und globalen Werte, die in der aktuellen Sitzung verwendet wurden.

Parameter. Weitere Informationen finden Sie im Thema „Stream-, Sitzungs- und Superknotenparameter“ auf Seite 78.

Globale Werte. Weitere Informationen finden Sie im *IBM SPSS Modeler-Handbuch zu Quellenknoten, Prozessknoten und Ausgabeknoten*.

Anzeigen oder Auswählen von Werten

Feldwerte können an verschiedenen Stellen im System angezeigt werden. Dazu gehören Expression Builder, Data Audit-Berichte und beim Bearbeiten zukünftiger Werte in einem Zeitintervallknoten. Beachten Sie, dass die Daten vollständig in einem Quellen- oder Typknoten instanziiert sein müssen (Speichertypen, Typen und Werte sind bekannt), um diese Funktion verwenden zu können.

Wählen Sie das gewünschte Feld aus und klicken Sie auf die Schaltfläche zur Wertauswahl, um Werte für ein Feld über Expression Builder oder einen Zeitintervallknoten anzuzeigen. Es wird ein Dialogfeld mit Werten für das ausgewählte Feld geöffnet. Sie können dann einen Wert auswählen und auf **Einfügen** klicken, um den Wert in den aktuellen Ausdruck oder die aktuelle Liste einzufügen.



Abbildung 14. Schaltfläche zur Wertauswahl

Bei Flagfeldern und nominalen Feldern werden alle definierten Werte aufgeführt. Bei stetigen Feldern (numerischer Bereich) werden die Minimal- und Maximalwerte angezeigt.

Überprüfen von CLEM-Ausdrücken

Klicken Sie in Expression Builder auf **Überprüfen** (unten rechts), um den Ausdruck zu validieren. Ausdrücke, die nicht geprüft wurden, werden rot angezeigt. Bei Fehlern gibt eine Nachricht die Fehlerursache an.

Folgende Elemente werden überprüft:

- Richtige Anführungszeichen für Werte und Feldnamen
- Richtige Verwendung von Parametern und globalen Variablen
- Zulässige Verwendung von Operatoren
- Vorhandensein von referenzierten Feldern
- Vorhandensein und Definition von referenzierten Globalwerten

Wenn Sie Fehler in der Syntax bemerken, versuchen Sie, den Ausdruck mithilfe der Listen- und Operator-schaltflächen zu erstellen, anstatt den Ausdruck manuell einzugeben. Mit dieser Methode werden automatisch die richtigen Anführungszeichen für Felder und Werte hinzugefügt.

Suchen und Ersetzen

Das Dialogfeld "Suchen/Ersetzen" ist in Situationen verfügbar, in denen Sie Script- oder Ausdruckstext bearbeiten, u. a. im Script-Editor, im CLEM Expression Builder und bei der Definition von Vorlagen im Berichtsknoten. Während der Bearbeitung von Text in einem dieser Bereiche können Sie durch Drücken von **Strg+F** das Dialogfeld aufrufen. Achten Sie dabei darauf, dass sich der Cursor über einem Textbereich befindet. So können Sie beispielsweise bei der Arbeit in einem Füllerknoten das Dialogfeld aus einem der Textbereiche auf der Registerkarte "Einstellungen" aufrufen oder über das Textfeld im Expression Builder.

1. Achten Sie darauf, dass sich der Cursor in einem Textfeld befindet, und drücken Sie Strg+F, um das Dialogfeld "Suchen/Ersetzen" aufzurufen.
2. Geben Sie den Text ein, nach dem Sie suchen möchten, oder treffen Sie eine Auswahl aus der Drop-down-Liste der kürzlich durchsuchten Elemente.
3. Geben Sie, falls erforderlich, den Ersatztext ein.
4. Klicken Sie auf **Weitersuchen**, um die Suche zu starten.
5. Klicken Sie auf **Ersetzen**, um die aktuelle Auswahl zu ersetzen, oder **Alle ersetzen**, um alle bzw. die ausgewählten Instanzen zu ersetzen.
6. Das Dialogfeld wird nach jedem Vorgang geschlossen. Drücken Sie in einem Textbereich die Taste F3, um den letzten Suchvorgang zu wiederholen, bzw. drücken Sie Strg+F, um erneut auf das Dialogfeld zuzugreifen.

Suchoptionen

Groß-/Kleinschreibung beachten. Gibt an, ob beim Suchvorgang die Groß- und Kleinschreibung berücksichtigt wird; beispielsweise, ob *myvar* als identisch mit *myVar* betrachtet wird. Ersetzungstext wird unabhängig von dieser Einstellung stets genau so eingefügt, wie er eingegeben wurde.

Nur ganze Wörter. Gibt an, ob beim Suchvorgang auch Wortteile gefunden werden. Wenn diese Option ausgewählt ist, werden bei einer Suche nach *Tag* Terme wie *Tagesordnung* oder *Tag-und-Nacht-Gleiche* nicht als Treffer ausgegeben.

Reguläre Ausdrücke. Gibt an, ob die Syntax für reguläre Ausdrücke verwendet werden soll (siehe nächsten Abschnitt). Bei Auswahl dieser Option wird die Option **Nur ganze Wörter** inaktiviert und ihr Wert ignoriert.

Nur ausgewählter Text. Legt den Suchumfang bei Verwendung der Option **Alle ersetzen** fest.

Syntax für reguläre Ausdrücke

Mithilfe von regulären Ausdrücken können Sie nach Sonderzeichen (z. B. Tabulatoren oder Zeilenumbrüche), Zeichenklassen bzw. -bereichen, wie *a* bis *d*, beliebige Ziffer oder Nichtziffer, und nach Begrenzungen, wie beispielsweise Zeilenanfang bzw. Zeilenende, suchen. Folgende Arten von Ausdrücken werden unterstützt:

Tabelle 6. Zeichenübereinstimmungen.

Zeichen	Übereinstimmung/Bedeutung
x	Das Zeichen x
\\	Umgekehrter Schrägstrich
\0n	Das Zeichen mit dem Oktalwert 0n (0 <= n <= 7)
\0nn	Das Zeichen mit dem Oktalwert 0nn (0 <= n <= 7)
\0mnn	Das Zeichen mit dem Oktalwert 0mnn (0 <= m <= 3; 0 <= n <= 7)
\xhh	Das Zeichen mit dem Hexadezimalwert 0xhh
\uhhhh	Das Zeichen mit dem Hexadezimalwert 0xhhhh
\t	Das Tabulatorzeichen ('\u0009')
\n	Das Zeilenvorschubzeichen ('\u000A')
\r	Das Rücklaufzeichen ('\u000D')
\f	Das Formularvorschubzeichen ('\u000C')
\a	Das Alarmzeichen ('\u0007')
\e	Das Escapezeichen ('\u001B')

Table 6. Zeichenübereinstimmungen (Forts.).

Zeichen	Übereinstimmung/Bedeutung
\cx	Das Steuerzeichen, das x entspricht

Table 7. Übereinstimmende Zeichenklassen.

Zeichenklassen	Übereinstimmung/Bedeutung
[abc]	a, b oder c (einfache Klasse)
[^abc]	Beliebiges Zeichen mit Ausnahme von a, b oder c (Differenzmenge)
[a-zA-Z]	a bis einschließlich z bzw. A bis einschließlich Z (Bereich)
[a-d[m-p]]	a bis d bzw. m bis P (Vereinigungsmenge) Alternative Angabemöglichkeit: [a-dm-p]
[a-z&&[def]]	a bis z und d, e bzw. f (Schnittmenge)
[a-z&&[^bc]]	a bis z mit Ausnahme von b und c (Differenzmenge) Alternative Angabemöglichkeit: [ad-z]
[a-z&&[^m-p]]	a bis z mit Ausnahme von m bis p (Differenzmenge) Alternative Angabemöglichkeit: [a-lq-z]

Table 8. Vordefinierte Zeichenklassen.

Vordefinierte Zeichenklassen	Übereinstimmung/Bedeutung
.	Beliebiges Zeichen (evtl. Übereinstimmung mit Zeilenabschlusszeichen)
\d	Beliebige Ziffer: [0-9]
\D	Nichtziffer: [^0-9]
\s	Ein Leerzeichen: [\t\n\r]
\S	Ein Nichtleerzeichen: [^\s]
\w	Ein Zeichen: [a-zA-Z_0-9]
\W	Ein Nichtzeichen: [^\w]

Table 9. Übereinstimmungen mit Begrenzungszeichen.

Zeichen(folgen) für Begrenzungszeichen	Übereinstimmung/Bedeutung
^	Zeilenanfang
\$	Zeilenende
\b	Wortgrenze
\B	Nichtwortgrenze
\A	Beginn der Eingabe
\Z	Ende der Eingabe mit Ausnahme des letzten Abschlusszeichens, sofern vorhanden
\z	Ende der Eingabe

Kapitel 8. CLEM-Sprachreferenz

CLEM-Referenz - Überblick

In diesem Abschnitt wird CLEM (Control Language for Expression Manipulation) beschrieben, ein leistungsstarkes Tool zur Analyse und Bearbeitung der in IBM SPSS Modeler-Streams verwendeten Daten. CLEM kann in Knoten zur Ausführung verschiedener Aufgaben eingesetzt werden, vom Auswerten von Bedingungen über das Ableiten von Werten bis hin zum Einfügen von Daten in Berichte.

Ein Subset der CLEM-Sprache kann außerdem in Scripts in der Benutzerschnittstelle verwendet werden. Dadurch können Sie viele gleichartige Datenbearbeitungen automatisieren.

CLEM-Ausdrücke bestehen aus Werten, Feldnamen, Operatoren und Funktionen. Mit der richtigen Syntax können Sie eine Vielzahl leistungsstarker Datenoperationen erstellen.

CLEM-Datentypen

CLEM-Datentypen können aus folgenden Elementen bestehen:

- Ganze Zahlen
- Reelle Zahlen
- Zeichen
- Zeichenfolgen
- Listen
- Felder
- Datum/Uhrzeit

Regeln für die Verwendung von Anführungszeichen

IBM SPSS Modeler ist zwar flexibel bei der Bestimmung der in einem CLEM-Ausdruck verwendeten Felder, Werte, Parameter und Zeichenfolgen. Dennoch ist es ratsam, sich beim Erstellen von Ausdrücken an die unten angegebenen allgemeinen Regeln zu halten.

- Zeichenfolgen - Setzen Sie Zeichenfolgen immer in doppelte Anführungszeichen, beispielsweise "Typ 2". Bei einfachen Anführungszeichen besteht die Gefahr, dass sie mit in Anführungszeichen gesetzten Feldern verwechselt werden.
- Felder - Verwenden Sie einfache Anführungszeichen nur, wenn es erforderlich ist, um Leerzeichen oder andere Sonderzeichen einzuschließen. Beispiel: 'Bestellung Nummer'. Felder, die in Anführungszeichen gesetzt sind, im Dataset jedoch nicht definiert sind, werden irrtümlich als Zeichenfolgen interpretiert.
- Parameter - Setzen Sie Parameter immer in einfache Anführungszeichen, wie beispielsweise '\$P-threshold'.
- Zeichen - Zeichen müssen stets in einfache umgekehrte Anführungszeichen (``) gesetzt werden. Beispiel: stripchar(`d`, "drugA").

Diese Regeln werden in den folgenden Themen genauer behandelt.

Ganze Zahlen

Ganze Zahlen werden als Folge von Dezimalziffern dargestellt. Optional können Sie der ganzen Zahl ein Minuszeichen (-) voranstellen, um eine negative Zahl anzugeben. Beispiele: 1.234, 999, -77.

Die CLEM-Sprache kann mit ganzen Zahlen beliebiger Genauigkeit umgehen. Die maximale Größe der ganzen Zahlen hängt von Ihrer Plattform ab. Wenn die Werte zu groß für die Anzeige in einem Feld ganzer Zahlen sind, kann der Wert normalerweise durch Ändern des Feldtyps in `Reelle Zahl` wiederhergestellt werden.

Reelle Zahlen

Reelle Zahlen sind Gleitkommazahlen. Reelle Zahlen bestehen aus einer oder mehreren Ziffern, gefolgt von einem Punkt als Dezimaltrennzeichen, gefolgt von mindestens einer weiteren Ziffer. Bei CLEM gilt für die reellen Zahlen doppelte Genauigkeit.

Optional können Sie der reellen Zahl ein Minuszeichen (–) voranstellen, um eine negative Zahl anzugeben. Beispiele: 1.234, 0,999, –77.001. Mit dem Format `<Zahl> e <Exponent>` können Sie eine reelle Zahl in Exponentialnotation angeben. Beispiel `1234,0e5`, `1,7e–2`. Wenn die IBM SPSS Modeler-Anwendung numerische Zeichenfolgen aus Dateien einliest und sie automatisch in Zahlen umwandelt, werden Zahlen ohne führende Ziffer vor dem Dezimalpunkt oder ohne Ziffer nach dem Dezimaltrennzeichen akzeptiert. Beispiel: `999`, oder `,11`. Bei CLEM-Ausdrücken sind diese Formate jedoch nicht zulässig.

Hinweis: Bei der Verwendung reeller Zahlen in CLEM-Ausdrücken muss ein Punkt als Dezimaltrennzeichen verwendet werden, unabhängig von den verwendeten Ländereinstellungen und den Einstellungen für den aktuellen Stream. So müssen Sie beispielsweise angeben:

```
Na > 0.6
```

Und nicht:

```
Na > 0,6
```

Dies gilt auch dann, wenn im Dialogfeld für die Streameigenschaften das Komma als Dezimaltrennzeichen ausgewählt wurde. Diese Vorgehensweise entspricht der allgemeinen Richtlinie, dass Codesyntax unabhängig von bestimmten Ländereinstellungen bzw. Konventionen sein sollte.

Zeichen

Zeichen (in der Regel als ZEICHEN angezeigt) werden in CLEM-Ausdrücken normalerweise zur Durchführung von Tests an Zeichenfolgen verwendet. Mit der Funktion `isuppercode` können Sie beispielsweise ermitteln, ob das erste Zeichen einer Zeichenfolge ein Großbuchstabe ist. Im folgenden CLEM-Ausdruck wird ein Zeichen verwendet, um anzugeben, dass der Test für das erste Zeichen der Zeichenfolge durchgeführt werden soll.

```
isuppercode(subscrs(1, "MyString"))
```

Um den Code (im Gegensatz zur Position) eines bestimmten Zeichens in einem CLEM-Ausdruck anzugeben, verwenden Sie einzelne umgekehrte Anführungszeichen in der Form ``<Zeichen>``. Beispiel: ``A``, ``Z``.

Hinweis: Bei Feldern gibt es keinen Speichertyp CHAR. Wenn also ein Feld abgeleitet oder mit einem Ausdruck ausgefüllt wird, der zu CHAR führt, wird das Ergebnis in eine Zeichenfolge umgewandelt.

Zeichenfolgen

Im Allgemeinen sollten Zeichenfolgen in doppelte Anführungszeichen gesetzt werden. Beispiele für Zeichenfolgen sind `"c35Produkt2"` und `"ReferenzID"`. Zur Kennzeichnung von Sonderzeichen in Zeichenfolgen werden umgekehrte Schrägstriche verwendet, z. B.: `"\$65443"`. (Zur Kennzeichnung eines umgekehrten Schrägstrichs wird ein doppelter umgekehrter Schrägstrich verwendet: `\\.`) Zeichenfolgen können auch in einfache Anführungszeichen gesetzt werden. Das Ergebnis lässt sich jedoch nicht von einem in Anführungszeichen gesetzten Feld unterscheiden (`'ReferenzID'`). Weitere Informationen finden Sie im Thema „Zeichenfolgefunktionen“ auf Seite 105.

Listen

Eine Liste ist eine geordnete Abfolge von Elementen, die verschiedenen Typen angehören können. Listen werden in eckige Klammern ([]) eingeschlossen. Beispiele für Listen sind [1 2 4 16] und ["abc" "def"]. Listen werden nicht als Wert von IBM SPSS Modeler-Feldern verwendet. Sie werden verwendet, um Argumente für Funktionen, beispielsweise member und oneof, bereitzustellen

Felder

Namen in CLEM-Ausdrücken, bei denen es sich nicht um Namen von Funktionen handelt, werden als Feldnamen betrachtet. Namen können einfach als Energie, val27, state_flag usw. geschrieben werden, aber wenn der Name mit einer Ziffer beginnt oder nicht alphabetische Zeichen, beispielsweise Leerzeichen, (Ausnahme: Unterstrich) enthält, muss der Name in einfache Anführungszeichen gesetzt werden. Beispiel: 'Anstieg der Energie', '2. Antwort', '#101', '\$P-NextField'.

Hinweis: Felder, die in Anführungszeichen gesetzt sind, im Dataset jedoch nicht definiert sind, werden irrtümlich als Zeichenfolgen interpretiert.

Datumsangaben

Die Datumsberechnungen beruhen auf einem "Basisdatum", das im Dialogfeld "Streameigenschaften" angegeben wird. Standardmäßig wird als Basisdatum der 1. Januar 1900 verwendet.

Die CLEM-Sprache unterstützt folgende Datumsformate.

Table 10. Datumsformate der CLEM-Sprache

Format	Beispiele
TTMMJJ	150163
MMTTJJ	011563
JJMMTT	630115
JJJJMMTT	19630115
JJJJTTT	Vierstellige Jahreszahl, gefolgt von einer dreistelligen Zahl für den Tag im Jahr. So steht 2000032 für den 32. Tag des Jahres 2000, also für den 1. Februar 2000.
DAY	Tag der Woche in der aktuellen Ländereinstellung - z. B. Montag, Dienstag, ..., für Deutsch.
MONTH	Monat in der aktuellen Ländereinstellung - z. B. Januar, Februar, ...
TT/MM/JJ	15/01/63
TT/MM/JJJJ	15/01/1963
MM/TT/JJ	01/15/63
MM/TT/JJJJ	01/15/1963
TT-MM-JJ	15-01-63
TT-MM-JJJJ	15-01-1963
MM-TT-JJ	01-15-63
MM-TT-JJJJ	01-15-1963
TT.MM.JJ	15.01.63
TT.MM.JJJJ	15.01.1963
MM.TT.JJ	01.15.63
MM.TT.JJJJ	01.15.1963
TT-MON-JJ	15-JAN-63, 15-jan-63, 15-Jan-63

Tabelle 10. Datumsformate der CLEM-Sprache (Forts.)

Format	Beispiele
TT/MON/JJ	15/JAN/63, 15/jan/63, 15/Jan/63
TT.MON.JJ	15.JAN.63, 15.jan.63, 15.Jan.63
TT-MON-JJJJ	15-JAN-1963, 15-jan-1963, 15-Jan-1963
TT/MON/JJJJ	15/JAN/1963, 15/jan/1963, 15/Jan/1963
TT.MON.JJJJ	15.JAN.1963, 15.jan.1963, 15.Jan.1963
MON JJJJ	Jan 2004
q Q JJJJ	Datum als Ziffer (1-4) für das Quartal, gefolgt vom Buchstaben Q und einer vierstelligen Jahresangabe. So wird beispielsweise das Datum 25. Dez. 2004 als 4 Q 2004 dargestellt.
ww WK JJJJ	Zweistellige Zahl für die Kalenderwoche, gefolgt von den Buchstaben WK und dann von einer vierstelligen Jahreszahl. Bei der Berechnung der Kalenderwoche wird davon ausgegangen, dass Montag der erste Wochentag ist, und dass mindestens ein Tag in der ersten Kalenderwoche liegt.

Zeit

Die CLEM-Sprache unterstützt folgende Zeitformate.

Tabelle 11. Zeitformate der CLEM-Sprache

Format	Beispiele
HHMMSS	120112, 010101, 221212
HHMM	1223, 0745, 2207
MMSS	5558, 0100
HH:MM:SS	12:01:12, 01:01:01, 22:12:12
HH:MM	12:23, 07:45, 22:07
MM:SS	55:58, 01:00
(H)H:(M)M:(S)S	12:1:12, 1:1:1, 22:12:12
(H)H:(M)M	12:23, 7:45, 22:7
(M)M:(S)S	55:58, 1:0
HH.MM.SS	12.01.12, 01.01.01, 22.12.12
HH.MM	12.23, 07.45, 22.07
MM.SS	55.58, 01.00
(H)H.(M)M.(S)S	12.1.12, 1.1.1, 22.12.12
(H)H.(M)M	12.23, 7.45, 22.7
(M)M.(S)S	55.58, 1.0

CLEM-Operatoren

Folgende Operatoren stehen zur Verfügung.

Tabelle 12. CLEM-Sprachoperatoren.

Operation	Kommentare	Rangfolge (siehe nächsten Abschnitt)
or	Wird zwischen zwei CLEM-Ausdrücken verwendet. Ergibt den Wert "Wahr", wenn mindestens einer der beiden Terme wahr ist.	10
and	Wird zwischen zwei CLEM-Ausdrücken verwendet. Ergibt den Wert "Wahr", wenn beide Terme "wahr" sind.	9
=	Wird zwischen zwei beliebigen miteinander vergleichbaren Elementen verwendet. Ergibt den Wert "Wahr", wenn ELEMENT1 gleich ELEMENT2 ist.	7
==	Identisch mit =.	7
/=	Wird zwischen zwei beliebigen miteinander vergleichbaren Elementen verwendet. Ergibt den Wert "Wahr", wenn ELEMENT1 <i>nicht</i> gleich ELEMENT2 ist.	7
/==	Identisch mit /=.	7
>	Wird zwischen zwei beliebigen miteinander vergleichbaren Elementen verwendet. Ergibt den Wert "Wahr", wenn ELEMENT1 größer als ELEMENT2 ist.	6
>=	Wird zwischen zwei beliebigen miteinander vergleichbaren Elementen verwendet. Ergibt den Wert "Wahr", wenn ELEMENT1 größer oder gleich ELEMENT2 ist.	6
<	Wird zwischen zwei beliebigen miteinander vergleichbaren Elementen verwendet. Ergibt den Wert "Wahr", wenn ELEMENT1 kleiner als ELEMENT2 ist.	6
<=	Wird zwischen zwei beliebigen miteinander vergleichbaren Elementen verwendet. Ergibt den Wert "Wahr", wenn ELEMENT1 kleiner oder gleich ELEMENT2 ist.	6
&&=_0	Wird zwischen zwei ganzen Zahlen verwendet. Entspricht dem booleschen Ausdruck GANZZ1 && GANZZ2 = 0.	6
&&/=_0	Wird zwischen zwei ganzen Zahlen verwendet. Entspricht dem booleschen Ausdruck GANZZ1 && GANZZ2 /= 0.	6
+	Addiert zwei Zahlen: ZAHL1 + ZAHL2.	5
><	Verkettet zwei Zeichenfolgen; beispielsweise ZEICHENFOLGE1 >< ZEICHENFOLGE2.	5
-	Subtrahiert eine Zahl von einer anderen Zahl: ZAHL1 - ZAHL2. Kann auch vor einer Zahl verwendet werden: - NUM.	5
*	Dient zur Multiplikation zweier Zahlen: ZAHL1 * ZAHL2.	4

Tabelle 12. CLEM-Sprachoperatoren (Forts.).

Operation	Kommentare	Rangfolge (siehe nächsten Abschnitt)
&&	Wird zwischen zwei ganzen Zahlen verwendet. Das Ergebnis liefert das bitweise "Und" der ganzen Zahlen GANZZ1 und GANZZ2.	4
&&~	Wird zwischen zwei ganzen Zahlen verwendet. Das Ergebnis liefert das bitweise "Und" von GANZZ1 und das bitweise Komplement von GANZZ2.	4
	Wird zwischen zwei ganzen Zahlen verwendet. Das Ergebnis liefert das bitweise "Oder" von GANZZ1 und GANZZ2.	4
~	Wird vor einer ganzen Zahl verwendet. Liefert das bitweise Komplement von GANZZ.	4
/&	Wird zwischen zwei ganzen Zahlen verwendet. Das Ergebnis liefert das bitweise "exklusive Oder" von GANZZ1 und GANZZ2.	4
GANZ1 << N	Wird zwischen zwei ganzen Zahlen verwendet. Ergibt das Bitmuster von GANZZ, um N Positionen nach links verschoben.	4
GANZ1 >> N	Wird zwischen zwei ganzen Zahlen verwendet. Ergibt das Bitmuster von GANZZ, um N Positionen nach rechts verschoben.	4
/	Dient zur Division zweier Zahlen: ZAHL1 / ZAHL2.	4
**	Wird zwischen zwei Zahlen verwendet: BASIS ** POTENZ. Ergibt BASIS hoch POTENZ.	3
rem	Wird zwischen zwei ganzen Zahlen verwendet: GANZZ1 rem GANZZ2. Ergibt den Rest: GANZZ1 - (GANZZ1 div GANZZ2) * GANZZ2.	Z
div	Wird zwischen zwei ganzen Zahlen verwendet: GANZZ1 div GANZZ2. Führt eine Division zwischen zwei ganzen Zahlen durch.	Z

Rangfolge der Operatoren

Rangfolgen bestimmen die Analyse komplexer Ausdrücke, insbesondere von Ausdrücken ohne Klammern mit mehreren Infix-Operatoren. Beispiel:

$$3 + 4 * 5$$

wird als $3 + (4 * 5)$ und nicht als $(3 + 4) * 5$ analysiert, da die relative Rangfolge festlegt, dass * vor + verarbeitet werden muss. Jedem Operator in der CLEM-Sprache ist ein Rangfolgewert zugeordnet; je niedriger dieser Wert ist, desto wichtiger ist er auf der Analyseliste, was bedeutet, dass er vor anderen Operatoren mit höheren Rangfolgewerten verarbeitet wird.

Funktionsreferenz

Folgende CLEM-Funktionen stehen bei der Arbeit mit Daten in IBM SPSS Modeler zur Verfügung. Diese Funktionen können in einer Reihe von Dialogfeldern (z. B. "Ableitungsknoten" und "Dichotomknoten") als Code eingegeben werden. Alternativ können Sie mit dem Expression Builder gültige CLEM-Ausdrücke erstellen, ohne sich Funktionslisten oder Feldnamen einprägen zu müssen.

Table 13. CLEM-Funktionen für die Verwendung mit IBM SPSS Modeler-Daten

Funktionstyp	Beschreibung
Informationsfunktionen	Dienen dazu, Erkenntnisse über Feldwerte zu gewinnen. Beispiel: Die Funktion <code>is_string</code> gibt für alle Datensätze vom Typ <code>string</code> den Wert "Wahr" zurück.
Umwandlung	Dient zur Erstellung neuer Felder oder zur Konvertierung des Speichertyps. Die Funktion <code>to_timestamp</code> konvertiert das ausgewählte Feld beispielsweise in eine Zeitmarke.
Vergleichsfunktionen	Dienen zum Vergleichen von Feldwerten untereinander oder mit einer angegebenen Zeichenfolge. Beispiel: Mit <code><=</code> verglichen, ob die Werte zweier Felder kleiner oder gleich sind.
Logische Funktionen	Dienen zur Durchführung logischer Operationen wie <code>if</code> (wenn), <code>then</code> (dann), <code>else</code> (sonst).
Numerische Funktionen	Dienen zur Durchführung numerischer Berechnungen, wie dem natürlichen Logarithmus von Feldwerten.
Trigonometrisch	Dient zur Durchführung trigonometrischer Berechnungen, wie Berechnung des Arkuskosinus eines bestimmten Winkels.
Wahrscheinlichkeit	Rückgabe von Wahrscheinlichkeiten auf der Grundlage verschiedener Verteilungen, z. B. die Wahrscheinlichkeit, dass ein Wert aus der <i>t</i> -Verteilung in der Studentenversion kleiner ist als ein bestimmter Wert.
Bitweise	Dient zur Bearbeitung von ganzen Zahlen als Bitmuster.
Zufallsfunktionen	Dienen zur zufälligen Auswahl von Elementen oder Generierung von Zahlen.
Zeichenfolge	Dient zur Durchführung diverser Operationen an Zeichenfolgen, wie beispielsweise <code>stripchar</code> zum Entfernen eines angegebenen Zeichens.
SoundEx	Dient zum Auffinden von Zeichenfolgen, deren genaue Schreibweise nicht bekannt ist, und zwar auf der Grundlage fonetischer Annahmen, wie bestimmte Buchstaben ausgesprochen werden.
Datums- und Uhrzeitfunktionen	Dienen zur Durchführung verschiedenster Operationen von Datums-, Zeit- und Zeitmarkenfeldern.
Sequenz	Dient zur Gewinnung von Erkenntnissen über die Datensatzsequenz eines Datasets oder zur Durchführung von Operationen basierend auf dieser Sequenz.
Globalwert	Ermöglicht den Zugriff auf Globalwerte, die von einem Globalwerteknoten erstellt wurden. Beispiel: <code>@MEAN</code> wird als Verweis auf den mittleren Durchschnitt aller Werte für ein Feld im gesamten Dataset verwendet.
Funktionen zum Umgang mit Leerstellen und Nullwerten	Dienen zum Zugreifen, Kennzeichnen und häufig zum Ausfüllen benutzerdefinierter Leerstellen oder systemdefiniert fehlender Werte. Beispiel: <code>@BLANK(FELD)</code> wird verwendet, um Datensätze mit Leerstellen als "wahr" zu kennzeichnen.
Sonderfelder	Dienen zur Angabe der gesonderten Felder, die untersucht werden. <code>@FIELD</code> wird beispielsweise beim Ableiten mehrerer Felder verwendet.

Konventionen bei Funktionsbeschreibungen

Folgende Konventionen werden überall in diesem Dokument für Elemente in einer Funktion verwendet.

Table 14. Konventionen bei Funktionsbeschreibungen.

Konvention	Beschreibung
<i>BOOL</i>	Ein boolescher Wert bzw. ein boolesches Flag, wie "wahr" oder "falsch".
<i>ZAHL, ZAHL1, ZAHL2</i>	Eine beliebige Zahl.
<i>REELL, REELL1, REELL2</i>	Eine reelle Zahl, wie 1,234 oder -77,01.
<i>GANZZ, GANZZ1, GANZZ2</i>	Eine ganze Zahl, wie 1 oder -77.
<i>ZEICHEN</i>	Eine Zeichencode, wie beispielsweise `A`.
<i>ZEICHENFOLGE</i>	Ein Zeichenfolge, wie beispielsweise "ReferenzID".
<i>LISTE</i>	Eine Liste mit Elementen, wie beispielsweise ["abc" "def"].
<i>ELEMENT</i>	Ein Feld, wie beispielsweise Kunde oder Konzept_extrahieren.
<i>DATUM</i>	Ein Datumsfeld, beispielsweise Startdatum, mit Werten in Formaten wie TT-MON-JJJJ.
<i>ZEIT</i>	Ein Zeitfeld, beispielsweise Energiefluss, mit Werten in Formaten wie HHMMSS.

Die Funktionen in diesem Handbuch sind wie folgt aufgeführt: Die Funktion in der ersten Spalte, der Ergebnistyp (ganze Zahl, Zeichenfolge usw.) in der zweiten und gegebenenfalls eine Beschreibung in der dritten Spalte. Folgendes Beispiel ist eine Beschreibung der Funktion *rem*.

Table 15. Beschreibung der Funktion "rem".

Funktion	Ergebnis	Beschreibung
<i>GANZZ1 rem GANZZ2</i>	<i>Zahl</i>	Ergibt den Rest des Quotienten aus <i>GANZZ1</i> und <i>GANZZ2</i> . Beispiel: $GANZZ1 - (GANZZ1 \text{ div } GANZZ2) * GANZZ2$.

Einzelheiten zur den Verwendungskonventionen, beispielsweise zur Auflistung von Elementen oder zur Angabe von Zeichen in einer Funktion, werden an anderer Stelle beschrieben. Weitere Informationen finden Sie im Thema „CLEM-Datentypen“ auf Seite 89.

Informationsfunktionen

Informationsfunktionen werden verwendet, um einen Einblick in die Werte eines bestimmten Felds zu gewinnen. Sie werden üblicherweise verwendet, um Flagfelder abzuleiten. Beispiel: Mit der Funktion @BLANK können Sie ein Flagfeld erstellen, das die Datensätze angibt, deren Werte für das ausgewählte Feld leer sind. In ähnlicher Weise können Sie den Speichertyp für ein Feld mit einer der Speichertypfunktionen überprüfen, beispielsweise mit *is_string*.

Table 16. CLEM-Informationsfunktionen

Funktion	Ergebnis	Beschreibung
@BLANK(FELD)	<i>boolesch</i>	Ergibt den Wert "Wahr" für alle Datensätze, deren Werte gemäß den Regeln zum Umgang mit Leerstellen, die in einem weiter oben im Stream gelegenen Typknoten oder Quellenknoten (Registerkarte "Typen") festgelegt wurden, Leerstellen sind. Beachten Sie, dass diese Funktion nicht über ein Script aufgerufen werden kann.

Tabelle 16. CLEM-Informationsfunktionen (Forts.)

Funktion	Ergebnis	Beschreibung
@NULL(ELEMENT)	<i>boolesch</i>	Ergibt den Wert "Wahr" für alle Datensätze, deren Werte nicht definiert sind. Nicht definierte Werte sind systemdefinierte Nullwerte, die in IBM SPSS Modeler als \$null\$ angezeigt werden. Beachten Sie, dass diese Funktion nicht über ein Script aufgerufen werden kann.
is_date(ELEMENT)	<i>boolesch</i>	Ergibt den Wert "Wahr" für alle Datensätze, deren Typ ein Datum ist.
is_datetime(ELEMENT)	<i>boolesch</i>	Ergibt den Wert "Wahr" für alle Datensätze, deren Typ "Datum", "Zeit" oder "Zeitmarke" ist.
is_integer(ELEMENT)	<i>boolesch</i>	Ergibt den Wert "Wahr" für alle Datensätze, deren Typ eine ganze Zahl ist.
is_number(ELEMENT)	<i>boolesch</i>	Ergibt den Wert "Wahr" für alle Datensätze, deren Typ eine Zahl ist.
is_real(ELEMENT)	<i>boolesch</i>	Ergibt den Wert "Wahr" für alle Datensätze, deren Typ eine reelle Zahl ist.
is_string(ELEMENT)	<i>boolesch</i>	Ergibt den Wert "Wahr" für alle Datensätze, deren Typ eine Zeichenfolge ist.
is_time(ELEMENT)	<i>boolesch</i>	Ergibt den Wert "Wahr" für alle Datensätze, deren Typ eine Zeitangabe ist.
is_timestamp(ELEMENT)	<i>boolesch</i>	Ergibt den Wert "Wahr" für alle Datensätze, deren Typ eine Zeitmarke ist.

Konvertierungsfunktionen

Mit Konvertierungsfunktionen können Sie neue Felder erstellen und den Speichertyp bestehender Felder umwandeln. Beispielsweise können Sie neue Zeichenfolgen bilden, indem Sie Zeichenfolgen miteinander verbinden oder zerlegen. Zwei Zeichenfolgen können mit dem Operator >< miteinander verbunden werden. Wenn das Feld Standort beispielsweise den Wert "EISENACH" aufweist, ergibt "xx" >< Standort den Wert "xxEISENACH". Das Ergebnis von >< ist immer eine Zeichenfolge, selbst wenn die einzelnen Argumente keine Zeichenfolgen sind. Wenn also Feld V1 den Wert 3 und Feld V2 den Wert 5 aufweist, ergibt V1 >< V2 den Wert "35" (eine Zeichenfolge, keine Zahl).

Konvertierungsfunktionen (und alle anderen Funktionen, für die ein spezieller Eingabetyp, wie beispielsweise ein Wert für Datum oder Uhrzeit, erforderlich ist) hängen von den aktuell im Dialogfeld für die Streamoptionen angegebenen Formaten ab. Wenn Sie beispielsweise ein Zeichenfolgenfeld mit den Werten *Jan 2003*, *Feb 2003* usw. konvertieren möchten, müssen Sie das zugehörige Datumsformat, **MON JJJJ**, als Standarddatumsformat für den Stream auswählen.

Tabelle 17. CLEM-Konvertierungsfunktionen

Funktion	Ergebnis	Beschreibung
ELEMENT1 >< ELEMENT2	<i>Zeichenfolge</i>	Zieht die Werte der beiden Felder zusammen und ergibt die resultierende Zeichenfolge als <i>ELEMENT1ELEMENT2</i> .
to_integer(ELEMENT)	<i>ganze Zahl</i>	Konvertiert den Speichertyp des angegebenen Felds in eine ganze Zahl.
to_real(ELEMENT)	<i>reelle Zahl</i>	Konvertiert den Speichertyp des angegebenen Felds in eine reelle Zahl.
to_number(ELEMENT)	<i>Zahl</i>	Konvertiert den Speichertyp des angegebenen Felds in eine Zahl.
to_string(ELEMENT)	<i>Zeichenfolge</i>	Konvertiert den Speichertyp des angegebenen Felds in eine Zeichenfolge.

Tabella 17. CLEM-Konvertierungsfunktionen (Forts.)

Funktion	Ergebnis	Beschreibung
to_time(ELEMENT)	Zeit	Konvertiert den Speichertyp des angegebenen Felds in eine Zeit.
to_date(ELEMENT)	Datum	Konvertiert den Speichertyp des angegebenen Felds in ein Datum.
to_timestamp(ELEMENT)	Zeitmarke	Konvertiert den Speichertyp des angegebenen Felds in eine Zeitmarke.
to_datetime(ELEMENT)	Datum/Uhrzeit	Konvertiert den Speichertyp des angegebenen Felds in einen Wert vom Typ "Datum", "Zeit" oder "Zeitmarke".
datetime_date(ELEMENT)	Datum	Gibt den Datumswert für eine Zahl, eine Zeichenfolge oder Zeitmarke zurück. Beachten Sie, dass dies die einzige Funktion ist, mit der Sie eine Zahl (in wenigen Sekunden) wieder in ein Datum zurückkonvertieren können. Wenn es sich bei ELEMENT um eine Zeichenfolge handelt, wird ein Datum erstellt, indem eine Zeichenfolge im aktuellen Datumsformat analysiert wird. Das im Dialogfeld für die Stromeigenschaften angegebene Datumsformat muss korrekt sein, damit diese Funktion erfolgreich ausgeführt werden kann. Wenn es sich bei ELEMENT um eine Zahl handelt, wird sie als Anzahl von Sekunden seit dem Basisdatum (Epoche) interpretiert. Die Bruchteile eines Tages werden gekürzt. Wenn es sich bei ELEMENT um eine Zeitmarke handelt, wird der Datumsteil der Zeitmarke ausgegeben. Wenn es sich bei ELEMENT um ein Datum handelt, wird es unverändert ausgegeben.

Vergleichsfunktionen

Mit Vergleichsfunktionen werden Feldwerte miteinander oder mit einer angegebenen Zeichenfolge verglichen. Beispielsweise können Sie Zeichenfolgen mit = auf Gleichheit überprüfen. Ein Beispiel für die Überprüfung der Zeichenfolgengleichheit ist: Klasse = "Klasse 1".

Bei numerischen Vergleichen bedeutet *größer* näher an der positiven Unendlichkeit und *kleiner* näher an der negativen Unendlichkeit. Negativen Zahlen sind also stets kleiner als jede positive Zahl.

Tabella 18. CLEM-Vergleichsfunktionen

Funktion	Ergebnis	Beschreibung
count_equal(ELEMENT1, LISTE)	ganze Zahl	Ergibt die Anzahl der Werte aus einer Felderliste, die gleich ELEMENT1 sind, bzw. den Wert null, wenn ELEMENT1 gleich null ist.
count_greater_than(ELEMENT1, LISTE)	ganze Zahl	Ergibt die Anzahl der Werte aus einer Felderliste, die größer als ELEMENT1 sind, bzw. den Wert null, wenn ELEMENT1 gleich null ist.
count_less_than(ELEMENT1, LISTE)	ganze Zahl	Ergibt die Anzahl der Werte aus einer Felderliste, die kleiner als ELEMENT1 sind, bzw. den Wert null, wenn ELEMENT1 gleich null ist.
count_not_equal(ELEMENT1, LISTE)	ganze Zahl	Ergibt die Anzahl der Werte aus einer Felderliste, die ungleich ELEMENT1 sind, bzw. den Wert null, wenn ELEMENT1 gleich null ist.
count_nulls(LISTE)	ganze Zahl	Ergibt die Anzahl der Nullwerte aus einer Felderliste.
count_non_nulls(LISTE)	ganze Zahl	Ergibt die Anzahl der Nichtnullwerte aus einer Felderliste.
date_before(DATUM1, DATUM2)	boolesch	Dient zur Überprüfung der alphabetischen Sortierung von Datumswerten. Ergibt den Wert "Wahr", wenn DATUM1 vor DATUM2 liegt.

Tabelle 18. CLEM-Vergleichsfunktionen (Forts.)

Funktion	Ergebnis	Beschreibung
first_index(ELEMENT, LISTE)	ganze Zahl	Ergibt den Index des ersten Felds aus einer LISTE von Feldern, das ELEMENT enthält, oder "0", wenn der Wert nicht gefunden wird. Nur für die Typen "Zeichenfolge", "Ganze Zahl" und "Reelle Zahl" unterstützt.
first_non_null(LISTE)	beliebig	Ergibt den ersten Nichtnullwert in der angegebenen Felderliste. Alle Speichertypen werden unterstützt.
first_non_null_index(LISTE)	ganze Zahl	Ergibt den Index des ersten Felds in der angegebenen LISTE, das einen Nichtnullwert enthält, oder "0", wenn alle Werte null sind. Alle Speichertypen werden unterstützt.
ELEMENT1 = ELEMENT2	boolesch	Ergibt den Wert "Wahr" für Datensätze, bei denen ELEMENT1 gleich ELEMENT2 ist.
ELEMENT1 /= ELEMENT2	boolesch	Ergibt den Wert "Wahr", wenn die beiden Zeichenfolgen nicht identisch sind, oder "0", wenn sie identisch sind.
ELEMENT1 < ELEMENT2	boolesch	Ergibt den Wert "Wahr" für Datensätze, bei denen ELEMENT1 kleiner als ELEMENT2 ist.
ELEMENT1 <= ELEMENT2	boolesch	Ergibt den Wert "Wahr" für Datensätze, bei denen ELEMENT1 kleiner oder gleich ELEMENT2 ist.
ELEMENT1 > ELEMENT2	boolesch	Ergibt den Wert "Wahr" für Datensätze, bei denen ELEMENT1 größer als ELEMENT2 ist.
ELEMENT1 >= ELEMENT2	boolesch	Ergibt den Wert "Wahr" für Datensätze, bei denen ELEMENT1 größer oder gleich ELEMENT2 ist.
last_index(ELEMENT, LISTE)	ganze Zahl	Ergibt den Index des letzten Felds aus einer LISTE von Feldern, das ELEMENT enthält, oder "0", wenn der Wert nicht gefunden wird. Nur für die Typen "Zeichenfolge", "Ganze Zahl" und "Reelle Zahl" unterstützt.
last_non_null(LISTE)	beliebig	Ergibt den letzten Nichtnullwert in der angegebenen Felderliste. Alle Speichertypen werden unterstützt.
last_non_null_index(LISTE)	ganze Zahl	Ergibt den Index des letzten Felds in der angegebenen LISTE, das einen Nichtnullwert enthält, oder "0", wenn alle Werte null sind. Alle Speichertypen werden unterstützt.
max(ELEMENT1, ELEMENT2)	beliebig	Ergibt das größere der beiden Elemente: ELEMENT1 oder ELEMENT2.
max_index(LISTE)	ganze Zahl	Ergibt den Index des Felds mit dem größten Wert aus einer Liste numerischer Felder bzw. "0", wenn alle Feldwerte null sind. Wenn beispielsweise das dritte Feld in der Liste den größten Wert enthält, wird der Indexwert 3 ausgegeben. Wenn mehrere Felder den größten Wert enthalten, wird das zuerst aufgelistete (von links nach rechts) ausgegeben.
max_n(LISTE)	Zahl	Ergibt den größten Wert aus einer Liste numerischer Felder bzw. den Wert null, wenn alle Feldwerte gleich null sind.
member(ELEMENT, LISTE)	boolesch	Ergibt den Wert "Wahr", wenn ELEMENT ein Mitglied der angegebenen LISTE ist. Andernfalls ergibt sich ein Falsch-Wert. Auch eine Liste mit Feldnamen kann angegeben werden.
min(ELEMENT1, ELEMENT2)	beliebig	Ergibt das kleinere der beiden Elemente: ELEMENT1 oder ELEMENT2.

Tabella 18. CLEM-Vergleichsfunktionen (Forts.)

Funktion	Ergebnis	Beschreibung
min_index(LISTE)	ganze Zahl	Ergibt den Index des Felds mit dem kleinsten Wert aus einer Liste numerischer Felder bzw. "0", wenn alle Feldwerte null sind. Wenn beispielsweise das dritte Feld in der Liste den kleinsten Wert enthält, wird der Indexwert 3 ausgegeben. Wenn mehrere Felder den kleinsten Wert enthalten, wird das zuerst aufgelistete (von links nach rechts) ausgegeben.
min_n(LISTE)	Zahl	Ergibt den kleinsten Wert aus einer Liste numerischer Felder bzw. den Wert null, wenn alle Feldwerte gleich null sind.
time_before(ZEIT1, ZEIT2)	boolesch	Dient zur Überprüfung der alphabetischen Sortierung von Zeitwerten. Ergibt den Wert "Wahr", wenn ZEIT1 vor ZEIT2 liegt.
value_at(GANZZ, LISTE)		Gibt den Wert jedes aufgelisteten Felds bei Offset GANZZ zurück oder NULL, wenn das Offset außerhalb des Bereichs der gültigen Werte liegt (d. h. kleiner als 1 oder größer als die Anzahl der aufgelisteten Felder ist). Alle Speichertypen werden unterstützt.

Logische Funktionen

CLEM-Ausdrücke können zur Durchführung logischer Operationen verwendet werden.

Tabella 19. Logische Funktionen von CLEM

Funktion	Ergebnis	Beschreibung
BEDG1 and BEDG2	boolesch	Bei dieser Operation handelt es sich um eine logische Konjunktion, die den Wert "Wahr" ergibt, wenn sowohl BEDG1 als auch BEDG2 wahr sind. Wenn BEDG1 falsch ist, wird BEDG2 nicht ausgewertet; dadurch sind Konjunktionen möglich, bei denen BEDG1 zuerst getestet, ob eine Operation in BEDG2 zulässig ist. Beispiele length(Label) >=6 und Label(6) = 'x'.
BEDG1 or BEDG2	boolesch	Bei dieser Operation handelt es sich um eine logische (inklusive) Disjunktion, die den Wert "Wahr" ergibt, wenn entweder BEDG1 oder BEDG2 wahr ist oder wenn beide wahr sind. Wenn BEDG1 wahr ist, wird BEDG2 nicht ausgewertet.
not(BEDG)	boolesch	Bei dieser Operation handelt es sich um eine logische Negation, die den Wert "Wahr" ergibt, wenn BEDG falsch ist. Andernfalls ergibt diese Operation den Wert 0.
if BEDG then AUSDR1 else AUSDR2 endif	beliebig	Bei dieser Operation handelt es sich um eine bedingte Auswertung. Wenn BEDG wahr ist, ergibt die Operation das Ergebnis von AUSDR1. Andernfalls ergibt sich das Ergebnis der Auswertung von AUSDR2.
if BEDG1 then AUSDR1 elseif BEDG2 then AUSDR2 else AUSDR_N endif	beliebig	Bei dieser Operation handelt es sich um eine bedingte Auswertung mit mehreren Zweigen. Wenn BEDG1 wahr ist, ergibt die Operation das Ergebnis von AUSDR1. Andernfalls, wenn BEDG2 wahr ist, ergibt die Operation das Ergebnis der Auswertung von AUSDR2. Andernfalls ergibt sich das Ergebnis der Auswertung von AUSDR_N.

Numerische Funktionen

CLEM enthält eine Reihe häufig verwendeter numerischer Funktionen.

Tabelle 20. Numerische Funktionen von CLEM.

Funktion	Ergebnis	Beschreibung
-ZAHL	Zahl	Wird zur Umkehrung des Vorzeichens von ZAHL verwendet. Ergibt die entsprechende Zahl mit dem entgegengesetzten Vorzeichen.
ZAHL1 + ZAHL2	Zahl	Ergibt die Summe von ZAHL1 und ZAHL2.
ZAHL1 – ZAHL2	Zahl	Ergibt die Differenz von ZAHL1 und ZAHL2.
ZAHL1 * ZAHL2	Zahl	Ergibt das Produkt aus ZAHL1 und ZAHL2.
ZAHL1 / ZAHL2	Zahl	Ergibt den Quotienten aus ZAHL1 und ZAHL2.
GANZZ1 div GANZZ2	Zahl	Wird zur Durchführung einer Division zwischen zwei ganzen Zahlen verwendet. Ergibt den Wert von GANZZ1 dividiert durch GANZZ2.
GANZZ1 rem GANZZ2	Zahl	Ergibt den Rest des Quotienten aus GANZZ1 und GANZZ2. Beispiel: GANZZ1 - (GANZZ1 div GANZZ2) * GANZZ2.
GANZZ1 mod GANZZ2	Zahl	Diese Funktion wird nicht mehr verwendet. Verwenden Sie stattdessen die Funktion rem.
BASIS ** POTENZ	Zahl	Ergibt den Wert von BASIS hoch POTENZ. Für beide Werte ist eine beliebige Zahl möglich (mit der Ausnahme, dass BASIS nicht null sein darf, wenn POTENZ einen anderen Nullwert als die ganze Zahl 0 aufweist). Wenn POTENZ eine ganze Zahl ist, erfolgt die Berechnung, indem BASIS mehrmals nacheinander mit sich selbst multipliziert wird. Wenn BASIS also eine ganze Zahl ist, ist das Ergebnis ebenfalls eine ganze Zahl. Wenn POTENZ die ganze Zahl 0 ist, ist das Ergebnis immer eine 1 mit demselben Typ wie BASIS. Andernfalls, wenn POTENZ keine ganze Zahl ist, wird das Ergebnis als $\exp(\text{POTENZ} * \log(\text{BASIS}))$ berechnet.
abs(ZAHL)	Zahl	Ergibt den absoluten Wert von ZAHL, bei dem es sich immer um eine Zahl desselben Typs handelt.
exp(ZAHL)	reelle Zahl	Ergibt den Wert von e hoch ZAHL. Dabei ist e die Basis eines natürlichen Logarithmus.
fracof(ZAHL)	reelle Zahl	Ergibt den Bruchbereich (Nachkommastellen) von ZAHL, der definiert ist als $\text{ZAHL} - \text{intof}(\text{ZAHL})$.
intof(ZAHL)	ganze Zahl	Kürzt das Argument zu einer ganzen Zahl. Ergibt eine ganze Zahl mit demselben Vorzeichen wie Zahl und mit dem größten Absolutwert, sodass $\text{abs}(\text{GANZZ}) \leq \text{abs}(\text{ZAHL})$.
log(ZAHL)	reelle Zahl	Ergibt den natürlichen Logarithmus (Basis e) von ZAHL (darf keine Art von Nullwert sein).
log10(ZAHL)	reelle Zahl	Ergibt den Logarithmus zur Basis 10 von ZAHL (darf keine Art von Nullwert sein). Diese Funktion ist als $\log(\text{ZAHL}) / \log(10)$ definiert.
negate(ZAHL)	Zahl	Wird zur Umkehrung des Vorzeichens von ZAHL verwendet. Ergibt die entsprechende Zahl mit dem entgegengesetzten Vorzeichen.
round(ZAHL)	ganze Zahl	Dient zum Runden von ZAHL auf eine ganze Zahl. Dabei wird $\text{intof}(\text{ZAHL}+0,5)$ verwendet, wenn ZAHL positiv ist, oder $\text{intof}(\text{ZAHL}-0,5)$, wenn ZAHL negativ ist.

Tabelle 20. Numerische Funktionen von CLEM (Forts.).

Funktion	Ergebnis	Beschreibung
sign(ZAHL)	Zahl	Wird zur Ermittlung des Vorzeichens von ZAHL verwendet. Diese Operation ergibt -1, 0 oder 1, wenn ZAHL eine ganze Zahl ist. Wenn ZAHL eine reelle Zahl ist, ergibt sich -1,0; 0,0 oder 1,0, je nachdem ob ZAHL negativ, null oder positiv ist.
sqrt(ZAHL)	reelle Zahl	Ergibt die Quadratwurzel von ZAHL. ZAHL muss positiv sein.
sum_n(LISTE)	Zahl	Ergibt die Summe der Werte aus einer Liste numerischer Felder bzw. den Wert null, wenn alle Feldwerte gleich null sind.
mean_n(LISTE)	Zahl	Ergibt den Mittelwert aus einer Liste numerischer Felder bzw. den Wert null, wenn alle Feldwerte gleich null sind.
sdev_n(LISTE)	Zahl	Ergibt die Standardabweichung aus einer Liste numerischer Felder bzw. den Wert null, wenn alle Feldwerte gleich null sind.

Trigonometrische Funktionen

Alle Funktionen in diesem Abschnitt verwenden entweder einen Winkel als Argument oder ergeben einen Winkel als Ergebnis. In beiden Fällen richtet sich die für den Winkel verwendete Einheit (Radianten oder Grad) nach die Einstellung in der entsprechenden Streamoption.

Tabelle 21. Trigonometrische Funktionen von CLEM

Funktion	Ergebnis	Beschreibung
arccos(ZAHL)	reelle Zahl	Berechnet den Arcuscosinus des angegebenen Winkels.
arccosh(ZAHL)	reelle Zahl	Berechnet den hyperbolischen Arcuscosinus des angegebenen Winkels.
arcsin(ZAHL)	reelle Zahl	Berechnet den Arcussinus des angegebenen Winkels.
arcsinh(ZAHL)	reelle Zahl	Berechnet den hyperbolischen Arcussinus des angegebenen Winkels.
arctan(ZAHL)	reelle Zahl	Berechnet den Arcustangens des angegebenen Winkels.
arctan2(ZAHL_Y, ZAHL_X)	reelle Zahl	Berechnet den Arcustangens von ZAHL_Y / ZAHL_X und verwendet die Vorzeichen der beiden Zahlen zur Ableitung von Quadranteninformationen. Das Ergebnis ist eine reelle Zahl im Bereich $-\pi < \text{WINKEL} \leq \pi$ (rad) - $180 < \text{ANGLE} \leq 180$ (deg)
arctanh(ZAHL)	reelle Zahl	Berechnet den hyperbolischen Arcustangens des angegebenen Winkels.
cos(ZAHL)	reelle Zahl	Berechnet den Cosinus des angegebenen Winkels.
cosh(ZAHL)	reelle Zahl	Berechnet den hyperbolischen Cosinus des angegebenen Winkels.
pi	reelle Zahl	Diese Konstante ist die beste reelle Annäherung an Pi.
sin(ZAHL)	reelle Zahl	Berechnet den Sinus des angegebenen Winkels.
sinh(ZAHL)	reelle Zahl	Berechnet den hyperbolischen Sinus des angegebenen Winkels.
tan(ZAHL)	reelle Zahl	Berechnet den Tangens des angegebenen Winkels.
tanh(ZAHL)	reelle Zahl	Berechnet den hyperbolischen Tangens des angegebenen Winkels.

Wahrscheinlichkeitsfunktionen

Wahrscheinlichkeitsfunktionen ergeben Wahrscheinlichkeiten auf der Grundlage verschiedener Verteilungen, z. B. die Wahrscheinlichkeit, dass ein Wert aus der *t*-Verteilung in der Studentenversion kleiner ist als ein bestimmter Wert.

Tabelle 22. CLEM-Wahrscheinlichkeitsfunktionen

Funktion	Ergebnis	Beschreibung
<code>cdf_chisq(ZAHL, DF)</code>	reelle Zahl	Ergibt die Wahrscheinlichkeit, dass ein Wert aus der Chi-Quadrat-Verteilung mit den angegebenen Freiheitsgraden kleiner ist als die angegebene Zahl.
<code>cdf_f(ZAHL, DF1, DF2)</code>	reelle Zahl	Ergibt die Wahrscheinlichkeit, dass ein Wert aus der <i>F</i> -Verteilung mit den angegebenen Freiheitsgraden <i>DF1</i> und <i>DF2</i> kleiner ist als die angegebene Zahl.
<code>cdf_normal(ZAHL, MITTELWERT, STDABW)</code>	reelle Zahl	Ergibt die Wahrscheinlichkeit, dass ein Wert aus der Normalverteilung mit dem angegebenen Mittelwert und der angegebenen Standardabweichung kleiner ist als die angegebene Zahl.
<code>cdf_t(ZAHL, DF)</code>	reelle Zahl	Ergibt die Wahrscheinlichkeit, dass ein Wert aus der Chi-Quadrat-Verteilung aus einer <i>t</i> -Verteilung in der Studentenversion mit den angegebenen Freiheitsgraden kleiner ist als die angegebene Zahl.

Bitweise Operationen mit ganzen Zahlen

Mit dieser Funktion können Ganzzahlen als Bitmuster behandelt werden, die die Zweierkomplementwerte darstellen, bei denen die Bitposition *N* den Wert 2^{*N} aufweist. Diese Operationen funktionieren so, als ob das Vorzeichenbit einer ganzen Zahl endlos nach links erweitert wird. Daher weisen positive ganze Zahlen überall oberhalb des werthöchsten 0-Bits auf, negative ganze Zahlen 1-Bits.

Hinweis: Bitweise Funktionen können nicht über Scripts aufgerufen werden.

Tabelle 23. Bitweise Ganzzahloperationen von CLEM.

Funktion	Ergebnis	Beschreibung
<code>~~ GANZZ1</code>	ganze Zahl	Liefert das bitweise Komplement der ganzen Zahl <i>GANZZ1</i> . Im Ergebnis steht also eine 1 für jede Bitposition, bei der <i>GANZZ1</i> den Wert 0 aufweist. Es gilt immer: <code>~~ GANZZ = -(GANZZ + 1)</code> . Beachten Sie, dass diese Funktion nicht über ein Script aufgerufen werden kann.
<code>GANZZ1 GANZZ2</code>	ganze Zahl	Das Ergebnis dieser Operation ist das bitweise "inklusive Oder" von <i>GANZZ1</i> und <i>GANZZ2</i> . Im Ergebnis steht also eine 1 für jede Bitposition, bei der bei <i>GANZZ1</i> und/oder bei <i>GANZZ2</i> der Wert 1 steht.
<code>GANZZ1 /& GANZZ2</code>	ganze Zahl	Das Ergebnis dieser Operation ist das bitweise "exklusive Oder" von <i>GANZZ1</i> und <i>GANZZ2</i> . Im Ergebnis steht also eine 1 für jede Bitposition, bei der entweder bei <i>GANZZ1</i> oder bei <i>GANZZ2</i> der Wert 1 steht, nicht jedoch bei beiden.
<code>GANZZ1 && GANZZ2</code>	ganze Zahl	Liefert das bitweise "Und" der ganzen Zahlen <i>GANZZ1</i> und <i>GANZZ2</i> . Im Ergebnis steht also eine 1 für jede Bitposition, bei der sowohl bei <i>GANZZ1</i> als auch bei <i>GANZZ2</i> der Wert 1 steht.
<code>GANZZ1 &&~~ GANZZ2</code>	ganze Zahl	Liefert das bitweise "Und" von <i>GANZZ1</i> und das bitweise Komplement von <i>GANZZ2</i> . Im Ergebnis steht also eine 1 für jede Bitposition, bei der bei <i>GANZZ1</i> der Wert 1 und bei <i>GANZZ2</i> der Wert 0 steht. Diese Operation ist identisch mit <code>GANZZ1 && (~~GANZZ2)</code> und ist sinnvoll zum Bereinigen von in <i>GANZZ2</i> gesetzten Bits von <i>GANZZ1</i> .
<code>GANZZ << N</code>	ganze Zahl	Ergibt das Bitmuster von <i>GANZZ1</i> , um <i>N</i> Positionen nach links verschoben. Ein negativer Wert von <i>N</i> führt zu einer Verschiebung nach rechts.

Tabelle 23. Bitweise Ganzzahloperationen von CLEM (Forts.).

Funktion	Ergebnis	Beschreibung
<code>GANZZ >> N</code>	<i>ganze Zahl</i>	Ergibt das Bitmuster von <code>GANZZ1</code> , um <code>N</code> Positionen nach rechts verschoben. Ein negativer Wert von <code>N</code> führt zu einer Verschiebung nach links.
<code>GANZZ1 &&=_0 GANZZ2</code>	<i>boolesch</i>	Entspricht dem booleschen Ausdruck <code>GANZZ1 && GANZZ2 /!= 0</code> , ist jedoch effizienter.
<code>GANZZ1 &&/=_0 GANZZ2</code>	<i>boolesch</i>	Entspricht dem booleschen Ausdruck <code>GANZZ1 && GANZZ2 == 0</code> , ist jedoch effizienter.
<code>integer_bitcount(GANZZ)</code>	<i>ganze Zahl</i>	Zählt die Anzahl der 1- oder 0-Bits in der Zweierkomplementdarstellung von <code>GANZZ</code> . Wenn <code>GANZZ</code> nicht negativ ist, ist <code>N</code> die Anzahl der 1-Bits. Wenn <code>GANZZ</code> negativ ist, gibt dieser Wert die Anzahl der 0-Bits an. Aufgrund der Vorzeichenerweiterung gibt es unendlich viele 0-Bits in nicht negativen ganzen Zahlen bzw. 1-Bits in negativen ganzen Zahlen. Es gilt immer: <code>integer_bitcount(GANZZ) = integer_bitcount(-(GANZZ+1))</code> .
<code>integer_leastbit(GANZZ)</code>	<i>ganze Zahl</i>	Gibt die Bitposition <code>N</code> des niedrigstwertigen Bits an, das in der ganzen Zahl <code>GANZZ</code> gesetzt wurde. <code>N</code> ist die höchste Potenz von 2, durch die sich <code>GANZZ</code> ohne Rest teilen lässt.
<code>integer_length(GANZZ)</code>	<i>ganze Zahl</i>	Ergibt die Länge von <code>GANZZ</code> als ganze Zweierkomplementzahl in Bit. <code>N</code> ist also die kleinste ganze Zahl, sodass gilt: <code>GANZZ < (1 << N)</code> , wenn <code>GANZZ >= 0</code> . <code>GANZZ >= (-1 << N)</code> , wenn <code>GANZZ < 0</code> . Wenn <code>GANZZ</code> nicht negativ ist, ist für die Darstellung von <code>GANZZ</code> als ganze Zahl ohne Vorzeichen ein Feld mit mindestens <code>N</code> Bit erforderlich. Alternativ sind, unabhängig vom Vorzeichen, mindestens <code>N+1</code> Bit erforderlich, um <code>GANZZ</code> als ganze Zahl mit Vorzeichen darzustellen.
<code>testbit(GANZZ, N)</code>	<i>boolesch</i>	Testet das Bit an Position <code>N</code> in der ganzen Zahl <code>GANZZ</code> und ergibt den Status von Bit <code>N</code> als booleschen Wert, der bei 1 "wahr" und bei 0 "falsch" ist.

Zufallsfunktionen

Folgende Funktionen dienen zur zufälligen Auswahl von Elementen oder zur Generierung von Zufallszahlen.

Tabelle 24. CLEM-Zufallsfunktionen.

Funktion	Ergebnis	Beschreibung
<code>oneof(LISTE)</code>	<i>beliebig</i>	Ergibt ein zufällig ausgewähltes Element von <code>LISTE</code> . Die Listenelemente sollten wie folgt eingegeben werden <code>[ELEMENT1, ELEMENT2, ..., ELEMENT_N]</code> . Auch eine Liste mit Feldnamen kann angegeben werden.
<code>random(ZAHL)</code>	<i>Zahl</i>	Ergibt eine gleichmäßig verteilte Zufallszahl desselben Typs (<code>GANZZ</code> oder <code>REELL</code>). Es gilt der Bereich von 1 bis <code>ZAHL</code> . Bei Verwendung einer ganzen Zahl ergeben sich nur ganze Zahlen. Bei Verwendung einer reellen Zahl (Dezimalzahl), ergeben sich reelle Zahlen (die Anzahl der Dezimalstellen werden in den Streamoptionen festgelegt). Die größte Zufallszahl, die die Funktion ergibt, kann gleich <code>ZAHL</code> sein.
<code>random0(ZAHL)</code>	<i>Zahl</i>	Diese Operation hat dieselben Eigenschaften wie <code>random(ZAHL)</code> , der Bereich beginnt jedoch bei 0. Die größte Zufallszahl, die die Funktion ergibt, ist nie gleich <code>ZAHL</code> .

Zeichenfolgefunktionen

In CLEM können Sie folgende Operationen an Zeichenfolgen durchführen:

- Vergleichen von Zeichenfolgen
- Erstellen von Zeichenfolgen
- Zugreifen auf Zeichen

In CLEM-Ausdrücken ist eine Zeichenfolge jede Folge von Zeichen zwischen zusammengehörenden (doppelten) Anführungszeichen ("Zeichenfolge in Anführungszeichen"). Zeichen (ZEICHEN) können beliebige alphanumerische Zeichen sein. Sie werden in CLEM-Ausdrücken mithilfe einfacher umgekehrter Anführungsstriche in der Form ``<Zeichen>`` deklariert, wie beispielsweise ``z``, ``A`` oder ``2``. Zeichen, die außerhalb des zulässigen Bereichs liegen, oder negative Indizes bei einer Zeichenfolge führen zu einem undefinierten Verhalten.

Anmerkung. Vergleiche zwischen Zeichenfolgen, die SQL-Pushback verwenden oder nicht, können unterschiedliche Ergebnisse generieren, wenn nachfolgende Leerzeichen vorhanden sind.

Table 25. CLEM-Zeichenfolgefunktionen.

Funktion	Ergebnis	Beschreibung
<code>allbutfirst(N, ZEICHENFOLGE)</code>	<i>Zeichenfolge</i>	Ergibt eine Zeichenfolge, bei der es sich um <code>ZEICHENFOLGE</code> nach der Entfernung der ersten <code>N</code> Zeichen handelt.
<code>allbutlast(N, ZEICHENFOLGE)</code>	<i>Zeichenfolge</i>	Ergibt eine Zeichenfolge, bei der es sich um <code>ZEICHENFOLGE</code> nach der Entfernung der letzten <code>N</code> Zeichen handelt.
<code>alphabefore(ZEICHENFOLGE1, ZEICHENFOLGE2)</code>	<i>boolesch</i>	Dient zur Überprüfung der alphabetischen Sortierung von Zeichenfolgen. Ergibt den Wert "Wahr", wenn <code>ZEICHENFOLGE1</code> der <code>ZEICHENFOLGE2</code> vorausgeht.
<code>endstring(LÄNGE, ZEICHENFOLGE)</code>	<i>Zeichenfolge</i>	Extrahiert die letzten <code>N</code> Zeichen aus der angegebenen Zeichenfolge. Ist eine Zeichenfolge kleiner oder gleich der angegebenen Länge, dann bleibt die Zeichenfolge unverändert.
<code>hasendstring(ZEICHENFOLGE, TEILZEICHENFOLGE)</code>	<i>ganze Zahl</i>	Diese Funktion entspricht <code>isendstring(TEILZEICHENFOLGE, ZEICHENFOLGE)</code> .
<code>hasmidstring(ZEICHENFOLGE, TEILZEICHENFOLGE)</code>	<i>ganze Zahl</i>	Diese Funktion entspricht <code>ismidstring(TEILZEICHENFOLGE, ZEICHENFOLGE)</code> (eingebettete Teilzeichenfolge).
<code>hasstartstring(ZEICHENFOLGE, TEILZEICHENFOLGE)</code>	<i>ganze Zahl</i>	Diese Funktion entspricht <code>isstartstring(TEILZEICHENFOLGE, ZEICHENFOLGE)</code> .
<code>hassubstring(ZEICHENFOLGE, N, TEILZEICHENFOLGE)</code>	<i>ganze Zahl</i>	Diese Funktion entspricht <code>issubstring(TEILZEICHENFOLGE, N, ZEICHENFOLGE)</code> . Dabei hat <code>N</code> standardmäßig den Wert 1.
<code>count_substring(ZEICHENFOLGE, TEILZEICHENFOLGE)</code>	<i>ganze Zahl</i>	Ergibt den Wert, wie oft die angegebene Teilzeichenfolge in der Zeichenfolge auftritt. Beispiel: <code>count_substring("foooo.txt", "oo")</code> ergibt den Wert 3.
<code>hassubstring(ZEICHENFOLGE, TEILZEICHENFOLGE)</code>	<i>ganze Zahl</i>	Diese Funktion entspricht <code>issubstring(TEILZEICHENFOLGE, 1, ZEICHENFOLGE)</code> . Dabei hat <code>N</code> standardmäßig den Wert 1.

Tabelle 25. CLEM-Zeichenfolgefunktionen (Forts.).

Funktion	Ergebnis	Beschreibung
isalphacode(ZEICHEN)	<i>boolesch</i>	Ergibt den Wert "Wahr", wenn ZEICHEN ein Zeichen in der angegebenen Zeichenfolge (häufig ein Feldname) ist, dessen Zeichencode ein Buchstabe ist. Andernfalls ergibt die Funktion den Wert 0. Beispiel: isalphacode(produce_num(1)).
isendstring(TEILZEICHENFOLGE, ZEICHENFOLGE)	<i>ganze Zahl</i>	Wenn die Zeichenfolge ZEICHENFOLGE mit der Teilzeichenfolge TEILZEICHENFOLGE endet, ergibt diese Funktion den ganzzahligen Index von TEILZEICHENFOLGE in ZEICHENFOLGE. Andernfalls ergibt die Funktion den Wert 0.
islowercode(ZEICHEN)	<i>boolesch</i>	Ergibt den Wert "Wahr", wenn ZEICHEN ein Kleinbuchstabe der angegebenen Zeichenfolge (häufig ein Feldname) ist. Andernfalls ergibt die Funktion den Wert 0. Beispielsweise sind sowohl islowercode(``) als auch islowercode(country_name(2)) gültige Ausdrücke.
ismidstring(TEILZEICHENFOLGE, ZEICHENFOLGE)	<i>ganze Zahl</i>	Wenn TEILZEICHENFOLGE eine Teilzeichenfolge von ZEICHENFOLGE ist, jedoch nicht beim ersten Zeichen von ZEICHENFOLGE beginnt oder mit dem letzten Zeichen endet, ergibt diese Funktion den Index, bei dem die Teilzeichenfolge beginnt. Andernfalls ergibt die Funktion den Wert 0.
isnumbercode(ZEICHEN)	<i>boolesch</i>	Ergibt den Wert "Wahr", wenn ZEICHEN bei der angegebenen Zeichenfolge (häufig ein Feldname) ein Zeichen ist, dessen Zeichencode eine Ziffer ist. Andernfalls ergibt die Funktion den Wert 0. Beispiel: isnumbercode(product_id(2)).
isstartstring(TEILZEICHENFOLGE, ZEICHENFOLGE)	<i>ganze Zahl</i>	Wenn die Zeichenfolge ZEICHENFOLGE mit der Teilzeichenfolge TEILZEICHENFOLGE beginnt, ergibt diese Funktion den Index 1. Andernfalls ergibt die Funktion den Wert 0.
issubstring(TEILZEICHENFOLGE, N, ZEICHENFOLGE)	<i>ganze Zahl</i>	Durchsucht die Zeichenfolge ZEICHENFOLGE beginnend beim N. Zeichen nach einer Teilzeichenfolge, die mit TEILZEICHENFOLGE übereinstimmt. Wenn die Teilzeichenfolge gefunden wird, ergibt diese Funktion den Index, bei dem die Teilzeichenfolge beginnt. Andernfalls ergibt diese Funktion den Wert 0. Wenn N nicht angegeben ist, wird standardmäßig der Wert 1 verwendet.
issubstring(TEILZEICHENFOLGE, ZEICHENFOLGE)	<i>ganze Zahl</i>	Durchsucht die Zeichenfolge ZEICHENFOLGE beginnend beim N. Zeichen nach einer Teilzeichenfolge, die mit TEILZEICHENFOLGE übereinstimmt. Wenn die Teilzeichenfolge gefunden wird, ergibt diese Funktion den Index, bei dem die Teilzeichenfolge beginnt. Andernfalls ergibt diese Funktion den Wert 0. Wenn N nicht angegeben ist, wird standardmäßig der Wert 1 verwendet.

Tabelle 25. CLEM-Zeichenfolgefunktionen (Forts.).

Funktion	Ergebnis	Beschreibung
<code>issubstring_count</code> (TEILZEICHENFOLGE, N, ZEICHENFOLGE):	<i>ganze Zahl</i>	Ergibt den Index für das <i>N</i> -te Vorkommen von <i>TEILZEICHENFOLGE</i> in der angegebenen <i>ZEICHENFOLGE</i> . Falls <i>TEILZEICHENFOLGE</i> weniger als <i>N</i> -mal auftritt, ergibt sich der Wert 0.
<code>issubstring_lim</code> (TEILZEICHENFOLGE, N, STARTLIM, ENDLIM, ZEICHENFOLGE)	<i>ganze Zahl</i>	Diese Funktion entspricht <code>issubstring</code> , der Treffer muss jedoch beim oder vor dem Index <i>STARTLIM</i> beginnen und beim oder vor dem Index <i>ENDLIM</i> enden. Die Beschränkungen <i>STARTLIM</i> bzw. <i>ENDLIM</i> können inaktiviert werden, indem ein Falsch-Wert für eines der Argumente angegeben wird. So ist beispielsweise <code>issubstring_lim</code> (TEILZEICHENFOLGE, N, false, false, ZEICHENFOLGE) mit <code>issubstring</code> identisch.
<code>isuppercode</code> (ZEICHEN)	<i>boolesch</i>	Ergibt den Wert "Wahr", wenn <i>ZEICHEN</i> ein Großbuchstabe ist. Andernfalls ergibt die Funktion den Wert 0. Beispielsweise sind sowohl <code>isuppercode(`)`)</code> als auch <code>isuppercode(country_name(2))</code> gültige Ausdrücke.
<code>last</code> (ZEICHEN)	<i>Zeichenfolge</i>	Ergibt das letzte Zeichen <i>ZEICHEN</i> von <i>ZEICHENFOLGE</i> (muss mindestens ein Zeichen lang sein).
<code>length</code> (ZEICHENFOLGE)	<i>ganze Zahl</i>	Ergibt die Länge der Zeichenfolge <i>ZEICHENFOLGE</i> , d. h. die Anzahl der Zeichen in der Zeichenfolge.
<code>locchar</code> (ZEICHEN, N, ZEICHENFOLGE)	<i>ganze Zahl</i>	Mit dieser Funktion kann die Position der Zeichen in symbolischen Feldern ermittelt werden. Mit dieser Funktion wird die Zeichenfolge <i>ZEICHENFOLGE</i> nach dem Zeichen <i>ZEICHEN</i> durchsucht, wobei der Suchvorgang beim <i>N</i> . Zeichen von <i>ZEICHENFOLGE</i> gestartet wird. Diese Funktion ergibt einen Wert (mit Startwert <i>N</i>), der die Position angibt, an der das Zeichen gefunden wurde. Wenn das Zeichen nicht gefunden wurde, ergibt die Funktion den Wert 0. Wenn die Funktion ein ungültiges Offset (<i>N</i>) aufweist (z. B. ein Offset, das außerhalb der Länge der Zeichenfolge liegt), ergibt die Funktion <code>\$null\$</code> . Beispiel: <code>locchar(`n`, 2, Webseite)</code> durchsucht das Feld <i>Webseite</i> nach dem Zeichen <code>`n`</code> und beginnt mit der Suche beim zweiten Zeichen im Feldwert. <i>Hinweis:</i> Achten Sie darauf, das angegebene Zeichen in einzelne umgekehrte Anführungszeichen einzuschließen.

Tabella 25. CLEM-Zeichenfolgefunktionen (Forts.).

Funktion	Ergebnis	Beschreibung
locchar_back(ZEICHEN, N, ZEICHENFOLGE)	ganze Zahl	Ähnlich wie locchar, außer dass die Suche ausgehend vom N-ten Zeichen rückwärts durchgeführt wird. Beispiel: occhar_back(`n`, 9, Webseite) durchsucht das Feld <i>Webseite</i> ausgehend vom 9. Zeichen rückwärts bis zum Anfang der Zeichenfolge. Wenn die Funktion ein ungültiges Offset aufweist, (z. B. ein Offset, das außerhalb der Länge der Zeichenfolge liegt), ergibt die Funktion \$null\$. Idealerweise sollte locchar_back in Verbindung mit der Funktion length(<Field>) verwendet werden, um die Länge des aktuellen Werts des Felds dynamisch zu verwenden. Beispiel: locchar_back(`n`, (length(Webseite)), Webseite).
lowertoupper (ZEICHEN) lowertoupper (ZEICHENFOLGE)	ZEICHEN oder Zeichenfolge	Die Eingabe kann entweder eine Zeichenfolge oder ein Zeichen sein. Diese wird in dieser Funktion verwendet, um ein neues Element desselben Typs auszugeben, wobei alle Kleinbuchstaben in ie entsprechenden Großbuchstaben konvertiert wurden. Beispiele für gültige Ausdrücke sind: lowertoupper(`a`), lowertoupper("Meine Zeichenfolge") und lowertoupper(feldname(2)).
matches	boolesch	Ergibt den Wert "wahr", wenn eine Zeichenfolge mit einem bestimmten Muster übereinstimmt. Das Muster muss aus einer Literalzeichenfolge bestehen; der Name eines Felds, das ein Muster enthält, ist nicht zulässig. Das Fragezeichen (?) als Platzhalterzeichen kann die Stelle von genau einem beliebigen Zeichen einnehmen, der Stern (*) entspricht keinem, einem oder mehreren Zeichen. Soll ein Fragezeichen oder ein Stern nicht als Platzhalter fungieren, sondern bei der Suche berücksichtigt werden, geben Sie den umgekehrten Schrägstrich (\) als Escapezeichen ein.
replace(TEILZEICHENFOLGE, NEUE_TEILZEICHENFOLGE, ZEICHENFOLGE)	Zeichenfolge	Innerhalb der angegebenen ZEICHENFOLGE werden alle Instanzen von TEILZEICHENFOLGE durch NEUE_TEILZEICHENFOLGE ersetzt.
replicate(ANZAHL, ZEICHENFOLGE)	Zeichenfolge	Ergibt eine Zeichenfolge, die aus der ursprünglichen Zeichenfolge besteht, die so oft wie angegeben kopiert wurde.
stripchar(ZEICHEN, ZEICHENFOLGE)	Zeichenfolge	Mit dieser Funktion können Sie die angegebenen Zeichen aus einer Zeichenfolge oder einem Feld entfernen. Mit dieser Funktion können Sie beispielsweise zusätzliche Symbole, wie Währungsangaben, aus den Daten entfernen, um eine einfache Zahl oder einen einfachen Namen zu erhalten. Mit der Syntax stripchar(`€`, 'Kosten') beispielsweise erhalten Sie ein neues Feld, bei dem das Euro-Zeichen aus allen Werten entfernt ist. <i>Hinweis:</i> Achten Sie darauf, das angegebene Zeichen in einzelne umgekehrte Anführungszeichen einzuschließen.

Tabelle 25. CLEM-Zeichenfolgefunktionen (Forts.).

Funktion	Ergebnis	Beschreibung
skipchar(ZEICHEN, N, ZEICHENFOLGE)	ganze Zahl	Durchsucht die Zeichenfolge <i>ZEICHENFOLGE</i> nach allen Zeichen außer <i>ZEICHEN</i> und beginnt dabei beim <i>N</i> . Zeichen. Diese Funktion ergibt eine ganzzahlige Teilzeichenfolge, die den Punkt angibt, an dem ein solches Zeichen gefunden wurde, oder 0, wenn es sich bei jedem Zeichen vom <i>N</i> . Zeichen an um <i>ZEICHEN</i> handelt. Wenn die Funktion ein ungültiges Offset aufweist, (z. B. ein Offset, das außerhalb der Länge der Zeichenfolge liegt), ergibt die Funktion <code>\$null\$</code> . <code>locchar</code> wird häufig in Verbindung mit den Funktionen vom Typ <code>skipchar</code> verwendet, um den Wert von <i>N</i> (die Stelle, an der mit der Suche begonnen werden soll) zu bestimmen. Beispiel: <code>skipchar(`s`, (locchar(`s`, 1, "MeinString")), "MeinString")</code> .
skipchar_back(ZEICHEN, N, ZEICHENFOLGE)	ganze Zahl	Ähnlich wie <code>skipchar</code> , außer dass die Suche ausgehend vom <i>N</i> -ten Zeichen rückwärts durchgeführt wird.
startstring(LÄNGE, ZEICHENFOLGE)	Zeichenfolge	Extrahiert die ersten <i>N</i> Zeichen aus der angegebenen Zeichenfolge. Ist eine Zeichenfolge kleiner oder gleich der angegebenen Länge, dann bleibt die Zeichenfolge unverändert.
strmember(ZEICHEN, ZEICHENFOLGE)	ganze Zahl	Entspricht <code>locchar(ZEICHEN, 1, ZEICHENFOLGE)</code> . Ergibt eine ganzzahlige Teilzeichenfolge, die entweder den Punkt angibt, an dem <i>CHAR</i> zum ersten Mal auftritt, oder den Wert 0. Wenn die Funktion ein ungültiges Offset aufweist (z. B. ein Offset, das außerhalb der Länge der Zeichenfolge liegt), ergibt diese Funktion <code>\$null\$</code> .
subscrs(N, ZEICHENFOLGE)	ZEICHEN	Ergibt das <i>N</i> . Zeichen <i>ZEICHEN</i> der Eingabezeichenfolge <i>ZEICHENFOLGE</i> . Diese Funktion kann auch in einer abgekürzten Form geschrieben werden, nämlich als <code>ZEICHENFOLGE(N)</code> . Der Ausdruck <code>lowertoupper("name"(1))</code> ist beispielsweise ein gültiger Ausdruck.
substring(N, LÄNGE, ZEICHENFOLGE)	Zeichenfolge	Ergibt die Zeichenfolge <i>TEILZEICHENFOLGE</i> , die aus den <i>LÄNGE</i> Zeichen der Zeichenfolge <i>ZEICHENFOLGE</i> besteht, wobei mit dem Zeichen bei Index <i>N</i> begonnen wird.
substring_between(N1, N2, ZEICHENFOLGE)	Zeichenfolge	Ergibt die Teilzeichenfolge von <i>ZEICHENFOLGE</i> , die bei Index <i>N1</i> beginnt und bei Index <i>N2</i> endet.
trim(ZEICHENFOLGE)	Zeichenfolge	Entfernt führende und nachgestellte Leerzeichen aus der angegebenen Zeichenfolge.
trim_start(ZEICHENFOLGE)	Zeichenfolge	Entfernt führende Leerzeichen aus der angegebenen Zeichenfolge.
trimend(ZEICHENFOLGE)	Zeichenfolge	Entfernt nachgestellte Leerzeichen aus der angegebenen Zeichenfolge.
unicode_char(ZAHL)	ZEICHEN	Ergibt das Zeichen mit dem Unicode-Wert von <i>ZAHL</i> .

Tabelle 25. CLEM-Zeichenfolgefunktionen (Forts.).

Funktion	Ergebnis	Beschreibung
unicode_value(ZEICHEN)	ZAHL	Ergibt den Unicode-Wert für das Zeichen ZEICHEN.
uppertolower (ZEICHEN) uppertolower (ZEICHENFOLGE)	ZEICHEN oder Zeichenfolge	Die Eingabe kann entweder eine Zeichenfolge oder ein Zeichen sein. Diese wird in dieser Funktion verwendet, um ein neues Element desselben Typs auszugeben, wobei alle Großbuchstaben in die entsprechenden Kleinbuchstaben konvertiert wurden. <i>Hinweis:</i> Achten Sie darauf, Zeichenfolgen in doppelte Anführungszeichen und Zeichen in einzelne, umgekehrte Anführungszeichen zu setzen. Einzelne Feldnamen sollten ohne Anführungszeichen angegeben werden.

SoundEx-Funktionen

SoundEx ist eine Methode zum Auffinden von Zeichenfolgen, deren Klang bekannt ist, nicht jedoch deren genaue Schreibweise. Die Methode wurde im Jahr 1918 entwickelt. Hiermit werden Wörter mit ähnlichen Lauten ermittelt und zwar auf der Grundlage phonetischer Annahmen, wie bestimmte Buchstaben ausgesprochen werden. Suchen Sie beispielsweise Namen in einer Datenbank, bei denen die Schreibweise und die Betonung ähnlicher Namen voneinander abweichen können. Der grundlegende SoundEx-Algorithmus ist in zahlreichen Schriften dokumentiert. Trotz bekannter Einschränkungen (führende Buchstaben und Buchstabenfolgen wie ph und f gelten beispielsweise nicht als Übereinstimmung, obwohl sie gleich klingen) wird diese Methode in den meisten Datenbanken in irgendeiner Form genutzt.

Tabelle 26. Soundex-Funktionen von CLEM

Funktion	Ergebnis	Beschreibung
soundex(ZEICHENFOLGE)	ganze Zahl	Ergibt den vierbuchstabigen SoundEx-Code für die angegebene ZEICHENFOLGE.
soundex_difference(ZEICHENFOLGE1, ZEICHENFOLGE2)	ganze Zahl	Ergibt eine ganze Zahl zwischen 0 und 4, aus der die Anzahl der Zeichen hervorgeht, die im SoundEx-Code für die beiden Zeichenfolgen übereinstimmen. Der Wert 0 bedeutet dabei, dass gar keine Ähnlichkeit zwischen den Zeichenfolgen vorliegt, der Wert 4 weist auf sehr ähnliche oder sogar identische Zeichenfolgen hin.

Datums- und Uhrzeitfunktionen

CLEM beinhaltet eine Gruppe von Funktionen zum Umgang mit Feldern mit dem Speichertyp "Datum/ Uhrzeit" von Zeichenfolgevariablen, die für Datums- und Zeitangaben stehen. Die Formate für Datum und Uhrzeit sind für jeden Stream spezifisch und werden im Dialogfeld "Streameigenschaften" angegeben. Die Datums- und Zeitfunktionen analysieren Datums- und Zeitzeichenfolgen gemäß dem aktuell ausgewählten Format.

Wenn Sie ein Jahr in einem Datumsformat angeben, bei dem nur zwei Ziffern verwendet werden (d. h., das Jahrhundert ist nicht angegeben), verwendet IBM SPSS Modeler das im Dialogfeld "Streameigenschaften" angegebene Standardjahrhundert.

Hinweis: Datums- und Zeitfunktionen können nicht über Scripts aufgerufen werden.

Tabelle 27. CLEM-Datums- und Uhrzeitfunktionen.

Funktion	Ergebnis	Beschreibung
@TODAY	Zeichenfolge	Bei Auswahl von Tage/Minuten übertragen im Dialogfeld "Streameigenschaften" ergibt diese Funktion das aktuelle Datum als Zeichenfolge im aktuellen Datumsformat. Wenn Sie ein zweistelliges Datumsformat verwenden und nicht die Option Tage/Minuten übertragen verwenden, ergibt diese Funktion \$null\$ auf dem aktuellen Server. Beachten Sie, dass diese Funktion nicht über ein Script aufgerufen werden kann.
to_time(ELEMENT)	Zeit	Konvertiert den Speichertyp des angegebenen Felds in eine Zeit.
to_date(ELEMENT)	Datum	Konvertiert den Speichertyp des angegebenen Felds in ein Datum.
to_timestamp(ELEMENT)	Zeitmarke	Konvertiert den Speichertyp des angegebenen Felds in eine Zeitmarke.
to_datetime(ELEMENT)	Datum/Uhrzeit	Konvertiert den Speichertyp des angegebenen Felds in einen Wert vom Typ "Datum", "Zeit" oder "Zeitmarke".
datetime_date(ELEMENT)	Datum	Gibt den Datumswert für eine <i>Zahl</i> , eine <i>Zeichenfolge</i> oder <i>Zeitmarke</i> zurück. Beachten Sie, dass dies die einzige Funktion ist, mit der Sie eine Zahl (in wenigen Sekunden) wieder in ein Datum zurückkonvertieren können. Wenn es sich bei ELEMENT um eine Zeichenfolge handelt, wird ein Datum erstellt, indem eine Zeichenfolge im aktuellen Datumsformat analysiert wird. Das im Dialogfeld für die Streameigenschaften angegebene Datumsformat muss korrekt sein, damit diese Funktion erfolgreich ausgeführt werden kann. Wenn es sich bei ELEMENT um eine Zahl handelt, wird sie als Anzahl von Sekunden seit dem Basisdatum (Epoche) interpretiert. Die Bruchteile eines Tages werden gekürzt. Wenn es sich bei ELEMENT um eine Zeitmarke handelt, wird der Datumsteil der Zeitmarke ausgegeben. Wenn es sich bei ELEMENT um ein Datum handelt, wird es unverändert ausgegeben.
date_before(DATUM1, DATUM2)	boolesch	Ergibt den Wert "Wahr", wenn <i>DATUM1</i> ein Datum oder eine Zeitmarke darstellt, das/der vor dem durch <i>DATUM2</i> dargestellten Datum liegt. Andernfalls ergibt die Funktion den Wert 0.
date_days_difference(DATUM1, DATUM2)	ganze Zahl	Ergibt die Zeit in Tagen (als ganze Zahl), die zwischen dem durch <i>DATUM1</i> und dem durch <i>DATUM2</i> angegebenen Datum bzw. der Zeitmarke liegt. Wenn <i>DATUM2</i> vor <i>DATUM1</i> liegt, ergibt diese Funktion eine negative Zahl.
date_in_days(DATUM)	ganze Zahl	Ergibt die Zeit in Tagen (als ganze Zahl), die zwischen dem Basisdatum und dem durch <i>DATUM</i> angegebenen Datum oder der Zeitmarke liegt. Wenn <i>DATUM</i> vor dem Basisdatum liegt, ergibt diese Funktion eine negative Zahl. Es muss ein gültiges Datum angegeben werden, damit die Berechnung ordnungsgemäß erfolgen kann. Beispielsweise sollten Sie nicht den 29. Februar 2001 als Datum angeben. Da 2001 kein Schaltjahr war, gibt es dieses Datum nicht.

Tabelle 27. CLEM-Datums- und Uhrzeitfunktionen (Forts.).

Funktion	Ergebnis	Beschreibung
date_in_months(DATUM)	reelle Zahl	Ergibt die Zeit in Monaten (als reelle Zahl), die zwischen dem Basisdatum und dem durch <i>DATUM</i> angegebenen Datum oder der Zeitmarke liegt. Dies ist ein Näherungswert, der auf einem Monat mit 30.4375 Tagen beruht. Wenn <i>DATUM</i> vor dem Basisdatum liegt, ergibt diese Funktion eine negative Zahl. Es muss ein gültiges Datum angegeben werden, damit die Berechnung ordnungsgemäß erfolgen kann. Beispielsweise sollten Sie nicht den 29. Februar 2001 als Datum angeben. Da 2001 kein Schaltjahr war, gibt es dieses Datum nicht.
date_in_weeks(DATUM)	reelle Zahl	Ergibt die Zeit in Wochen (als reelle Zahl), die zwischen dem Basisdatum und dem durch <i>DATUM</i> angegebenen Datum oder der Zeitmarke liegt. Der Wert basiert auf einer Woche mit 7,0 Tagen. Wenn <i>DATUM</i> vor dem Basisdatum liegt, ergibt diese Funktion eine negative Zahl. Es muss ein gültiges Datum angegeben werden, damit die Berechnung ordnungsgemäß erfolgen kann. Beispielsweise sollten Sie nicht den 29. Februar 2001 als Datum angeben. Da 2001 kein Schaltjahr war, gibt es dieses Datum nicht.
date_in_years(DATUM)	reelle Zahl	Ergibt die Zeit in Jahren (als reelle Zahl), die zwischen dem Basisdatum und dem durch <i>DATUM</i> angegebenen Datum oder der Zeitmarke liegt. Dies ist ein Näherungswert, der auf einem Jahr mit 365.25 Tagen beruht. Wenn <i>DATUM</i> vor dem Basisdatum liegt, ergibt diese Funktion eine negative Zahl. Es muss ein gültiges Datum angegeben werden, damit die Berechnung ordnungsgemäß erfolgen kann. Beispielsweise sollten Sie nicht den 29. Februar 2001 als Datum angeben. Da 2001 kein Schaltjahr war, gibt es dieses Datum nicht.
date_months_difference (DATUM1, DATUM2)	reelle Zahl	Ergibt die Zeit in Monaten (als reelle Zahl), die zwischen dem durch <i>DATUM1</i> und dem durch <i>DATUM2</i> angegebenen Datum bzw. der Zeitmarke liegt. Dies ist ein Näherungswert, der auf einem Monat mit 30.4375 Tagen beruht. Wenn <i>DATUM2</i> vor <i>DATUM1</i> liegt, ergibt diese Funktion eine negative Zahl.
datetime_date(JAHR, MONAT, TAG)	Datum	Erstellt einen Datumswert für die angegebenen Werte für <i>JAHR</i> , <i>MONAT</i> und <i>TAG</i> . Bei den Argumenten muss es sich um ganze Zahlen handeln.
datetime_day(DATUM)	ganze Zahl	Ergibt den Tag im Monat ausgehend von einem angegebenen <i>DATUM</i> oder einer Zeitmarke. Das Ergebnis ist eine ganze Zahl im Bereich von 1 bis 31.
datetime_day_name(TAG)	Zeichenfolge	Ergibt den vollständigen Namen für den angegebenen <i>TAG</i> . Das Argument muss eine ganze Zahl im Bereich von 1 (Sonntag) bis 7 (Samstag) sein.
datetime_hour(ZEIT)	ganze Zahl	Ergibt die Stunde aus einer <i>ZEIT</i> oder einer Zeitmarke. Das Ergebnis ist eine ganze Zahl im Bereich von 0 bis 23.
datetime_in_seconds(ZEIT)	reelle Zahl	Ergibt den in <i>ZEIT</i> gespeicherten Sekundenabschnitt.
datetime_in_seconds(DATUM), datetime_in_seconds(DATUMZEIT)	reelle Zahl	Ergibt die aus der Differenz zwischen dem aktuellen <i>DATUM</i> oder <i>DATUM/UHRZEIT</i> und dem Basisdatum (1900-01-01) berechneten Gesamtwert (umgewandelt in Sekunden).
datetime_minute(ZEIT)	ganze Zahl	Ergibt die Minute aus einer <i>ZEIT</i> oder einer Zeitmarke. Das Ergebnis ist eine ganze Zahl im Bereich von 0 bis 59.

Tabelle 27. CLEM-Datums- und Uhrzeitfunktionen (Forts.).

Funktion	Ergebnis	Beschreibung
datetime_month(DATUM)	ganze Zahl	Ergibt den Monat aus einem <i>DATUM</i> oder einer Zeitmarke. Das Ergebnis ist eine ganze Zahl im Bereich von 1 bis 12.
datetime_month_name (MONAT)	Zeichenfolge	Ergibt den vollständigen Namen für den angegebenen <i>MONAT</i> . Das Argument muss eine ganze Zahl im Bereich von 1 bis 12 sein.
datetime_now	Zeitmarke	Ergibt die aktuelle Zeit als Zeitmarke.
datetime_second(ZEIT)	ganze Zahl	Ergibt die Sekunde aus einer <i>ZEIT</i> oder einer Zeitmarke. Das Ergebnis ist eine ganze Zahl im Bereich von 0 bis 59.
datetime_day_short_name (TAG)	Zeichenfolge	Ergibt den abgekürzten Namen für den angegebenen <i>TAG</i> . Das Argument muss eine ganze Zahl im Bereich von 1 (Sonntag) bis 7 (Samstag) sein.
datetime_month_short_name (MONAT)	Zeichenfolge	Ergibt den abgekürzten Namen für den angegebenen <i>MONAT</i> . Das Argument muss eine ganze Zahl im Bereich von 1 bis 12 sein.
datetime_time(STUNDE, MINUTE, SEKUNDE)	Zeit	Ergibt den Zeitwert für die angegebene <i>STUNDE</i> , <i>MINUTE</i> und <i>SEKUNDE</i> . Bei den Argumenten muss es sich um ganze Zahlen handeln.
datetime_time(ELEMENT)	Zeit	Ergibt den Zeitwert für das angegebene <i>ELEMENT</i> .
datetime_timestamp(JAHR, MONAT, TAG, STUNDE, MINUTE, SEKUNDE)	Zeitmarke	Ergibt den Zeitmarkenwert für die angegebenen Werte für <i>JAHR</i> , <i>MONAT</i> , <i>TAG</i> , <i>STUNDE</i> , <i>MINUTE</i> und <i>SEKUNDE</i> .
datetime_timestamp(DATUM, ZEIT)	Zeitmarke	Ergibt den Zeitmarkenwert für das angegebene <i>DATUM</i> und die angegebene <i>ZEIT</i> .
datetime_timestamp (ZAHL)	Zeitmarke	Ergibt den Zeitmarkenwert für die angegebene Anzahl von Sekunden.
datetime_weekday(DATUM)	ganze Zahl	Ergibt den Tag in der Woche <i>t</i> ausgehend von einem angegebenen <i>DATUM</i> oder einer Zeitmarke.
datetime_year(DATUM)	ganze Zahl	Ergibt das Jahr aus einem <i>DATUM</i> oder einer Zeitmarke. Das Ergebnis ist eine ganze Zahl, wie beispielsweise 2002.
date_weeks_difference (DATUM1, DATUM2)	reelle Zahl	Ergibt die Zeit in Wochen (als reelle Zahl), die zwischen dem durch <i>DATUM1</i> und dem durch <i>DATUM2</i> angegebenen Datum bzw. der Zeitmarke liegt. Der Wert basiert auf einer Woche mit 7,0 Tagen. Wenn <i>DATUM2</i> vor <i>DATUM1</i> liegt, ergibt diese Funktion eine negative Zahl.
date_years_difference (DATUM1, DATUM2)	reelle Zahl	Ergibt die Zeit in Jahren (als reelle Zahl), die zwischen dem durch <i>DATUM1</i> und dem durch <i>DATUM2</i> angegebenen Datum bzw. der Zeitmarke liegt. Dies ist ein Näherungswert, der auf einem Jahr mit 365,25 Tagen beruht. Wenn <i>DATUM2</i> vor <i>DATUM1</i> liegt, ergibt diese Funktion eine negative Zahl.
time_before(ZEIT1, ZEIT2)	boolesch	Ergibt den Wert "Wahr", wenn <i>ZEIT1</i> eine Uhrzeit oder eine Zeitmarke darstellt, die/der vor der durch <i>ZEIT2</i> dargestellten Zeit liegt. Andernfalls ergibt die Funktion den Wert 0.

Tabelle 27. CLEM-Datums- und Uhrzeitfunktionen (Forts.).

Funktion	Ergebnis	Beschreibung
time_hours_difference (ZEIT1, ZEIT2)	reelle Zahl	Ergibt den Zeitunterschied in Stunden (als reelle Zahl) zwischen den durch ZEIT1 und ZEIT2 angegebenen Uhrzeiten oder Zeitmarken. Bei Auswahl von Tage/Minuten übertragen im Dialogfeld "Streameigenschaften" wird ein höherer Wert von ZEIT1 als Verweis auf den vorhergehenden Tag interpretiert. Wenn die Übertragungsoption nicht ausgewählt wird, führt ein höherer Wert von ZEIT1 dazu, dass der ausgegebene Wert negativ ist.
time_in_hours(ZEIT)	reelle Zahl	Ergibt die durch ZEIT angegebene Zeit in Stunden als reelle Zahl. So wird beispielsweise beim Zeitformat HHMM der Ausdruck time_in_hours('0130') als 1,5 ausgewertet. ZEIT kann für eine Uhrzeit oder eine Zeitmarke stehen.
time_in_mins(ZEIT)	reelle Zahl	Ergibt die durch ZEIT angegebene Zeit in Minuten als reelle Zahl. ZEIT kann für eine Uhrzeit oder eine Zeitmarke stehen.
time_in_secs(ZEIT)	ganze Zahl	Ergibt die durch ZEIT angegebene Zeit in Sekunden als ganze Zahl. ZEIT kann für eine Uhrzeit oder eine Zeitmarke stehen.
time_mins_difference(ZEIT1, ZEIT2)	reelle Zahl	Ergibt den Zeitunterschied in Minuten (als reelle Zahl) zwischen den durch ZEIT1 und ZEIT2 angegebenen Uhrzeiten oder Zeitmarken. Bei Auswahl von Tage/Minuten übertragen im Dialogfeld "Streameigenschaften" wird ein höherer Wert von ZEIT1 als Verweis auf den vorhergehenden Tag (bzw. die vorangegangene Stunde, sofern nur Minuten und Sekunden im aktuellen Format angegeben werden) interpretiert. Wenn die Übertragungsoption nicht ausgewählt wird, führt ein höherer Wert von ZEIT1 dazu, dass der ausgegebene Wert negativ ist.
time_secs_difference(ZEIT1, ZEIT2)	ganze Zahl	Ergibt den Zeitunterschied in Sekunden (als ganze Zahl) zwischen den durch ZEIT1 und ZEIT2 angegebenen Uhrzeiten oder Zeitmarken. Bei Auswahl von Tage/Minuten übertragen im Dialogfeld "Streameigenschaften" wird ein höherer Wert von ZEIT1 als Verweis auf den vorhergehenden Tag (bzw. die vorangegangene Stunde, sofern nur Minuten und Sekunden im aktuellen Format angegeben werden) interpretiert. Wenn die Übertragungsoption nicht ausgewählt wird, führt ein höherer Wert von ZEIT1 dazu, dass der ausgegebene Wert negativ ist.

Konvertieren von Datums- und Zeitwerten.

Beachten Sie, dass die Konvertierungsfunktionen (und alle anderen Funktionen, für die ein spezieller Eingabetyp, wie beispielsweise ein Wert für Datum oder Uhrzeit, erforderlich ist) von den aktuell im Dialogfeld für die Streamoptionen angegebenen Formaten abhängen. Ein Feld mit der Bezeichnung *DATUM* beispielsweise, das als Zeichenfolge mit den Werten *Jan 2003*, *Feb 2003* usw. gespeichert ist, könnte wie folgt in einen Datumsspeicher konvertiert werden:

```
to_date(DATUM)
```

Damit diese Konvertierung funktionieren kann, müssen Sie das entsprechende Datumsformat **MON JJJJ** als Standarddatumsformat für den Stream festlegen.

Ein Beispiel, bei dem Zeichenfolgewerte mithilfe eines Füllerknotens in Datumswerte konvertiert werden, finden Sie im Stream *broadband_create_models.str*, der im Ordner *\Demos* unter dem Unterordner *streams* installiert ist.

Als Zahlen gespeicherte Datumsangaben. Beachten Sie, dass *DATUM* im vorherigen Beispiel der Name eines Felds ist, während es sich bei *to_date* um eine CLEM-Funktion handelt. Wenn Datumsangaben als Zahlen gespeichert wurden, können Sie sie mithilfe der Funktion *datetime_date* konvertieren. Dabei wird die Zahl als Anfang von Sekunden seit dem Basisdatum (Epoche) interpretiert.

```
datetime_date(DATUM)
```

Indem Sie ein Datum in eine Sekundenzahl (und zurück) konvertieren, können Sie Berechnungen wie die Addition oder Subtraktion einer Anzahl von Tagen zum bzw. vom aktuellen Datum durchführen. Beispiel:

```
datetime_date((date_in_days(DATUM)-7)*60*60*24)
```

Sequenzfunktionen

Bei einigen Operationen ist die Abfolge (Sequenz) der Ereignisse von Bedeutung. Die Anwendung ermöglicht die Arbeit mit folgenden Datensatzsequenzen:

- Sequenzen und Zeitreihen
- Sequenzfunktionen
- Datensatzindizierung
- Durchschnittsbildung, Summierung und Vergleich von Werten
- Überwachen von Änderungen - Differenzierung
- @SINCE
- Offset-Werte
- Zusätzliche Sequenzfunktionen

Bei vielen Anwendungen kann jeder Datensatz, der einen Stream durchläuft, als Einzelfall betrachtet werden, der von allen anderen Fällen unabhängig ist. In solchen Situationen ist die Reihenfolge der Datensätze normalerweise nicht von Bedeutung.

Bei einigen Problemklassen jedoch ist die Datensatzsequenz höchst wichtig. Dies ist üblicherweise bei Zeitreihen der Fall, bei denen die Sequenz der Datensätze eine geordnete Sequenz von Ereignissen oder Vorkommen darstellt. Jeder Datensatz ist eine Momentaufnahme zu einem bestimmten Zeitpunkt; ein Großteil der aussagekräftigsten Informationen sind jedoch möglicherweise nicht in Augenblickswerten enthalten, sondern in der Art und Weise, wie sich diese Werte im Laufe der Zeit ändern und verhalten.

Der entscheidende Parameter muss natürlich nicht die Zeit sein. So könnten die Datensätze Analysen darstellen, die in Abständen entlang einer Linie durchgeführt werden. Doch auch hier gelten dieselben Prinzipien.

Sequenz- und Sonderfunktionen lassen sich sofort an folgenden Merkmalen erkennen:

- Sie enthalten alle das Präfix @.
- Die Namen dieser Funktionen bestehen aus Großbuchstaben.

Sequenzfunktionen können sich auf den derzeit von einem Knoten verarbeiteten Datensatz, auf die Datensätze, die einen Knoten bereits durchlaufen haben, und sogar, in einem Fall, auf Datensätze beziehen, die einen Knoten noch durchlaufen müssen. Sequenzfunktionen können nach Belieben mit anderen Komponenten von CLEM-Ausdrücken gemischt werden; bei einigen gelten jedoch Einschränkungen in Bezug auf die verwendeten Argumente.

Beispiele

Es kann interessant sein zu wissen, wie lange es zurückliegt, dass ein bestimmtes Ereignis eintrat oder eine Bedingung wahr war. Dies kann mit der Funktion @SINCE ermittelt werden. Beispiel:

```
@SINCE(Einkommen > Ausgaben)
```

Diese Funktion ergibt das Offset des letzten Datensatzes, bei dem diese Bedingung erfüllt war, also die Anzahl der Datensätze vor dem aktuellen Datensatz, bei denen die Bedingung wahr war. Wenn die Bedingung niemals wahr war, ergibt @SINCE den Wert @INDEX + 1.

Manchmal möchte man sich in dem von @SINCE verwendeten Ausdruck auf einen Wert des aktuellen Datensatzes beziehen. Dies ist mit der Funktion @THIS möglich, die angibt, dass ein Feldname immer für den aktuellen Datensatz gilt. Um das Offset des letzten Datensatzes zu ermitteln, bei dem der Wert des Felds Konzentration mehr als zweimal so hoch war wie der des aktuellen Datensatzes können Sie folgende Funktion verwenden:

```
@SINCE(Konzentration > 2 * @THIS(Konzentration))
```

In einigen Fällen ist die für @SINCE angegebene Bedingung beim aktuellen Datensatz per definitionem wahr. Beispiel:

```
@SINCE(ID == @THIS(ID))
```

Aus diesem Grund wertet @SINCE die Bedingung nicht für den aktuellen Datensatz aus. Mit einer ähnlichen Funktion @SINCE0 können Sie die Bedingung für den aktuellen Datensatz und die vorangegangenen auswerten; wenn die Bedingung im aktuellen Datensatz wahr ist, ergibt @SINCE0 den Wert 0.

Hinweis: @-Funktionen können nicht über Scripts aufgerufen werden.

Tabelle 28. CLEM-Sequenzfunktionen.

Funktion	Ergebnis	Beschreibung
@MEAN(FELD)	reelle Zahl	Ergibt den mittleren Durchschnitt der Werte für das angegebene FELD bzw. die angegebenen FELDER.
@MEAN(FELD, AUSDR)	reelle Zahl	Ergibt den mittleren Durchschnitt der Werte für FELD in den letzten AUSDR Datensätzen, die vom aktuellen Knoten erhalten wurden, einschließlich des aktuellen Datensatzes. FELD muss der Name eines numerischen Felds sein. AUSDR kann jeder Ausdruck sein, dessen Auswertung eine ganze Zahl größer als 0 ergibt. Wenn AUSDR weggelassen wird oder die Anzahl der bisher erhaltenen Datensätze übersteigt, ergibt sich der Durchschnitt aus den bisher erhaltenen Datensätzen. Beachten Sie, dass diese Funktion nicht über ein Script aufgerufen werden kann.
@MEAN(FELD, AUSDR, GANZZ)	reelle Zahl	Ergibt den mittleren Durchschnitt der Werte für FELD in den letzten AUSDR Datensätzen, die vom aktuellen Knoten erhalten wurden, einschließlich des aktuellen Datensatzes. FELD muss der Name eines numerischen Felds sein. AUSDR kann jeder Ausdruck sein, dessen Auswertung eine ganze Zahl größer als 0 ergibt. Wenn AUSDR weggelassen wird oder die Anzahl der bisher erhaltenen Datensätze übersteigt, ergibt sich der Durchschnitt aus den bisher erhaltenen Datensätzen. GANZZ gibt die maximale Anzahl von zurückliegenden Werten an, die betrachtet werden sollen. Dies ist wesentlich effizienter als die Verwendung von nur zwei Argumenten.
@DIFF1(FELD)	reelle Zahl	Ergibt das erste Differenzial von FELD1. Das Format mit einem einzelnen Argument ergibt also einfach die Differenz zwischen dem aktuellen Wert und dem früheren Wert des Feldes. Ergibt 0, wenn die relevanten vorhergehenden Datensätze nicht vorhanden sind.
@DIFF1(FELD1, FELD2)	reelle Zahl	Das Format mit einem zwei Argumenten ergibt das erste Differenzial von FELD1 in Bezug auf FELD2. Ergibt 0, wenn die relevanten vorhergehenden Datensätze nicht vorhanden sind.

Tabelle 28. CLEM-Sequenzfunktionen (Forts.).

Funktion	Ergebnis	Beschreibung
@DIFF2(FELD)	reelle Zahl	Ergibt das zweite Differenzial von <i>FELD1</i> . Das Format mit einem einzelnen Argument ergibt also einfach die Differenz zwischen dem aktuellen Wert und dem früheren Wert des Feldes. Ergibt 0, wenn die relevanten vorhergehenden Datensätze nicht vorhanden sind.
@DIFF2(FELD1, FELD2)	reelle Zahl	Das Format mit einem zwei Argumenten ergibt das erste Differenzial von <i>FELD1</i> in Bezug auf <i>FELD2</i> . Ergibt 0, wenn die relevanten vorhergehenden Datensätze nicht vorhanden sind.
@INDEX	ganze Zahl	Ergibt den Index des aktuellen Datensatzes. Indizes werden den Datensätzen zugeordnet, wenn sie am aktuellen Knoten ankommen. Der erste Datensatz erhält den Index 1. Dieser Index wird für jeden nachfolgenden Datensatz um den Wert 1 erhöht.
@LAST_NON_BLANK(FELD)	beliebig	Ergibt den letzten nicht leeren Wert für <i>FELD</i> , wie in einem aufwärts liegenden Quellen- oder Typknoten definiert. Wenn in den bisher gelesenen Datensätzen keine nicht leeren Werte für <i>FELD</i> vorhanden sind, wird \$null\$ zurückgegeben. Leerwerte (auch als benutzerdefiniert fehlende Werte bezeichnet) können für jedes Feld einzeln definiert werden.
@MAX(FELD)	Zahl	Ergibt den Höchstwert für das angegebene <i>FELD</i> .
@MAX(FELD, AUSDR)	Zahl	Ergibt den Höchstwert für <i>FELD</i> in den letzten <i>AUSDR</i> Datensätzen, die bisher erhalten wurden, einschließlich des aktuellen Datensatzes. <i>FELD</i> muss der Name eines numerischen Felds sein. <i>AUSDR</i> kann jeder Ausdruck sein, dessen Auswertung eine ganze Zahl größer als 0 ergibt.
@MAX(FELD, AUSDR, GANZZ)	Zahl	Ergibt den Höchstwert für <i>FELD</i> in den letzten <i>AUSDR</i> Datensätzen, die bisher erhalten wurden, einschließlich des aktuellen Datensatzes. <i>FELD</i> muss der Name eines numerischen Felds sein. <i>AUSDR</i> kann jeder Ausdruck sein, dessen Auswertung eine ganze Zahl größer als 0 ergibt. Wenn <i>AUSDR</i> weggelassen wird oder die Anzahl der bisher erhaltenen Datensätze übersteigt, ergibt sich der Höchstwert aus den bisher erhaltenen Datensätzen. <i>GANZZ</i> gibt die maximale Anzahl von zurückliegenden Werten an, die betrachtet werden sollen. Dies ist wesentlich effizienter als die Verwendung von nur zwei Argumenten.
@MIN(FELD)	Zahl	Ergibt den Mindestwert für das angegebene <i>FELD</i> .
@MIN(FELD, AUSDR)	Zahl	Ergibt den Mindestwert für <i>FELD</i> in den letzten <i>AUSDR</i> Datensätzen, die bisher erhalten wurden, einschließlich des aktuellen Datensatzes. <i>FELD</i> muss der Name eines numerischen Felds sein. <i>AUSDR</i> kann jeder Ausdruck sein, dessen Auswertung eine ganze Zahl größer als 0 ergibt.

Tabelle 28. CLEM-Sequenzfunktionen (Forts.).

Funktion	Ergebnis	Beschreibung
@MIN(FELD, AUSDR, GANZZ)	Zahl	Ergibt den Mindestwert für <i>FELD</i> in den letzten <i>AUSDR</i> Datensätzen, die bisher erhalten wurden, einschließlich des aktuellen Datensatzes. <i>FELD</i> muss der Name eines numerischen Felds sein. <i>AUSDR</i> kann jeder Ausdruck sein, dessen Auswertung eine ganze Zahl größer als 0 ergibt. Wenn <i>AUSDR</i> weggelassen wird oder die Anzahl der bisher erhaltenen Datensätze übersteigt, ergibt sich der Mindestwert aus den bisher erhaltenen Datensätzen. <i>GANZZ</i> gibt die maximale Anzahl von zurückliegenden Werten an, die betrachtet werden sollen. Dies ist wesentlich effizienter als die Verwendung von nur zwei Argumenten.
@OFFSET(FELD, AUSDR)	beliebig	Ergibt den Wert von <i>FELD</i> in dem Datensatz, der gegenüber dem aktuellen Datensatz ein Offset mit dem Wert von <i>AUSDR</i> aufweist. Ein positives Offset bezieht sich auf einen Datensatz, der den Knoten bereits durchlaufen hat, ein negatives Offset gibt ein "Lookahead" auf einen Datensatz an, der noch nicht angekommen ist. Beispiel: @OFFSET(Status, 1) ergibt den Wert des Felds Status im vorangegangenen Datensatz, während @OFFSET(Status, -4) vier Datensätze nach vorn in der Sequenz blickt (also auf Datensätze, die den Knoten noch nicht durchlaufen haben), um den Wert zu ermitteln. <i>Beachten Sie, dass ein negatives Offset (Lookahead-Offset) als Konstante angegeben sein muss.</i> Bei positiven Offsets gilt Folgendes: Bei <i>AUSDR</i> kann es sich auch um einen beliebigen CLEM-Ausdruck handeln, der für den aktuellen Datensatz ausgewertet wird, um das Offset zu erhalten. In diesem Fall sollte die Version dieser Funktion mit drei Argumenten die Leistung verbessern (siehe nächste Funktion). Wenn der Ausdruck einen anderen Wert als eine nicht negative ganze Zahl zurückgibt, wird ein Fehler ausgelöst. Berechnete Lookahead-Offsets sind also nicht zulässig. <i>Hinweis:</i> Eine selbstreferenzielle @OFFSET-Funktion kann kein Literal-Lookahead verwenden. In einem Füllerknoten beispielsweise kann der Wert von Feld1 nicht mithilfe eines Ausdrucks wie @OFFSET(Feld1, -2) ersetzt werden.
@OFFSET(FELD, AUSDR, GANZZ)	beliebig	Führt die gleiche Operation durch wie @OFFSET, allerdings mit einem zusätzlichen dritten Argument, <i>GANZZ</i> , das die maximale Anzahl der zurückliegenden Werte angibt, die betrachtet werden sollen. Wenn das Offset aus einem Ausdruck berechnet wird, sollte das dritte Argument zu einer Verbesserung der Leistung führen. In einem Ausdruck wie @OFFSET(Dings, Monat, 12) beispielsweise wird das System angewiesen, nur die letzten 12 Werte von Dings zu behalten. Andernfalls müsste sicherheitshalber jeder Wert gespeichert werden. In Fällen, in denen es sich beim Offset-Wert um eine Konstante handelt - einschließlich der negativen "Lookahead"-Offsets, die konstant sein müssen - ist das dritte Argument sinnlos, weshalb hier die Version dieser Funktion mit zwei Argumenten verwendet werden sollte. Siehe auch den Hinweis zu selbstreferenziellen Funktionen für die Version mit zwei Argumenten weiter oben.
@SDEV(FELD)	reelle Zahl	Ergibt die Standardabweichung der Werte für das angegebene <i>FELD</i> bzw. die angegebene <i>FELDER</i> .

Tabelle 28. CLEM-Sequenzfunktionen (Forts.).

Funktion	Ergebnis	Beschreibung
@SDEV(FELD, AUSDR)	reelle Zahl	Gibt die Standardabweichung der Werte für <i>FELD</i> in den letzten <i>AUSDR</i> Datensätzen zurück, die vom aktuellen Knoten erhalten wurden, einschließlich des aktuellen Datensatzes. <i>FELD</i> muss der Name eines numerischen Felds sein. <i>AUSDR</i> kann jeder Ausdruck sein, dessen Auswertung eine ganze Zahl größer als 0 ergibt. Wenn <i>AUSDR</i> weggelassen wird oder die Anzahl der bisher erhaltenen Datensätze übersteigt, ergibt sich die Standardabweichung aus allen bisher erhaltenen Datensätzen.
@SDEV(FELD, AUSDR, GANZZ)	reelle Zahl	Gibt die Standardabweichung der Werte für <i>FELD</i> in den letzten <i>AUSDR</i> Datensätzen zurück, die vom aktuellen Knoten erhalten wurden, einschließlich des aktuellen Datensatzes. <i>FELD</i> muss der Name eines numerischen Felds sein. <i>AUSDR</i> kann jeder Ausdruck sein, dessen Auswertung eine ganze Zahl größer als 0 ergibt. Wenn <i>AUSDR</i> weggelassen wird oder die Anzahl der bisher erhaltenen Datensätze übersteigt, ergibt sich die Standardabweichung aus allen bisher erhaltenen Datensätzen. <i>GANZZ</i> gibt die maximale Anzahl von zurückliegenden Werten an, die betrachtet werden sollen. Dies ist wesentlich effizienter als die Verwendung von nur zwei Argumenten.
@SINCE(AUSDR)	beliebig	Gibt an, wie viele Datensätze den Knoten durchlaufen haben, seit <i>AUSDR</i> , ein beliebiger CLEM-Ausdruck, wahr war.
@SINCE(AUSDR, GANZZ)	beliebig	Das zweite Argument, <i>GANZZ</i> , gibt die maximale Anzahl von zurückliegenden Werten an, die betrachtet werden sollen. Wenn <i>AUSDR</i> niemals wahr war, ist <i>GANZZ</i> gleich @INDEX+1.
@SINCE0(AUSDR)	beliebig	Berücksichtigt den aktuellen Datensatz, was bei @SINCE nicht der Fall ist; @SINCE0 ergibt 0, wenn <i>AUSDR</i> für den aktuellen Datensatz wahr ist.
@SINCE0(AUSDR, GANZZ)	beliebig	Das zweite Argument, <i>GANZZ</i> , gibt die maximale Anzahl von zurückliegenden Werten an, die betrachtet werden sollen.
@SUM(FELD)	Zahl	Ergibt die Summe der Werte für das angegebene <i>FELD</i> bzw. die angegebene <i>FELDER</i> .
@SUM(FELD, AUSDR)	Zahl	Ergibt die Summe der Werte für <i>FELD</i> in den letzten <i>AUSDR</i> Datensätzen, die vom aktuellen Knoten erhalten wurden, einschließlich des aktuellen Datensatzes. <i>FELD</i> muss der Name eines numerischen Felds sein. <i>AUSDR</i> kann jeder Ausdruck sein, dessen Auswertung eine ganze Zahl größer als 0 ergibt. Wenn <i>AUSDR</i> weggelassen wird oder die Anzahl der bisher erhaltenen Datensätze übersteigt, ergibt sich die Summe aus den bisher erhaltenen Datensätzen.

Tabelle 28. CLEM-Sequenzfunktionen (Forts.).

Funktion	Ergebnis	Beschreibung
@SUM(FELD, AUSDR, GANZZ)	Zahl	Ergibt die Summe der Werte für <i>FELD</i> in den letzten <i>AUSDR</i> Datensätzen, die vom aktuellen Knoten erhalten wurden, einschließlich des aktuellen Datensatzes. <i>FELD</i> muss der Name eines numerischen Felds sein. <i>AUSDR</i> kann jeder Ausdruck sein, dessen Auswertung eine ganze Zahl größer als 0 ergibt. Wenn <i>AUSDR</i> weggelassen wird oder die Anzahl der bisher erhaltenen Datensätze übersteigt, ergibt sich die Summe aus den bisher erhaltenen Datensätzen. <i>GANZZ</i> gibt die maximale Anzahl von zurückliegenden Werten an, die betrachtet werden sollen. Dies ist wesentlich effizienter als die Verwendung von nur zwei Argumenten.
@THIS(FELD)	beliebig	Ergibt den Wert des Felds namens <i>FELD</i> im aktuellen Datensatz. Wird nur bei @SINCE-Ausdrücken verwendet.

Globale Funktionen

Die Funktionen @MEAN,@SUM, @MIN, @MAX und @SDEV werden maximal auf alle Datensätze bis einschließlich des aktuellen Datensatzes angewendet. In einigen Fällen ist es jedoch von Vorteil, in der Lage zu sein, die Daten im aktuellen Datensatz mit Werten im gesamten Dataset zu vergleichen. Mit einem Globalwerteknoten können Sie Werte im gesamten Dataset generieren, auf die Sie dann mit den globalen Funktionen in einem CLEM-Ausdruck zugreifen können.

Beispiel:

@GLOBAL_MAX(Alter)

ergibt den höchsten Wert von Alter im Dataset, während der Ausdruck
(Wert - @GLOBAL_MEAN(Wert)) / @GLOBAL_SDEV(Wert)

die Differenz zwischen dem Wert dieses Datensatzes und dem globalen Mittelwert als Anzahl der Standardabweichungen angibt. Globale Werte können nur verwendet werden, nachdem sie von einem Globalwerteknoten berechnet wurden. Alle aktuellen Globalwerte können durch Klicken auf die Schaltfläche zum Löschen der Globalwerte auf der Registerkarte "Globalwerte" im Dialogfeld "Streameigenschaften" entfernt werden.

Hinweis: @-Funktionen können nicht über Scripts aufgerufen werden.

Tabelle 29. Globale Funktionen von CLEM.

Funktion	Ergebnis	Beschreibung
@GLOBAL_MAX(FELD)	Zahl	Ergibt den Höchstwert für <i>FELD</i> im gesamten Dataset, das zuvor durch einen Globalwerteknoten erstellt wurde. <i>FELD</i> muss der Name eines numerischen Felds sein. Wenn der entsprechende Globalwert nicht festgelegt wurde, wird ein Fehler ausgegeben. Beachten Sie, dass diese Funktion nicht über ein Script aufgerufen werden kann.
@GLOBAL_MIN(FELD)	Zahl	Ergibt den Höchstwert für <i>FELD</i> im gesamten Dataset, das zuvor durch einen Globalwerteknoten generiert wurde. <i>FELD</i> muss der Name eines numerischen Felds sein. Wenn der entsprechende Globalwert nicht festgelegt wurde, wird ein Fehler ausgegeben.

Tabelle 29. Globale Funktionen von CLEM (Forts.).

Funktion	Ergebnis	Beschreibung
@GLOBAL_SDEV(FELD)	Zahl	Ergibt die Standardabweichung der Werte für <i>FELD</i> im gesamten Dataset, das zuvor durch einen Globalwerteknoten generiert wurde. <i>FELD</i> muss der Name eines numerischen Felds sein. Wenn der entsprechende Globalwert nicht festgelegt wurde, wird ein Fehler ausgegeben.
@GLOBAL_MEAN(FELD)	Zahl	Ergibt den mittleren Durchschnitt der Werte für <i>FELD</i> im gesamten Dataset, das zuvor durch einen Globalwerteknoten generiert wurde. <i>FELD</i> muss der Name eines numerischen Felds sein. Wenn der entsprechende Globalwert nicht festgelegt wurde, wird ein Fehler ausgegeben.
@GLOBAL_SUM(FELD)	Zahl	Ergibt die Summe der Werte für <i>FELD</i> im gesamten Dataset, das zuvor durch einen Globalwerteknoten generiert wurde. <i>FELD</i> muss der Name eines numerischen Felds sein. Wenn der entsprechende Globalwert nicht festgelegt wurde, wird ein Fehler ausgegeben.

Funktionen zum Umgang mit Leerstellen und Nullwerten

Mit CLEM können Sie angeben, dass bestimmte Werte in einem Feld als "Leerstellen" oder fehlende Werte betrachtet werden sollen. Die nachstehenden Funktionen können auf Leerstellen angewendet werden.

Hinweis: @-Funktionen können nicht über Scripts aufgerufen werden.

Tabelle 30. CLEM-Funktionen zum Umgang mit Leerstellen und Nullwerten.

Funktion	Ergebnis	Beschreibung
@BLANK(FELD)	boolesch	Ergibt den Wert "Wahr" für alle Datensätze, deren Werte gemäß den Regeln zum Umgang mit Leerstellen, die in einem weiter oben im Stream gelegenen Typknoten oder Quellenknoten (Registerkarte "Typen") festgelegt wurden, Leerstellen sind. Beachten Sie, dass diese Funktion nicht über ein Script aufgerufen werden kann.
@LAST_NON_BLANK(FELD)	beliebig	Ergibt den letzten nicht leeren Wert für <i>FELD</i> , wie in einem aufwärts liegenden Quellen- oder Typknoten definiert. Wenn in den bisher gelesenen Datensätzen keine nicht leeren Werte für <i>FELD</i> vorhanden sind, wird \$null\$ zurückgegeben. Leerwerte (auch als benutzerdefiniert fehlende Werte bezeichnet) können für jedes Feld einzeln definiert werden.
@NULL(FELD)	boolesch	Ergibt den Wert "Wahr", wenn der Wert von <i>FELD</i> der systemdefiniert fehlende Wert \$null\$ ist. Ergibt den Falschwert für alle anderen Werte, einschließlich benutzerdefinierter Leerstellen. Wenn Sie die Daten auf beides prüfen möchten, verwenden Sie @BLANK(FELD) und @NULL(FELD).
undef	beliebig	Wird allgemein in CLEM zur Eingabe eines \$null\$-Wertes verwendet; beispielsweise, um Leerwerte im Füllerknoten mit Nullen aufzufüllen.

Mithilfe des Füllerknotens können leere Felder "aufgefüllt" werden. Bei Füller- und Ableitungsknoten (nur beim Mehrfachmodus) bezieht sich die CLEM-Sonderfunktion @FIELD auf die aktuell untersuchten Felder.

Sonderfelder

Sonderfunktionen dienen zur Kennzeichnung der speziellen untersuchten Felder oder zum Erzeugen einer Felderliste als Eingabe. Wenn Sie beispielsweise mehrere Felder gleichzeitig ableiten, sollten Sie mit @FIELD angeben, dass diese Ableitungsaktion für die ausgewählten Felder durchgeführt werden soll. Mit dem Ausdruck log(@FIELD) wird ein neues Log-Feld für jedes ausgewählte Feld abgeleitet.

Hinweis: @-Funktionen können nicht über Scripts aufgerufen werden.

Table 31. CLEM-Sonderfelder.

Funktion	Ergebnis	Beschreibung
@FIELD	<i>beliebig</i>	Führt eine Aktion für alle im Ausdruckskontext angegebenen Felder durch. Beachten Sie, dass diese Funktion nicht über ein Script aufgerufen werden kann.
@TARGET	<i>beliebig</i>	Wenn ein CLEM-Ausdruck in einer benutzerdefinierten Analysefunktion verwendet wird, steht @TARGET für das Zielfeld bzw. den "korrekten Wert" für das analysierte Paar aus Zielwert und vorhergesagtem Wert. Diese Funktion wird häufig in Analyseknöten verwendet.
@PREDICTED	<i>beliebig</i>	Wenn ein CLEM-Ausdruck in einer benutzerdefinierten Analysefunktion verwendet wird, steht @PREDICTED für den vorhergesagten Wert des analysierten Paares aus Zielwert und vorhergesagtem Wert. Diese Funktion wird häufig in Analyseknöten verwendet.
@PARTITION_FIELD	<i>beliebig</i>	Ersetzt den Namen des aktuellen Partitionsfelds.
@TRAINING_PARTITION	<i>beliebig</i>	Ergibt den Wert der aktuellen Trainingspartition. Beispiel: Wenn Sie einen Auswahlknöten verwenden wollen, um Trainingsdatensätze auszuwählen, verwenden Sie den CLEM-Ausdruck @PARTITION_FIELD = @TRAINING_PARTITION. Dadurch wird gewährleistet, dass der Auswahlknöten immer funktioniert, unabhängig davon, welche Werte zur Darstellung der einzelnen Partitionen in den Daten verwendet werden.
@TESTING_PARTITION	<i>beliebig</i>	Ergibt den Wert der aktuellen Testpartition.
@VALIDATION_PARTITION	<i>beliebig</i>	Ergibt den Wert der aktuellen Validierungspartition.
@FIELDS_BETWEEN(Anfang, Ende)	<i>beliebig</i>	Ergibt die Liste der Feldnamen, die zwischen dem angegebenen Anfangs- und Endfeld (einschließlich dieser Felder) liegen, und zwar auf der Grundlage der natürlichen Reihenfolge der Felder in den Daten, also gemäß der Einfügereihenfolge.

Tabelle 31. CLEM-Sonderfelder (Forts.).

Funktion	Ergebnis	Beschreibung
@FIELDS_MATCHING(Muster)	<i>beliebig</i>	Ergibt eine Liste der Feldnamen, die mit dem angegebenen Muster übereinstimmen. Das Fragezeichen (?) als Platzhalterzeichen kann die Stelle von genau einem beliebigen Zeichen einnehmen, der Stern (*) entspricht keinem, einem oder mehreren Zeichen. Soll ein Fragezeichen oder ein Stern nicht als Platzhalter fungieren, sondern bei der Suche berücksichtigt werden, geben Sie den umgekehrten Schrägstrich (\) als Escapezeichen ein.
@MULTI_RESPONSE_SET	<i>beliebig</i>	Ergibt die Liste der Felder im benannten Mehrfachantwortset.

Kapitel 9. Verwenden von IBM SPSS Modeler mit einem Repository

Informationen zu IBM SPSS Collaboration and Deployment Services Repository

Sie können IBM SPSS Modeler zusammen mit einem Repository von IBM SPSS Collaboration and Deployment Services verwenden und dadurch die Lebensdauer von Data-Mining-Modellen und verwandten Vorhersageobjekten verwalten. Dies ermöglicht, dass diese Objekte im Zusammenhang mit Unternehmensanwendungen, Tools und Lösungen verwendet werden. IBM SPSS Modeler-Objekte, die auf diese Weise gemeinsam benutzbar sind, umfassen Streams, Knoten, Streamausgaben, Szenarios, Projekte und Modelle. Die Objekte werden im zentralen Repository gespeichert, wo sie für andere Anwendungen freigegeben und mithilfe von umfassenden Versionserstellungs-, Metadaten- und Suchfunktionen aufgezeichnet werden können.

Hinweis: Für den Zugriff auf ein Repository von IBM SPSS Collaboration and Deployment Services ist eine separate Lizenz erforderlich. Weitere Informationen finden Sie unter <http://www.ibm.com/software/analytics/spss/products/deployment/cds/>.

Bevor Sie IBM SPSS Modeler mit dem Repository verwenden können, müssen Sie einen Adapter auf dem Repository-Host installieren. Ohne diesen Adapter sehen Sie eventuell folgende Nachricht beim Versuch, über bestimmte IBM SPSS Modeler-Knoten oder -Modelle auf Repository-Objekte zuzugreifen:

Das Repository benötigt evtl. Updates zur Unterstützung neuer Knoten-, Modell- und Ausgabetypen.

Anweisungen zum Installieren des Adapters enthält das Installationshandbuch zu IBM SPSS Modeler Deployment, das auf der IBM SPSS Modeler Deployment-DVD enthalten ist. Genauere Angaben dazu, wie Sie über IBM SPSS Collaboration and Deployment Services Deployment Manager auf IBM SPSS Modeler-Repository-Objekte zugreifen, finden Sie Bereitstellungshandbuch zu IBM SPSS Modeler.

Die folgenden Abschnitte enthalten Informationen zum Zugriff auf das Repository über IBM SPSS Modeler.

Umfassende Versionserstellung und Unterstützung der Suchfunktion

Das Repository bietet umfassende Funktionen für die Objektversionserstellung und Suchmöglichkeiten. Angenommen, Sie erstellen z. B. einen Stream und speichern ihn im Repository, wo Sie ihn mit Mitarbeitern aus anderen Abteilungen gemeinsam nutzen können. Wenn Sie den Stream später in IBM SPSS Modeler aktualisieren, können Sie die aktualisierte Version zum Repository hinzufügen, ohne die frühere Version zu überschreiben. Es können weiterhin alle Versionen aufgerufen und nach Name, Beschriftung, verwendeten Feldern und anderen Attributen durchsucht werden. So können Sie z. B. alle Modellversionen suchen, die die Nettoeinkünfte als Eingabe verwenden, oder alle von einem bestimmten Autor erstellten Modelle. (Mit einem herkömmlichen Dateisystem müssten Sie jede Version unter einem anderen Dateinamen speichern, und die Beziehungen zwischen den Versionen wären der Software nicht bekannt.)

Single Sign-on

Mit der Single-Sign-on-Funktion können sich Benutzer mit dem Repository verbinden, ohne jedes Mal Benutzername und Kennwort eingeben zu müssen. Die bestehenden Anmeldedaten für das lokale Netz sind als Authentifizierung für IBM SPSS Collaboration and Deployment Services ausreichend. Diese Funktion hängt von folgenden Bedingungen ab:

- IBM SPSS Collaboration and Deployment Services muss für die Verwendung eines Single-Sign-on-Anbieters konfiguriert sein.

- Der Benutzer muss in einem Host angemeldet sein, der mit dem Anbieter kompatibel ist.

Weitere Informationen finden Sie im Thema „Verbinden mit dem Repository“ auf Seite 127.

Speichern und Bereitstellen von Repository-Objekten

Streams, die in IBM SPSS Modeler erstellt werden, lassen sich so, wie sie sind, als Dateien mit der Erweiterung *.str* im Repository **speichern**. Auf diese Weise können mehrere Benutzer im gesamten Unternehmen auf einen einzelnen Stream zugreifen. Weitere Informationen finden Sie im Thema „Speichern von Objekten im Repository“ auf Seite 128.

Es ist auch möglich, einen Stream im Repository **bereitstellen**. Bereitgestellte Streams werden als Datei mit zusätzlichen Metadaten gespeichert. Bereitgestellte Streams können die Unternehmensfunktionen von IBM SPSS Collaboration and Deployment Services, wie automatisiertes Scoring und Modellaktualisierung, vollständig nutzen. So kann beispielsweise ein Modell automatisch in regelmäßigen Abständen aktualisiert werden, wenn neue Daten verfügbar werden. Alternativ kann ein Set von Streams zum Zweck der Champion-Challenger-Analyse verwendet werden, in der Streams verglichen werden, um zu ermitteln, welcher das wirksamste Vorhersagemodell enthält.

Sie haben zwei Möglichkeiten, einen Stream bereitzustellen: als Stream (mit der Erweiterung *.str*) oder als Szenario (mit der Erweiterung *.scn*). Bei der Bereitstellung als Stream kann der Stream durch die Thin-Client-Anwendung IBM SPSS Modeler Advantage aktiviert werden. Weitere Informationen finden Sie im Thema „Öffnen eines Streams in IBM SPSS Modeler Advantage“ auf Seite 145. Durch die Bereitstellung als Szenario kann der Stream von Predictive Applications Version 5, dem Vorgänger von IBM SPSS Modeler Advantage, verwendet werden.

Weitere Informationen finden Sie in „Bereitstellungsoptionen für den Stream“ auf Seite 139.

Anforderungen für Streams, die als Szenarios bereitgestellt werden

- Um einen konsistenten Zugriff auf Unternehmensdaten zu gewährleisten, muss der Zugriff auf die als Szenarios bereitgestellten Streams über die Enterprise View-Komponente von IBM SPSS Collaboration and Deployment Services erfolgen. Das bedeutet, dass in IBM SPSS Modeler für jede angegebene Scoring- bzw. Modellierungsverzweigung mindestens ein Quellenknoten vom Typ "Enterprise-Ansicht" (Enterprise View) vorhanden sein muss.
- Um den Enterprise-Ansichtsknoten verwenden zu können, muss IBM SPSS Collaboration and Deployment Services an Ihrem Standort installiert, konfiguriert und zugreifbar sein. Dabei müssen bereits eine Enterprise-Ansicht, Anwendungsansichten und DPDs (Data Provider Definitions) definiert sein. Weitere Informationen erhalten Sie bei Ihrem lokalen Administrator oder auf der Unternehmenswebsite unter <http://www.ibm.com/software/analytics/spss/products/deployment/cds/>.
- Eine DPD wird anhand einer bestimmten ODBC-Datenquelle definiert. Um eine DPD aus IBM SPSS Modeler zu verwenden, muss eine ODBC-Datenquelle auf dem IBM SPSS Modeler Server-Host definiert sein, der denselben Namen trägt und der eine Verbindung zu demselben Datenspeicher herstellt wie dem in der DPD referenzierten Datenspeicher.
- Außerdem muss auf jedem Computer, der zur Bearbeitung oder Ausführung des Streams verwendet wird, der IBM SPSS Collaboration and Deployment Services Enterprise View Driver installiert sein. Unter Windows installieren Sie den Treiber einfach auf dem Computer, auf dem SPSS Modeler bzw. SPSS Modeler Server installiert ist. Es ist keine weitere Konfiguration des Treibers erforderlich. Unter UNIX muss ein Verweis auf das Script *pev.sh* zum Startscript hinzugefügt werden. Ausführliche Informationen zur Installation von IBM SPSS Collaboration and Deployment Services Enterprise View Driver erhalten Sie bei Ihrem lokalen Administrator.

Andere Bereitstellungsoptionen

IBM SPSS Collaboration and Deployment Services bietet zwar bereits extrem umfassende Funktionen zur Verwaltung von Unternehmensinhalten, es steht aber dennoch eine Reihe weiterer Mechanismen zur Bereitstellung und zum Export von Streams zur Verfügung, darunter:

- Exportieren Sie den Stream zur späteren Verwendung in IBM SPSS Modeler Solution Publisher Runtime.
- Exportieren Sie ein oder mehrere Modelle in PMML, einem XML-basierten Format zur Codierung von Modellinformationen. Weitere Informationen finden Sie im Thema „Importieren und Exportieren von Modellen als PMML“ auf Seite 146.

Verbinden mit dem Repository

1. Um eine Verbindung mit dem Repository herzustellen, klicken Sie im IBM SPSS Modeler-Hauptmenü auf:

Tools > Repository > Optionen...

2. Geben Sie in das Feld **Repository-URL** den Verzeichnispfad zur oder die URL der Repository-Installation ein, auf die Sie zugreifen möchten, oder wählen Sie diesen Verzeichnispfad bzw. diese URL aus. Sie können immer nur die Verbindung zu einem einzigen Repository herstellen.

Die Einstellungen sind für jeden Standort oder jede Installation spezifisch. Für spezielle Anmeldedaten wenden Sie sich an den lokalen Systemadministrator.

Hinweis: Für den Zugriff auf ein Repository von IBM SPSS Collaboration and Deployment Services ist eine separate Lizenz erforderlich. Weitere Informationen finden Sie unter <http://www.ibm.com/software/analytics/spss/products/deployment/cds/>.

Berechtigungs nachweise festlegen. Belassen Sie dieses Kontrollkästchen nicht aktiviert, um die Funktion für **Single Sign-on** zu aktivieren. Diese versucht, Sie mithilfe Ihres lokalen Benutzernamens und Kennworts anzumelden. Falls ein Single Sign-on nicht möglich ist oder Sie das Kontrollkästchen zur Inaktivierung des Single Sign-on aktivieren (z. B. zur Anmeldung in ein Administratorkonto), wird ein weiteres Fenster angezeigt, in dem Sie aufgefordert werden, Ihre Berechtigungs nachweise einzugeben.

Eingabe der Berechtigungs nachweise für das Repository

In Abhängigkeit von Ihren Einstellungen sind möglicherweise folgende Felder im Repository erforderlich: Dialogfeld "Berechtigungs nachweise":

Benutzername und Kennwort. Geben Sie einen gültigen Benutzernamen und ein Kennwort für die Anmeldung an. Bitten Sie, falls erforderlich, Ihren lokalen Administrator um weitere Informationen.

Provider. Wählen Sie zur Authentifizierung einen Sicherheitsanbieter aus. Das Repository kann für die Verwendung verschiedener Sicherheitsanbieter konfiguriert werden; wenn Sie mehr Hilfe benötigen, wenden Sie sich an Ihren lokalen Administrator.

Repository und Benutzername merken. Speichert die aktuellen Einstellungen als Standard, sodass Sie diese nicht bei jeder neuen Verbindung erneut eingeben müssen.

Durchsuchen des Repository-Inhalts

Mit dem Repository können Sie die gespeicherten Inhalte ähnlich wie mit Windows Explorer durchsuchen. Außerdem können Sie verschiedene *Versionen* von jedem gespeicherten Objekt durchsuchen.

1. Um das Fenster von IBM SPSS Collaboration and Deployment Services Repository zu öffnen, klicken Sie in den SPSS Modeler-Menüs auf:

Tools > Repository > Explorative Datenanalyse...

1. Geben Sie gegebenenfalls Einstellungen für die Verbindung mit dem Repository ein. Weitere Informationen finden Sie im Thema „Verbinden mit dem Repository“ auf Seite 127. Wenden Sie sich für spezielle Port-, Kennwort- und andere Verbindungsdaten an den lokalen Systemadministrator.

Das Explorer-Fenster zeigt anfangs eine Baumansicht der Ordnerhierarchie. Klicken Sie auf einen Ordnernamen, um den Inhalt des Ordners anzuzeigen.

Objekte, die der aktuellen Auswahl oder Suchkriterien entsprechen, werden zusammen mit detaillierten Informationen über die ausgewählte Version, die im unteren rechten Fenster angezeigt wird, im rechten Fenster aufgelistet. Die angezeigten Attribute gelten für die aktuellste Version.

Speichern von Objekten im Repository

Sie können Streams, Knoten, Modelle, Modellpaletten, Projekte und Ausgabeobjekte im Repository speichern, von wo andere Benutzer und Anwendungen darauf zugreifen können.

Hinweis: Für den Zugriff auf ein Repository von IBM SPSS Collaboration and Deployment Services ist eine separate Lizenz erforderlich. Weitere Informationen finden Sie unter <http://www.ibm.com/software/analytics/spss/products/deployment/cds/>.

Sie können auch Streamausgaben im Repository in einem Format veröffentlichen, in dem andere Benutzer diese mithilfe von IBM SPSS Collaboration and Deployment Services Deployment Portal über das Internet betrachten können.

Festlegen der Objekteigenschaften

Beim Speichern eines Objekts wird das Dialogfeld "Repository: Speichern" angezeigt, in dem Sie die Werte von mehreren Eigenschaften für das Objekt einstellen können. Sie verfügen über folgende Möglichkeiten:

- Auswählen des Namens und des Repository-Ordners, unter dem das Objekt gespeichert werden soll
- Hinzufügen von Informationen über das Objekt, z. B. Versionsbeschriftungen und andere Eigenschaften, nach denen gesucht werden kann
- Hinzufügen eines oder mehrerer Klassifizierungsthemen zum Objekt
- Festlegen von Sicherheitsoptionen für das Objekt

In den folgenden Abschnitten werden die einstellbaren Eigenschaften beschrieben.

Auswählen des Speicherorts für Objekte

Geben Sie in das Dialogfeld "Repository: Speichern" Folgendes ein.

Speichern in. Zeigt den aktuellen Ordner an, in dem das Objekt gespeichert wird. Doppelklicken Sie auf einen Ordnernamen in der Liste, um diesen Ordner als aktuellen Ordner festzulegen. Mit der Schaltfläche "Ordner nach oben" navigieren Sie zum übergeordneten Ordner. Mit der Schaltfläche "Neuer Ordner" erstellen Sie einen Ordner auf der aktuellen Ebene.

Dateiname. Der Name, unter dem das Objekt gespeichert wird.

Speichern. Speichert das Objekt am aktuellen Speicherort.

Hinzufügen von Informationen über gespeicherte Objekte

Alle Felder auf der Registerkarte "Informationen" des Dialogfelds "Repository: Speichern" sind optional.

Autor. Der Benutzername des Benutzers, der das Objekt im Repository erstellt. Standardmäßig wird der Benutzername angezeigt, der für die Repository-Verbindung verwendet wurde, Sie können diesen Namen aber hier ändern.

Versionsbeschriftung. Wählen Sie eine Beschriftung aus der Liste aus, um die Objektversion anzugeben, oder klicken Sie auf **Hinzufügen**, um eine neue Beschriftung zu erstellen. Verwenden Sie in der Beschriftung nicht das Zeichen "[". Stellen Sie sicher, dass keine Kontrollkästchen aktiviert sind, wenn Sie dieser Objektversion keine Beschriftung zuweisen möchten. Weitere Informationen finden Sie im Thema „Anzeigen und Bearbeiten von Objekteigenschaften“ auf Seite 137.

Beschreibung. Eine Beschreibung des Objekts. Benutzer können Objekte nach Beschreibung suchen (siehe Hinweis).

Schlüsselwörter. Ein oder mehrere Stichwörter, die sich auf das Objekt beziehen und die zum Suchen verwendet werden können (siehe Hinweis).

Ablaufdatum. Ein Datum, nach dem das Objekt nicht mehr für allgemeine Benutzer sichtbar ist, obwohl es sein Eigentümer und der Repository-Administrator noch sehen können. Zum Festlegen des Ablaufdatums wählen Sie die Option **Datum** und geben Sie das Datum ein oder wählen Sie ein Datum mithilfe der Schaltfläche "Kalender".

Speichern. Speichert das Objekt am aktuellen Speicherort.

Hinweis: Die Informationen in den Feldern **Beschreibung** und **Schlüsselwörter** werden getrennt von den Angaben in SPSS Modeler auf der Registerkarte "Anmerkungen" des Objekts behandelt. Eine Repository-Suche nach Beschreibung oder Schlüsselwort liefert keine Informationen von der Registerkarte "Anmerkungen". Weitere Informationen finden Sie im Thema „Suchen nach Objekten im Repository“ auf Seite 134.

Zuweisen von Themen zu einem gespeicherten Objekt

Themen sind ein hierarchisches Klassifizierungssystem für den Inhalt, der im Repository gespeichert ist. Sie können beim Speichern von Objekten aus den verfügbaren Themen auswählen und Benutzer können auch Objekte nach Thema suchen. Die Liste verfügbarer Themen wird von Repository-Benutzern mit den geeigneten Berechtigungen festgelegt. (Weitere Informationen finden Sie im Benutzerhandbuch zu Deployment Manager).

Über die Registerkarte "Themen" im Dialogfeld "Repository: Speichern" können Sie dem Objekt ein Thema zuweisen:

1. Klicken Sie auf die Schaltfläche **Hinzufügen**.
2. Klicken Sie in der Liste verfügbarer Themen auf einen Themennamen.
3. Klicken Sie auf **OK**.

So entfernen Sie eine Themenzuweisung:

4. Wählen Sie das Thema in der Liste der zugewiesenen Themen aus.
5. Klicken Sie auf **Löschen**.

Festlegen der Sicherheitsoptionen für gespeicherte Objekte

Auf der Registerkarte "Sicherheit" des Dialogfelds "Repository: Speichern" können Sie eine Reihe von Sicherheitsoptionen für ein gespeichertes Objekt festlegen. Für ein oder mehrere **Prinzipale** (d. h. Benutzer oder Benutzergruppen) können Sie:

- dem Objekt Zugriffsrechte zuweisen
- Zugriffsrechte auf das Objekt ändern
- Zugriffsrechte auf das Objekt entfernen

Prinzipal. Der Repository-Benutzername des Benutzers oder der Gruppe mit Zugriffsrechten auf dieses Objekt.

Berechtigungen. Die Zugriffsrechte dieses Benutzers bzw. dieser Gruppe auf das Objekt.

Hinzufügen. Ermöglicht es Ihnen, einen oder mehrere Benutzer zu der Liste der Benutzer mit Zugriffsrechten auf dieses Objekt hinzuzufügen. Weitere Informationen finden Sie im Thema „Hinzufügen eines Benutzers zur Berechtigungsliste“.

Ändern. Ermöglicht es Ihnen, die Zugriffsrechte des ausgewählten Benutzers bzw. der ausgewählten Gruppe auf dieses Objekt zu ändern. Lesezugriff wird standardmäßig gewährt. Mithilfe dieser Option können Sie zusätzliche Zugriffsrechte gewähren, nämlich "Eigentümer", "Schreiben", "Löschen" und "Ändern".

Löschen. Löscht den ausgewählten Benutzer bzw. die ausgewählte Gruppe aus der Berechtigungsliste für dieses Objekt.

Hinzufügen eines Benutzers zur Berechtigungsliste: Folgende Felder sind verfügbar, wenn Sie auf der Registerkarte "Sicherheit" des Dialogfelds "Repository: Speichern" die Option **Hinzufügen** auswählen:

Anbieter auswählen. Wählen Sie zur Authentifizierung einen Sicherheitsanbieter aus. Das Repository kann für die Verwendung verschiedener Sicherheitsanbieter konfiguriert werden. Wenn Sie mehr Hilfe benötigen, wenden Sie sich an Ihren lokalen Administrator.

Suchen. Geben Sie den Repository-Benutzernamen des Benutzers bzw. der Gruppe ein, die Sie hinzufügen möchten, und klicken Sie auf **Suchen**, um diesen Namen in der Benutzerliste anzuzeigen. Um mehrere Benutzernamen gleichzeitig hinzuzufügen, lassen Sie dieses Feld leer und klicken Sie einfach auf **Suchen**, um eine Liste mit allen Repository-Benutzernamen anzuzeigen.

Benutzerliste. Wählen Sie einen oder mehrere Benutzernamen aus der Liste und klicken Sie auf OK, um sie der Berechtigungsliste hinzuzufügen.

Ändern der Zugriffsrechte für ein Objekt: Folgende Felder sind verfügbar, wenn Sie auf der Registerkarte "Sicherheit" des Dialogfelds "Repository: Speichern" die Option **Ändern** auswählen.

Eigentümer. Wählen Sie diese Option aus, um diesem Benutzer bzw. dieser Gruppe Zugriffsrechte auf das Objekt zu geben. Der Eigentümer verfügt über die kompletten Zugriffsrechte auf das Objekt einschließlich Lösch- und Änderungsrechten.

Lesen. Standardmäßig hat ein Benutzer bzw. eine Gruppe, die nicht Objekteigentümer ist, nur Lesezugriff für das Objekt. Markieren Sie die geeigneten Kontrollkästchen für Schreiben, Löschen und Ändern, um die entsprechenden Zugriffsrechte für diesen Benutzer bzw. diese Gruppe hinzuzufügen.

Speichern von Streams

Sie können einen Stream als *.str*-Datei im Repository speichern, von wo andere Benutzer darauf zugreifen können.

Hinweis: Weitere Informationen über das Bereitstellen eines Streams, um zusätzliche Repository-Funktionen zu nutzen, finden Sie unter „Bereitstellen von Streams“ auf Seite 139.

So speichern Sie den aktuellen Stream:

1. Klicken Sie im Hauptmenü auf Folgendes:
Datei > Speichern > Als Stream speichern...
2. Geben Sie gegebenenfalls Einstellungen für die Verbindung mit dem Repository ein. Weitere Informationen finden Sie im Thema „Verbinden mit dem Repository“ auf Seite 127. Wenden Sie sich für spezielle Port-, Kennwort- und andere Verbindungsdaten an den lokalen Systemadministrator.
3. Wählen Sie im Dialogfeld "Repository: Speichern" den Ordner aus, in dem Sie das Objekt speichern möchten, geben Sie alle weiteren Informationen an, die Sie aufzeichnen möchten, und klicken Sie auf die Schaltfläche **Speichern**. Weitere Informationen finden Sie im Thema „Festlegen der Objekteigenschaften“ auf Seite 128.

Temporäres Speichern von Projekten

Sie können ein vollständiges IBM SPSS Modeler-Projekt als *.cpj*-Datei im Repository speichern, damit andere Benutzer darauf zugreifen können.

Da eine Projektdatei ein Container für andere IBM SPSS Modeler-Objekte ist, müssen Sie angeben, wo IBM SPSS Modeler die Objekte des Projekts im Repository speichern soll. Dies geschieht über eine Einstellung im Dialogfeld "Projekteigenschaften". Weitere Informationen finden Sie im Thema „Festlegen der Projekteigenschaften“ auf Seite 152.

Sobald Sie ein Projekt zum Speichern von Objekten im Repository konfiguriert haben, werden Sie immer, wenn Sie ein neues Objekt zum Projekt hinzufügen, automatisch von IBM SPSS Modeler aufgefordert, das Objekt zu speichern.

Nach Abschluss der IBM SPSS Modeler-Sitzung müssen Sie eine neue Version der Projektdatei speichern, damit Ihre Änderungen erhalten bleiben. Die Projektdatei enthält automatisch die aktuellsten Versionen ihrer Objekte (und ruft sie ab). Wenn Sie einem Projekt während einer IBM SPSS Modeler-Sitzung keine Objekte hinzugefügt haben, brauchen Sie die Projektdatei nicht wiederherzustellen. Sie müssen jedoch neuere Versionen der Projektobjekte (Streams, Ausgaben usw.), die Sie geändert haben, temporär speichern.

So speichern Sie ein Projekt:

1. Wählen Sie das Projekt auf der Registerkarte "CRISP-DM" oder "Klassen" im Managerbereich in IBM SPSS Modeler aus. Klicken Sie dann im Hauptmenü auf:
Datei > Projekt > Projekt temporär speichern...
2. Geben Sie gegebenenfalls Einstellungen für die Verbindung mit dem Repository ein. Weitere Informationen finden Sie im Thema „Verbinden mit dem Repository“ auf Seite 127. Wenden Sie sich für spezielle Port-, Kennwort- und andere Verbindungsdaten an den lokalen Systemadministrator.
3. Wählen Sie im Dialogfeld "Repository: Speichern" den Ordner aus, in dem Sie das Objekt speichern möchten, geben Sie alle weiteren Informationen an, die Sie aufzeichnen möchten, und klicken Sie auf die Schaltfläche **Speichern**. Weitere Informationen finden Sie im Thema „Festlegen der Projekteigenschaften“ auf Seite 128.

Speichern von Knoten

Sie können eine individuelle Knotendefinition aus dem aktuellen Stream als *.nod*-Datei im Repository speichern, von wo andere Benutzer darauf zugreifen können.

So speichern Sie einen Knoten:

1. Klicken Sie im Streamerstellungsbereich mit der rechten Maustaste auf den Knoten und klicken Sie auf **Knoten speichern**.
2. Geben Sie gegebenenfalls Einstellungen für die Verbindung mit dem Repository ein. Weitere Informationen finden Sie im Thema „Verbinden mit dem Repository“ auf Seite 127. Wenden Sie sich für spezielle Port-, Kennwort- und andere Verbindungsdaten an den lokalen Systemadministrator.
3. Wählen Sie im Dialogfeld "Repository: Speichern" den Ordner aus, in dem Sie das Objekt speichern möchten, geben Sie alle weiteren Informationen an, die Sie aufzeichnen möchten, und klicken Sie auf die Schaltfläche **Speichern**. Weitere Informationen finden Sie im Thema „Festlegen der Projekteigenschaften“ auf Seite 128.

Speichern von Ausgabeobjekten

Sie können ein Ausgabeobjekt aus dem aktuellen Stream als *.cou*-Datei im Repository speichern, von wo andere Benutzer darauf zugreifen können.

So speichern Sie ein Ausgabeobjekt:

1. Klicken Sie auf der Registerkarte "Ausgaben" im Managerbereich in SPSS Modeler auf das Objekt und klicken Sie dann im Hauptmenü auf:
Datei > Ausgaben > Ausgabe speichern...
2. Klicken Sie als Alternative mit der rechten Maustaste auf ein Objekt auf der Registerkarte "Ausgaben" und klicken Sie auf **Speichern**.
3. Geben Sie gegebenenfalls Einstellungen für die Verbindung mit dem Repository ein. Weitere Informationen finden Sie im Thema „Verbinden mit dem Repository“ auf Seite 127. Wenden Sie sich für spezielle Port-, Kennwort- und andere Verbindungsdaten an den lokalen Systemadministrator.
4. Wählen Sie im Dialogfeld "Repository: Speichern" den Ordner aus, in dem Sie das Objekt speichern möchten, geben Sie alle weiteren Informationen an, die Sie aufzeichnen möchten, und klicken Sie auf die Schaltfläche **Speichern**. Weitere Informationen finden Sie im Thema „Festlegen der Objekteigenschaften“ auf Seite 128.

Speichern von Modellen und Modellpaletten

Sie können ein einzelnes Modell als *.gm*-Datei im Repository speichern, von wo andere Benutzer darauf zugreifen können. Sie können auch den vollständigen Inhalt der Modellpalette als *.gen*-Datei im Repository speichern.

Speichern eines Modells:

1. Klicken Sie in der Modellpalette in SPSS Modeler auf das Objekt und klicken Sie dann im Hauptmenü auf:
Datei > Modelle > Modell speichern...
2. Klicken Sie als Alternative mit der rechten Maustaste auf ein Objekt in der Modellpalette und klicken Sie dann auf **Modell speichern**.
3. Fahren Sie mit "Abschließen des Speichervorgangs" fort.

Speichern einer Modellpalette:

1. Klicken Sie mit der rechten Maustaste in den Hintergrund der Modellpalette.
2. Klicken Sie im Popup-Menü auf **Palette speichern**.
3. Fahren Sie mit "Abschließen des Speichervorgangs" fort.

Abschließen des Speichervorgangs:

1. Geben Sie gegebenenfalls Einstellungen für die Verbindung mit dem Repository ein. Weitere Informationen finden Sie im Thema „Verbinden mit dem Repository“ auf Seite 127. Wenden Sie sich für spezielle Port-, Kennwort- und andere Verbindungsdaten an den lokalen Systemadministrator.
2. Wählen Sie im Dialogfeld "Repository: Speichern" den Ordner aus, in dem Sie das Objekt speichern möchten, geben Sie alle weiteren Informationen an, die Sie aufzeichnen möchten, und klicken Sie auf die Schaltfläche **Speichern**. Weitere Informationen finden Sie im Thema „Festlegen der Objekteigenschaften“ auf Seite 128.

Abrufen von Objekten aus dem Repository

Sie können Streams, Modelle, Modellpaletten, Knoten, Projekte und Ausgabeobjekte abrufen, die temporär im Repository gespeichert wurden.

Hinweis: Sie können neben der hier beschriebenen Verwendung der Menüoptionen diese auch nutzen, um Streams, Ausgabeobjekte, Modelle und Modellpaletten abzurufen, indem Sie mit der rechten Maustaste in die jeweilige Registerkarte des Managerbereichs oben rechts im SPSS Modeler-Fenster klicken.

1. Zum Abrufen eines Streams klicken Sie im IBM SPSS Modeler-Hauptmenü auf:
Datei > Stream abrufen...

2. Um ein Modell, eine Modellpalette, ein Projekt oder ein Ausgabeobjekt abzurufen, klicken Sie im IBM SPSS Modeler-Hauptmenü auf:
 - Datei > Modelle > Modell abrufen...**
ODER
 - Datei > Modelle > Modellpalette abrufen...**
ODER
 - Datei > Projekte > Projekt abrufen...**
ODER
 - Datei > Ausgaben > Ausgabe abrufen...**
3. Klicken Sie als Alternative mit der rechten Maustaste in den Manager- oder Projektbereich und klicken Sie im Popup-Menü auf **Abrufen**.
4. Zum Abrufen eines Knotens klicken Sie im IBM SPSS Modeler-Hauptmenü auf:
 - Einfügen > Knoten (oder Superknoten) aus Repository...**
 - a. Geben Sie gegebenenfalls Einstellungen für die Verbindung mit dem Repository ein. Weitere Informationen finden Sie im Thema „Verbinden mit dem Repository“ auf Seite 127. Wenden Sie sich für spezielle Port-, Kennwort- und andere Verbindungsdaten an den lokalen Systemadministrator.
5. Navigieren Sie im Dialogfeld "Repository: Abrufen" zum Objekt, wählen Sie es aus und klicken Sie auf die Schaltfläche **Abrufen**. Weitere Informationen finden Sie im Thema .

Auswählen eines abzurufenden Objekts

Folgende Felder sind im Dialogfeld "Repository: Abrufen/Suchen" verfügbar:

Suchen in: Zeigt die Ordnerhierarchie für den aktuellen Ordner. Um zu einem anderen Ordner zu navigieren, wählen Sie einen aus dieser Liste aus, um direkt zu dem Ordner zu wechseln, oder navigieren Sie mithilfe der Objektliste unter diesem Feld.

Schaltfläche "Ordner nach oben": Wechselt in der Hierarchie auf die Ebene über dem aktuellen Ordner.

Schaltfläche "Neuer Ordner": Erstellt auf der aktuellen Ebene in der Hierarchie einen neuen Ordner.

Dateiname: Der Repository-Dateiname des ausgewählten Objekts. Klicken Sie zum Abrufen dieses Objekts auf **Abrufen**.

Dateien des Typs: Der Typ des Objekts, das Sie abrufen möchten. Nur Objekte dieses Typs werden zusammen mit Ordnern in der Objektliste angezeigt. Um Objekte eines anderen Typs zum Abrufen anzuzeigen, wählen Sie den Objekttyp aus der Liste aus.

Gesperrt öffnen: Ein Objekt wird beim Abrufen standardmäßig im Repository gesperrt, damit andere es nicht aktualisieren können. Wenn das Objekt beim Abrufen nicht gesperrt werden soll, inaktivieren Sie dieses Kontrollkästchen.

Beschreibung, Schlüsselwörter: Wenn beim Speichern des Objekts zusätzliche Details über das Objekt definiert wurden, werden diese Details hier angezeigt. Weitere Informationen finden Sie im Thema „Hinzufügen von Informationen über gespeicherte Objekte“ auf Seite 128.

Version: Um eine andere Objektversion als die aktuellste abzurufen, klicken Sie auf diese Schaltfläche. Es werden detaillierte Informationen für alle Versionen angezeigt, damit Sie die gewünschte Version auswählen können.

Auswählen einer Objektversion

So wählen Sie im Dialogfeld "Repository: Version auswählen" eine bestimmte Version eines Repository-Objekts aus:

1. (Optional) Sortieren Sie die Liste nach Version, Beschriftung, Größe, Erstellungsdatum oder Ersteller, indem Sie auf den Titel der entsprechenden Spalte doppelklicken.
2. Wählen Sie die Objektversion aus, mit der Sie arbeiten möchten.
3. Klicken Sie auf "Weiter".

Suchen nach Objekten im Repository

Sie können Objekten anhand von Namen, Ordner, Typ, Beschriftung, Datum oder anderen Kriterien suchen.

Suchen nach Name

So suchen Sie ein Objekt nach Name:

1. Klicken Sie im IBM SPSS Modeler-Hauptmenü auf:
Tools > Repository > Explorative Datenanalyse...
 - a. Geben Sie gegebenenfalls Einstellungen für die Verbindung mit dem Repository ein. Weitere Informationen finden Sie im Thema „Verbinden mit dem Repository“ auf Seite 127. Wenden Sie sich für spezielle Port-, Kennwort- und andere Verbindungsdaten an den lokalen Systemadministrator.
2. Klicken Sie auf die Registerkarte **Suchen**.
3. Geben Sie im Feld **Suchen nach Objekten mit Namen** den Namen des zu suchenden Objekts an.

Wenn Sie anhand des Namens nach Objekten suchen, kann ein Stern (*) als Platzhalterzeichen für eine beliebige Zeichenfolge oder ein Fragezeichen (?) als Platzhalterzeichen für ein einzelnes Zeichen verwendet werden. So werden beispielsweise mit **cluster** alle Objekte gefunden, die irgendwo im Namen die Zeichenfolge *cluster* enthalten. Mit der Suchzeichenfolge *m0?_** werden beispielsweise die Namen *M01_cluster.str* und *M02_cluster.str*, nicht jedoch *M01a_cluster.str* gefunden. Bei der Suche wird nicht zwischen Groß- und Kleinschreibung unterschieden (*cluster* entspricht *Cluster* und *CLUSTER*).

Hinweis: Bei einer großen Anzahl von Objekten kann der Suchvorgang einige Zeit in Anspruch nehmen.

Suchen nach anderen Kriterien

Sie können die Suche auf der Basis von Titel, Beschriftung, Datum, Autor, Schlüsselwort, Indexinhalt oder Beschreibung ausführen. Es werden lediglich Objekte gefunden, die *allen* angegebenen Suchkriterien entsprechen. Sie könnten beispielsweise alle Streams suchen, die mindestens ein Clustering-Modell enthalten und auf die außerdem eine bestimmte Beschriftung angewendet wurde und die nach einem bestimmten Datum bearbeitet wurden.

Objekttypen. Sie können die Suche auf Modelle, Streams, Ausgaben, Knoten, Superknoten, Projekte, Modellpaletten, Szenarios oder andere Objekttypen einschränken.

- **Modelle.** Sie können anhand der Kategorie (Klassifizierung, Näherung, Clusterbildung usw.) eines bestimmten Modellierungsalgorithmus, wie beispielsweise Kohonen, nach Modellen suchen.
Sie können die Suche auch nach verwendeten Feldern durchführen, z. B. nach allen Modellen, die ein Feld *Einkommen* als Eingabe- oder Ausgabefeld (Zielfeld) verwenden.
- **Streams.** Für Streams können Sie die Suche nach verwendeten Feldern oder Modelltyp (Kategorie oder Algorithmus) im Stream beschränken.

Themen. Sie können nach Modellen suchen, die bestimmten Themen aus einer Liste zugeordnet sind, die von Repository-Benutzern mit den entsprechenden Berechtigungen festgelegt wurde. (Weitere Informationen finden Sie im Benutzerhandbuch zu *Deployment Manager*). Um die Liste abzurufen, aktivieren Sie dieses Kontrollkästchen. Klicken Sie dann auf die Schaltfläche "Themen hinzufügen", die eingeblendet wird, wählen Sie ein oder mehrere Themen in der Liste aus und klicken Sie auf "OK".

Beschriftung. Beschränkt die Suche auf bestimmte Objektversionsbeschriftungen.

Daten. Sie können ein Erstellungs- oder Änderungsdatum angeben und nach Objekten vor, nach oder innerhalb des angegebenen Datumsbereichs suchen.

Autor. Beschränkt die Suche auf Objekte, die von einem bestimmten Benutzer erstellt wurden.

Schlüsselwörter. Suche nach bestimmten Schlüsselwörtern. In IBM SPSS Modeler werden Schlüsselwörter auf der Registerkarte "Anmerkung" für einen Stream, ein Modell oder ein Ausgabeobjekt angegeben.

Beschreibung. Suche nach bestimmten Termen im Beschreibungsfeld. In IBM SPSS Modeler wird die Beschreibung auf der Registerkarte "Anmerkung" für einen Stream, ein Modell oder ein Ausgabeobjekt angegeben. Mehrere Suchausdrücke können durch Semikolons voneinander getrennt werden, beispielsweise Einkommen; Feldfruchttyp; Subventionshöhe. (Beachten Sie, dass innerhalb von Suchausdrücken, Leerzeichen von Bedeutung sind. Zum Beispiel sind Höhe der Subventionen mit einem Leerzeichen vor "der" und Höhe der Subventionen mit zwei Leerzeichen vor "der" nicht dasselbe.)

Ändern von Repository-Objekten

Sie können bestehende Objekte im Repository direkt von SPSS Modeler aus ändern. Sie verfügen über folgende Möglichkeiten:

- Erstellen, Umbenennen oder Löschen von Ordnern
- Sperren oder Entsperren von Objekten
- Löschen von Objekten

Erstellen, Umbenennen und Löschen von Ordnern

1. Um Operationen an Ordnern im Repository auszuführen, klicken Sie im SPSS Modeler-Hauptmenü auf:

Tools > Repository > Explorative Datenanalyse...

- a. Geben Sie gegebenenfalls Einstellungen für die Verbindung mit dem Repository ein. Weitere Informationen finden Sie im Thema „Verbinden mit dem Repository“ auf Seite 127. Wenden Sie sich für spezielle Port-, Kennwort- und andere Verbindungsdaten an den lokalen Systemadministrator.
2. Vergewissern Sie sich, dass die Registerkarte **Ordner** aktiv ist.
 3. Um einen neuen Ordner zu erstellen, klicken Sie mit der rechten Maustaste auf den übergeordneten Ordner und klicken Sie dann auf **Neuer Ordner**.
 4. Um einen Ordner umzubenennen, klicken Sie mit der rechten Maustaste darauf und klicken Sie dann auf **Ordner umbenennen**.
 5. Um einen Ordner zu löschen, klicken Sie mit der rechten Maustaste darauf und klicken Sie dann auf **Ordner löschen**.

Sperren und Entsperren von Repository-Objekten

Sie können ein Objekt sperren, um zu verhindern, dass andere Benutzer seine bestehenden Versionen aktualisieren oder neue Versionen erstellen. Ein gesperrtes Objekt wird durch ein Vorhängeschlosssymbol auf dem Objektsymbol gekennzeichnet.



Abbildung 15. Gesperrtes Objekt

So sperren Sie ein Objekt:

1. Klicken Sie im Repository-Explorer-Fenster mit der rechten Maustaste auf das gewünschte Objekt.
2. Klicken Sie auf **Sperren**.

So entsperren Sie ein Objekt:

1. Klicken Sie im Repository-Explorer-Fenster mit der rechten Maustaste auf das gewünschte Objekt.
2. Klicken Sie auf **Entsperren**.

Löschen von Repository-Objekten

Bevor Sie ein Objekt aus dem Repository löschen, müssen Sie entscheiden, ob Sie alle Versionen oder nur eine bestimmte Version des Objekts löschen möchten.

So löschen Sie alle Versionen eines Objekts:

1. Klicken Sie im Repository-Explorer-Fenster mit der rechten Maustaste auf das gewünschte Objekt.
2. Klicken Sie auf **Objekte löschen**.

So löschen Sie die aktuellste Version eines Objekts:

1. Klicken Sie im Repository-Explorer-Fenster mit der rechten Maustaste auf das gewünschte Objekt.
2. Klicken Sie auf **Löschen**.

So löschen Sie eine frühere Version eines Objekts:

1. Klicken Sie im Repository-Explorer-Fenster mit der rechten Maustaste auf das gewünschte Objekt.
2. Klicken Sie auf **Versionen löschen**.
3. Wählen Sie die zu löschenden Versionen aus und klicken Sie auf **OK**.

Verwalten der Eigenschaften von Repository-Objekten

Sie können zahlreiche Objekteigenschaften von SPSS Modeler aus steuern. Sie verfügen über folgende Möglichkeiten:

- Anzeigen der Ordneigenschaften
- Anzeigen und Bearbeiten von Objekteigenschaften
- Erstellen, Zuweisen und Löschen von Versionsbeschriftungen für ein Objekt

Anzeigen von Ordneigenschaften

Um Eigenschaften für einen beliebigen Ordner im Repository-Fenster anzuzeigen, klicken Sie mit der rechten Maustaste auf den gewünschten Ordner. Klicken Sie auf **Ordneigenschaften**.

Registerkarte "Allgemein"

Diese Registerkarte zeigt den Ordnernamen, das Erstellungsdatum und das Änderungsdatum an.

Registerkarte "Berechtigungen"

Auf dieser Registerkarte geben Sie Lese- und Schreibberechtigungen für den Ordner an. Es werden alle Benutzer und Gruppen mit Zugriff auf den übergeordneten Ordner aufgelistet. Es gibt eine Hierarchie der Berechtigungen. Wenn Sie beispielsweise nicht über eine Leseberechtigung verfügen, können Sie sie auch nicht über eine Schreibberechtigung verfügen. Wenn Sie nicht über eine Schreibberechtigung verfügen, können Sie auch nicht über eine Löschberechtigung verfügen.

Benutzer und Gruppen. Listet die Repository-Benutzer und -Gruppen auf, die mindestens über Lesezugriff für diesen Ordner verfügen. Aktivieren Sie die Kontrollkästchen für Schreiben und Löschen, um die entsprechenden Zugriffsrechte für einen bestimmten Benutzer bzw. eine bestimmte Gruppe hinzuzufügen. Klicken Sie auf das Symbol **Benutzer/Gruppen hinzufügen** rechts auf der Registerkarte "Berechtigungen", um weiteren Benutzern und Gruppen den Zugriff zuzuweisen. Die Liste mit den verfügbaren Benutzern und Gruppen wird vom Administrator gesteuert.

Berechtigungen weiterleiten. Wählen Sie eine Option aus, um zu steuern, wie Änderungen am aktuellen Ordner seinen untergeordneten Ordnern (falls vorhanden) zugewiesen werden.

- **Alle Berechtigungen weiterleiten.** Leitet die Berechtigungseinstellungen vom aktuellen Ordner zu allen untergeordneten Ordnern weiter. Dies stellt eine schnelle Möglichkeit dar, Berechtigungen für mehrere Ordner gleichzeitig festzulegen. Legen Sie die Berechtigungen für den übergeordneten Ordner wie gewünscht fest und leiten Sie diese dann nach Bedarf weiter.
- **Nur Änderungen weiterleiten.** Leitet nur Änderungen weiter, die seit den letzten Änderungen vorgenommen wurden. Wenn z. B. eine neue Gruppe hinzugefügt wurde und sie dieser Gruppe Zugriff auf alle Ordner unter "Umsatz" zuweisen möchten, können Sie der Gruppe Zugriff auf den Stammordner "Umsatz" geben und die Änderung an alle Unterordner weiterleiten. Alle anderen Berechtigungen für bereits vorhandene Unterordner bleiben wie zuvor bestehen.
- **Nicht weiterleiten.** Alle vorgenommenen Änderungen gelten lediglich für den aktuellen Ordner und werden nicht an die untergeordneten Ordner weitergeleitet.

Anzeigen und Bearbeiten von Objekteigenschaften

Im Dialogfeld "Objekteigenschaften" können Sie Eigenschaften anzeigen und bearbeiten. Zwar können einige Eigenschaften nicht geändert werden, Sie können jedoch immer ein Objekt aktualisieren, indem Sie eine neue Version hinzufügen.

1. Klicken Sie im Repository-Fenster mit der rechten Maustaste auf das gewünschte Objekt.
2. Klicken Sie auf **Objekteigenschaften**.

Registerkarte "Allgemein"

Name. Der Name des Objekts, wie im Repository angezeigt.

Erstellt am: Datum, an dem das Objekt (nicht die Version) erstellt wurde.

Zuletzt geändert am: Datum, an dem die aktuellste Version geändert wurde.

Autor. Der Anmelde-name des Benutzers.

Beschreibung. Standardmäßig enthält diese Option die Beschreibung, die in SPSS Modeler auf der Registerkarte "Anmerkung" angegeben ist.

Verknüpfte Themen. Mit dem Repository können Modelle und verwandte Objekte nach Themen geordnet werden. Die Liste verfügbarer Themen wird von Repository-Benutzern mit den geeigneten Berechtigungen festgelegt. (Weitere Informationen finden Sie im Benutzerhandbuch zu Deployment Manager).

Schlüsselwörter. Schlüsselwörter werden auf der Registerkarte "Anmerkung" für einen Stream, ein Modell oder ein Ausgabeobjekt angegeben. Mehrere Schlüsselwörter sollten durch Leerzeichen voneinander getrennt werden, insgesamt sind bis zu 255 Zeichen zulässig. (Wenn Schlüsselwörter Leerzeichen enthalten, müssen Sie zur Trennung Ausrufezeichen verwenden.)

Registerkarte "Versionen"

Die im Repository gespeicherten Objekte weisen möglicherweise mehrere Versionen auf. Auf der Registerkarte "Versionen" werden Informationen über jede Version angezeigt.

Die folgenden Eigenschaften können für bestimmte Versionen eines gespeicherten Objekts angegeben oder geändert werden:

Version. Eindeutige ID für die Version, generiert basierend auf der Uhrzeit, zu der die Version gespeichert wurde.

Beschriftung. Aktuelle Beschriftung für die Version, falls vorhanden. Im Gegensatz zur Versions-ID können die Beschriftungen von einer Version eines Objekts zu einer anderen verschoben werden.

Die Dateigröße, das Erstellungsdatum und der Autor werden ebenfalls für jede Version angezeigt.

Beschriftungen bearbeiten. Klicken Sie auf das Symbol **Beschriftungen bearbeiten** rechts oben auf der Registerkarte "Versionen", um Beschriftungen für gespeicherte Objekte zu definieren, anzuwenden oder zu entfernen. Weitere Informationen finden Sie im Thema „Verwalten von Objektversionsbeschriftungen“.

Registerkarte "Berechtigungen"

Auf der Registerkarte **Berechtigungen** können Sie Lese- und Schreibberechtigungen für das Objekt festlegen. Es werden alle Benutzer und Gruppen mit Zugriff auf das aktuelle Objekt aufgelistet. Es gibt eine Hierarchie der Berechtigungen. Wenn Sie beispielsweise nicht über eine Leseberechtigung verfügen, können Sie sie auch nicht über eine Schreibberechtigung verfügen. Wenn Sie nicht über eine Schreibberechtigung verfügen, können Sie auch nicht über eine Löschberechtigung verfügen.

Benutzer und Gruppen. Listet die Repository-Benutzer und -Gruppen auf, die mindestens über Lesezugriff für dieses Objekt verfügen. Aktivieren Sie die Kontrollkästchen für Schreiben und Löschen, um die entsprechenden Zugriffsrechte für einen bestimmten Benutzer bzw. eine bestimmte Gruppe hinzuzufügen. Klicken Sie auf das Symbol **Benutzer/Gruppen hinzufügen** rechts auf der Registerkarte "Berechtigungen", um weiteren Benutzern und Gruppen den Zugriff zuzuweisen. Die Liste mit den verfügbaren Benutzern und Gruppen wird vom Administrator gesteuert.

Verwalten von Objektversionsbeschriftungen

Im Dialogfeld "Versionsbeschriftungen bearbeiten" haben Sie folgende Möglichkeiten:

- Anwenden von Beschriftungen auf das ausgewählte Objekt
- Entfernen von Beschriftungen vom ausgewählten Objekt
- Definieren einer neuen Beschriftung und Anwenden auf das Objekt

So wenden Sie Beschriftungen auf ein Objekt an:

1. Wählen Sie in der Liste **Verfügbare Beschriftungen** eine oder mehrere Beschriftungen aus.
2. Klicken Sie auf die Schaltfläche mit dem Rechtspfeil, um die ausgewählten Beschriftungen in die Liste **Angewendete Beschriftungen** zu verschieben.
3. Klicken Sie auf **OK**.

So entfernen Sie Beschriftungen von einem Objekt:

1. Wählen Sie in der Liste **Angewendete Beschriftungen** eine oder mehrere Beschriftungen aus.
2. Klicken Sie auf die Schaltfläche mit dem Linkspfeil, um die ausgewählten Beschriftungen in die Liste **Verfügbare Beschriftungen** zu verschieben.
3. Klicken Sie auf **OK**.

So definieren Sie eine neue Beschriftung und wenden sie auf das Objekt an:

1. Geben Sie den Beschriftungsnamen in das Feld **Neue Beschriftung** ein.
2. Klicken Sie auf die Schaltfläche mit dem Rechtspfeil, um die neue Beschriftung in die Liste **Angewendete Beschriftungen** zu verschieben.
3. Klicken Sie auf **OK**.

Bereitstellen von Streams

Damit ein Stream mit der Thin-Client-Anwendung IBM SPSS Modeler Advantage verwendet werden kann, muss er als Stream (*.str*-Datei) im Repository bereitgestellt werden.

Unabhängig davon, ob ein Stream als Stream (*.str*-Datei) oder als Szenario (*.scn*-Datei) bereitgestellt wird, kann das Objekt die Unternehmensfunktionen von IBM SPSS Collaboration and Deployment Services vollständig nutzen. Weitere Informationen finden Sie im Thema „Speichern und Bereitstellen von Repository-Objekten“ auf Seite 126.

So stellen Sie den aktuellen Stream bereit (Verfahren mit dem Menü "Datei"):

1. Klicken Sie im Hauptmenü auf Folgendes:

Datei > Bereitstellen

2. Wählen Sie den Bereitstellungstyp aus und füllen Sie das übrige Dialogfeld wie erforderlich aus.
3. Klicken Sie auf **Als Stream bereitstellen**, um den Stream für die Verwendung mit IBM SPSS Modeler Advantage oder IBM SPSS Collaboration and Deployment Services bereitzustellen. Klicken Sie auf **Als Szenario bereitstellen**, um den Stream für die Verwendung mit IBM SPSS Collaboration and Deployment Services oder Predictive Applications Version 5 bereitzustellen.
4. Klicken Sie auf **Speichern**. Wenn Sie Informationen wünschen, klicken Sie auf **Hilfe**.
5. Fahren Sie mit "Abschließen des Bereitstellungsvorgangs" fort.

So stellen Sie den aktuellen Stream bereit (Verfahren mit dem Menü "Extras"):

1. Klicken Sie im Hauptmenü auf Folgendes:

Tools > Streameigenschaften > Bereitstellung

2. Wählen Sie den Bereitstellungstyp aus, füllen Sie den Rest der Registerkarte "Bereitstellung" wie erforderlich aus und klicken Sie auf **Speichern**. Weitere Informationen finden Sie im Thema „Bereitstellungsoptionen für den Stream“.

Abschließen des Bereitstellungsvorgangs

1. Geben Sie gegebenenfalls Einstellungen für die Verbindung mit dem Repository ein. Weitere Informationen finden Sie im Thema „Verbinden mit dem Repository“ auf Seite 127. Wenden Sie sich für spezielle Port-, Kennwort- und andere Verbindungsdaten an den lokalen Systemadministrator.
2. Wählen Sie im Dialogfeld "Repository: Speichern" den Ordner aus, in dem Sie das Objekt speichern möchten, geben Sie alle weiteren Informationen an, die Sie aufzeichnen möchten, und klicken Sie auf die Schaltfläche **Speichern**. Weitere Informationen finden Sie im Thema „Festlegen der Objekteigenschaften“ auf Seite 128.

Bereitstellungsoptionen für den Stream

Auf der Registerkarte "Bereitstellung" im Dialogfeld "Streamoptionen" können Sie Optionen für die Bereitstellung des Streams festlegen. Die Bereitstellung ist als Stream oder als Szenario möglich.

Beim Bereitstellen als Stream können Sie den Stream in der Thin-Client-Anwendung IBM SPSS Modeler Advantage öffnen und ändern. Der Stream wird im Repository als Datei mit der Erweiterung *.str* gespeichert.

Beim Bereitstellen als Szenario wird der Stream im Repository als Datei mit der Erweiterung *.scn* gespeichert. Damit kann ein Stream auch von Predictive Applications Version 5 verwendet werden.

Unabhängig davon, ob die Bereitstellung als Stream oder als Szenario erfolgt, können Sie die zusätzliche Funktionalität nutzen, die in IBM SPSS Collaboration and Deployment Services zur Verfügung steht, z. B. Mehrbenutzerzugriff, automatisches Scoring, Modellaktualisierung und Champion-Challenger-Analyse.

Über die Registerkarte "Bereitstellung" können Sie auch eine Vorschau der Streambeschreibung anzeigen, die IBM SPSS Modeler für den Stream erstellt. Weitere Informationen finden Sie im Thema „Streambeschreibungen“ auf Seite 55.

Hinweis: Um konsistenten Zugriff auf Unternehmensdaten sicherzustellen, muss ein Stream, der als Szenario bereitgestellt ist, über IBM SPSS Collaboration and Deployment Services Enterprise View auf seine Quelldaten zugreifen. In einem solchen Fall muss der Stream in jeder angegebenen Scoring- bzw. Modellierungsverzweigung mindestens einen Quellenknoten vom Typ "Enterprise-Ansicht" (Enterprise View) enthalten.

Bereitstellungstyp. Wählen Sie aus, wie der Stream bereitgestellt werden soll. Alle Streams benötigen einen speziellen Scoring-Knoten, bevor sie bereitgestellt werden können; weitere Voraussetzungen und Optionen hängen vom Bereitstellungstyp ab.

- **<keine>.** Der Stream wird nicht für das Repository bereitgestellt. Alle Optionen mit Ausnahme der Vorschau der Streambeschreibung sind inaktiviert.
- **Nur Scoring.** Der Stream wird dem Repository bereitgestellt, wenn Sie auf die Schaltfläche **Speichern** klicken. Daten können mithilfe des Knotens gescort werden, den Sie im Feld **Scoring-Knoten** angeben.
- **Modellaktualisierung.** Wie "Nur Scoring", jedoch kann das Modell zusätzlich im Repository mit Objekten aktualisiert werden, die Sie in den Feldern **Modellierungsknoten** und **Modellnugget** festlegen.
Hinweis: Automatische Modellaktualisierung wird in IBM SPSS Collaboration and Deployment Services nicht standardmäßig unterstützt. Daher müssen Sie diesen Bereitstellungstyp auswählen, wenn Sie diese Funktion beim Ausführen eines Streams aus dem Repository verwenden wollen. Weitere Informationen finden Sie im Thema „Modellaktualisierung“ auf Seite 142.

Scoring-Knoten. Wählen Sie ein Diagramm, einen Ausgabe- oder einen Exportknoten aus, um die Streamverzweigung zu identifizieren, die zum Scoring der Daten verwendet werden soll. Der Stream kann zwar eine beliebige Anzahl an gültigen Verzweigungen, Modellen und Endknoten enthalten, zu Bereitstellungszwecken muss jedoch genau eine Scoring-Verzweigung angegeben werden. Dies ist die grundlegendste Voraussetzung für die Bereitstellung eines Streams.

Scoring-Parameter. Dient zur Angabe von Parametern, die bei der Ausführung der Scoring-Verzweigung geändert werden können. Weitere Informationen finden Sie im Thema „Scoring und Modellierung von Parametern“ auf Seite 141.

Modellierungsknoten. Gibt für die Modellaktualisierung den Modellierungsknoten an, der zum Generieren bzw. zur Aktualisierung des Modells im Repository verwendet wird. Dies muss ein Modellierungsknoten desselben Typs sein wie für **Modellnugget** angegeben wurde.

Modellerstellungsparameter. Dient zur Angabe von Parametern, die bei der Ausführung des Modellierungsknotens geändert werden können. Weitere Informationen finden Sie im Thema „Scoring und Modellierung von Parametern“ auf Seite 141.

Modellnugget. Gibt zum Zwecke der Modellaktualisierung das Modellnugget an, das bei jeder Aktualisierung des Streams im Repository aktualisiert bzw. neu generiert werden soll (üblicherweise im Rahmen einer geplanten Aufgabe). Das Modell muss sich in der Scoring-Verzweigung befinden. In der Scoring-Verzweigung können mehrere Modelle vorhanden sein, es kann jedoch nur eines angegeben werden. Beachten Sie: Bei der erstmaligen Erstellung des Streams kann es sich hierbei im Grunde um ein Platzhaltermodell handeln, das immer dann aktualisiert bzw. neu generiert wird, wenn neue Daten verfügbar sind.

Als Stream bereitstellen. Klicken Sie auf diese Option, wenn Sie den Stream mit IBM SPSS Modeler Advantage oder IBM SPSS Collaboration and Deployment Services verwenden möchten (siehe auch den folgenden Hinweis).

Als Szenario bereitstellen. Klicken Sie auf diese Option, wenn Sie den Stream mit IBM SPSS Collaboration and Deployment Services oder Predictive Applications Version 5 verwenden möchten (siehe auch den folgenden Hinweis).

Überprüfen. Klicken Sie auf diese Schaltfläche, um zu prüfen, ob es sich um einen gültigen Stream zur Bereitstellung handelt. Alle Streams müssen über einen speziellen Scoring-Knoten verfügen, bevor sie bereitgestellt werden können. Wenn Sie die Bereitstellung als Szenario vornehmen, muss der Stream außerdem einen gültigen Enterprise View-Quellenknoten enthalten. Wenn diese Bedingungen nicht erfüllt sind, werden Fehlermeldungen angezeigt.

Speichern. Stellt den Stream bereit, sofern dieser gültig ist. Andernfalls wird eine Fehlermeldung angezeigt. Klicken Sie auf die Schaltfläche **Korrigieren**, um den Fehler zu korrigieren und es erneut zu versuchen.

Vorschau der Streambeschreibung anzeigen. Ermöglicht die Anzeige des Inhalts einer Streambeschreibung, die IBM SPSS Modeler für den Stream erstellt. Weitere Informationen finden Sie im Thema „Streambeschreibungen“ auf Seite 55.

Hinweis: (Bereitstellung als Stream oder als Szenario) Innerhalb der Modellierungsverzweigung können mehrere Enterprise-Ansichtsknoten verwendet werden. Ist dies der Fall, sollte in den meisten Fällen eine einzelne Datenverbindung für alle Enterprise-Ansichtsknoten innerhalb der Verzweigung verwendet werden; für Champion/Challenger-Analyse ist sie sogar erforderlich.

- Wenn keine Champion/Challenger-Unterstützung benötigt wird, können innerhalb derselben Verzweigung verschiedene Enterprise-Ansichtsverbindungen verwendet werden, vorausgesetzt, die Verbindungen unterscheiden sich nur nach Datenproviderdefinition (DPD).
- Diese Einschränkungen gelten jeweils nur innerhalb einer bestimmten Verzweigung. Zwischen den Scoring- und den Modellerstellungsverzweigungen können verschiedene Enterprise-Ansichtsverbindungen ohne derartige Einschränkungen verwendet werden.

Scoring und Modellierung von Parametern

Bei der Bereitstellung eines Streams für IBM SPSS Collaboration and Deployment Services können Sie auswählen, welche Parameter angezeigt bzw. bearbeitet werden können, wenn das Modell aktualisiert oder gescort wird. Sie können beispielsweise Mindest- und Höchstwerte oder einen anderen Wert angeben, der sich bei jeder Ausführung eines Jobs ändern kann.

1. Um einen Parameter sichtbar zu machen, sodass er nach der Bereitstellung des Streams angezeigt bzw. bearbeitet werden kann, müssen Sie ihn aus der Liste im Dialogfeld "Scoring-Parameter" auswählen.

Die Liste der verfügbaren Parameter wird im Dialogfeld für die Streameigenschaften auf der Registerkarte "Parameter" definiert. Weitere Informationen finden Sie im Thema „Festlegen der Stream- und Sitzungsparameter“ auf Seite 51.

Scoring-Verzweigung

Wenn Sie einen Stream bereitstellen, muss eine Verzweigung des Streams als **Scoring-Verzweigung** (d. h. die Verzweigung, die den Scoring-Knoten enthält) festgelegt werden. Wenn Sie eine Verzweigung als Scoring-Verzweigung festlegen, wird im Streamerstellungsbereich diese Verzweigung sowie die Modellverknüpfung zum Nugget auf der Scoring-Verzweigung hervorgehoben. Die visuelle Repräsentation ist besonders nützlich in komplexen Streams mit mehreren Verzweigungen, in denen die Scoring-Verzweigung nicht sofort erkennbar ist.

Hinweis: Nur eine Streamverzweigung kann als Scoring-Verzweigung festgelegt werden.

Wenn für den Stream bereits eine Scoring-Verzweigung definiert wurde, wird diese durch die neu festgelegte Verzweigung als Scoring-Verzweigung ersetzt. Sie können die Farbe der Scoring-Verzweigung mithilfe einer Option für benutzerdefinierte Farben festlegen. Weitere Informationen finden Sie im Thema „Festlegen von Anzeigeeoptionen“ auf Seite 160.

Sie können die Anzeige der Scoring-Verzweigung mithilfe der Symbolleistenschaltfläche zum Ein-/Ausblenden von Stream-Markups ein- und ausblenden.



Abbildung 16. Symbolleistenschaltfläche zum Ein-/Ausblenden von Stream-Markups

Angeben der Scoring-Verzweigung für die Bereitstellung

Sie können die Scoring-Verzweigung über das Popup-Menü eines Endknotens oder über das Menü "Extras" festlegen. Bei Verwendung des Popup-Menüs wird der Scoring-Knoten automatisch auf der Registerkarte "Bereitstellung" der Streameigenschaften festgelegt.

So legen Sie eine Verzweigung als die Scoring-Verzweigung fest (Popup-Menü):

1. Verbinden Sie das Modellnugget mit einem Endknoten (einem Verarbeitungs- oder Ausgabeknoten unterhalb des Nuggets).
2. Klicken Sie mit der rechten Maustaste auf den Endknoten.
3. Klicken Sie im Menü auf die Option **Als Scoring-Verzweigung verwenden**.

So legen Sie eine Verzweigung als die Scoring-Verzweigung fest (Menü "Extras"):

1. Verbinden Sie das Modellnugget mit einem Endknoten (einem Verarbeitungs- oder Ausgabeknoten unterhalb des Nuggets).
2. Klicken Sie im Hauptmenü auf Folgendes:
Tools > Streameigenschaften > Bereitstellung
3. Klicken Sie in der Liste **Bereitstellungstyp** wie gewünscht auf die Option **Nur Scoring** oder **Modellaktualisierung**. Weitere Informationen finden Sie im Thema „Bereitstellungsoptionen für den Stream“ auf Seite 139.
4. Klicken Sie auf das Feld **Scoring-Knoten** und wählen Sie einen Endknoten aus der Liste.
5. Klicken Sie auf **OK**.

Modellaktualisierung

Eine Modellaktualisierung ist der Vorgang, bei dem ein vorhandenes Modell in einem Stream anhand von neueren Daten neu aufgebaut wird. Der Stream selbst ändert sich im Repository nicht. Beispielsweise bleiben Algorithmustyp und streamspezifische Einstellungen gleich, aber das Modell wird mit neuen Daten neu trainiert und dann aktualisiert, wenn die neue Version des Modells besser als die alte Version funktioniert.

Nur ein Modellnugget in einem Stream kann auf Aktualisierung eingestellt werden. Dies ist das sogenannte **Aktualisierungsmodell**. Wenn Sie auf die Option **Modellaktualisierung** auf der Registerkarte "Bereitstellung" klicken (siehe „Bereitstellungsoptionen für den Stream“ auf Seite 139), wird das zu diesem Zeitpunkt festgelegte Modellnugget zu Ihrem Aktualisierungsmodell. Sie können auch aus dem Popup-Menü eines Modellnuggets ein Modell als Aktualisierungsmodell festlegen. Das Nugget muss sich dafür bereits auf der Scoring-Verzweigung befinden.

Wenn Sie den Modellaktualisierungsstatus eines Nuggets inaktivieren, hat dies dieselbe Wirkung wie das Einstellen des Bereitstellungstyps des Streams auf "Nur Scoring" und die Registerkarte "Bereitstellung"

des Dialogfelds für die Streameigenschaften wird entsprechend aktualisiert. Sie können diesen Status mit der Option **Als Aktualisierungsmodell verwenden** im Popup-Menü des Nuggets auf der aktuellen Scoring-Verzweigung ein- und ausschalten.

Beim Entfernen der Modellverknüpfung auf der Scoring-Verzweigung wird auch der Aktualisierungsmodellstatus des Nuggets entfernt. Sie können das Entfernen der Modellverknüpfung über das Menü "Bearbeiten" oder die Symbolleiste widerrufen, aber damit wird auch der Aktualisierungsmodellstatus wieder eingestellt.

Auswählen des Aktualisierungsmodells

Wie die Scoring-Verzweigung wird auch die Verknüpfung zum Aktualisierungsmodell im Stream hervorgehoben. Das als Aktualisierungsmodell gewählte Modellnugget und die hervorgehobene Verknüpfung hängen davon ab, wie viele Nuggets sich im Stream befinden.

Einzelnes Modell im Stream

Wenn ein einzelnes, verknüpftes Modellnugget auf einer Scoring-Verzweigung als solches identifiziert wird, wird dieses Nugget das Aktualisierungsmodell des Streams.

Mehrere Modelle im Stream

Falls der Stream mehrere verknüpfte Nuggets enthält, wird das Aktualisierungsmodell wie folgt ausgewählt.

Wenn ein Modellnugget auf der Registerkarte "Bereitstellung" des Dialogfelds für die Streameigenschaften definiert wurde und sich im Stream befindet, wird dieses Nugget zum Aktualisierungsmodell.

Wenn auf der Registerkarte "Bereitstellung" kein Nugget definiert wurde oder wenn einer definiert wurde, sich aber nicht auf der Scoring-Verzweigung befindet, wird das Nugget, das sich am nächsten zum Endknoten befindet, zum Aktualisierungsmodell.

Wenn Sie anschließend die Auswahl aller Modellverknüpfungen als Aktualisierungsverknüpfungen aufheben, wird nur die Scoring-Verzweigung hervorgehoben, nicht die Verknüpfungen. Der Bereitstellungstyp wird auf "Nur Scoring" eingestellt.

Hinweis: Sie können bei Bedarf für eine der Verknüpfungen den Status "Ersetzen" festlegen, nicht aber für die andere. In diesem Fall ist das als Aktualisierungsmodell gewählte Modellnugget dasjenige, das über die Aktualisierungsverknüpfung verfügt und sich am nächsten zum Endknoten befindet, wenn die Scoring-Verzweigung festgelegt wird.

Keine Modelle im Stream

Wenn der Stream keine Modelle oder nur Modelle ohne Modellverknüpfungen enthält, ist der Bereitstellungstyp auf "Nur Scoring" eingestellt.

Überprüfen einer Scoring-Verzweigung auf Fehler

Wenn Sie die Scoring-Verzweigung festlegen, wird sie auf Fehler überprüft, z. B. bei der Bereitstellung als Szenario auf das Fehlen eines Enterprise-Ansichtsknotens im Stream.

Bei Auffinden eines Fehlers wird die Scoring-Verzweigung in der entsprechenden Fehlerfarbe markiert und eine Fehlermeldung angezeigt. Sie können die Fehlerfarbe mithilfe einer Option für benutzerdefinierte Farben festlegen. Weitere Informationen finden Sie im Thema „Festlegen von Anzeigeeoptionen“ auf Seite 160.

Gehen Sie bei Auffinden eines Fehlers wie folgt vor:

1. Korrigieren Sie den Fehler gemäß dem Inhalt der Fehlermeldung.

2. Führen Sie eine der folgenden Aktionen aus:
 - Klicken Sie mit der rechten Maustaste auf den Endknoten und klicken Sie im Popup-Menü auf **Szenario prüfen**.
 - Klicken Sie im Hauptmenü auf Folgendes:
Tools > Streameigenschaften > Bereitstellung
und klicken Sie auf **Überprüfen**.
3. Wiederholen Sie diesen Vorgang, bis keine Fehler mehr gefunden werden.

Kapitel 10. Exportieren in externe Anwendungen

Exportieren in externe Anwendungen

IBM SPSS Modeler bietet eine Reihe von Mechanismen zum Exportieren des gesamten Data-Mining-Prozesses in externe Anwendungen, damit Sie die durchgeführten Arbeiten zur Vorbereitung von Daten und zum Erstellen von Modellen auch außerhalb von IBM SPSS Modeler gewinnbringend einsetzen können.

Im vorherigen Abschnitt wurde gezeigt, wie Sie Streams für ein Repository von IBM SPSS Collaboration and Deployment Services bereitstellen können, um dessen Mehrbenutzerzugriff, Aufgabenplanung und andere Funktionen zu nutzen. Ebenso lassen sich IBM SPSS Modeler-Streams zusammen mit folgenden Anwendungen nutzen:

- IBM SPSS Modeler Advantage
- Anwendungen, die Dateien im PMML-Format importieren und exportieren können

Weitere Informationen über das Verwenden von Streams in IBM SPSS Modeler Advantage finden Sie unter „Öffnen eines Streams in IBM SPSS Modeler Advantage“.

Informationen zum Exportieren und Importieren von Modellen als PMML-Dateien, um Modelle mit beliebigen anderen Anwendungen, die dieses Format unterstützen, gemeinsam zu nutzen, finden Sie unter „Importieren und Exportieren von Modellen als PMML“ auf Seite 146.

Öffnen eines Streams in IBM SPSS Modeler Advantage

IBM SPSS Modeler-Streams können zusammen mit der Thin-Client-Anwendung IBM SPSS Modeler Advantage verwendet werden. Es ist zwar möglich, benutzerdefinierte Anwendungen komplett in IBM SPSS Modeler Advantage zu erstellen, jedoch können Sie auch einen bereits in IBM SPSS Modeler erstellten Stream als Basis eines Anwendungsablaufs verwenden.

So öffnen Sie einen Stream in IBM SPSS Modeler Advantage:

1. Stellen Sie den Stream im Repository von IBM SPSS Collaboration and Deployment Services bereit, wobei die Option **Als Stream bereitstellen** aktiviert sein muss. Weitere Informationen finden Sie im Thema „Bereitstellen von Streams“ auf Seite 139.
2. Klicken Sie in der Symbolleiste auf die Schaltfläche "In IBM SPSS Modeler Advantage öffnen" oder klicken Sie im Hauptmenü auf:

Datei > In IBM SPSS Modeler Advantage öffnen

1. Geben Sie gegebenenfalls Einstellungen für die Verbindung mit dem Repository ein. Weitere Informationen finden Sie im Thema „Verbinden mit dem Repository“ auf Seite 127. Wenden Sie sich für spezielle Port-, Kennwort- und andere Verbindungsdaten an den lokalen Systemadministrator.

Hinweis: Auf dem Repository-Server muss die IBM SPSS Modeler Advantage-Software ebenfalls installiert sein.

1. Wählen Sie im Dialogfeld "Repository: Speichern" den Ordner aus, in dem Sie das Objekt speichern möchten, geben Sie alle weiteren Informationen an, die Sie aufzeichnen möchten, und klicken Sie auf die Schaltfläche **Speichern**. Weitere Informationen finden Sie im Thema „Festlegen der Objekteigenschaften“ auf Seite 128.

Dadurch wird IBM SPSS Modeler Advantage gestartet, wobei der Stream bereits geöffnet ist. Der Stream wird in IBM SPSS Modeler geschlossen.

Importieren und Exportieren von Modellen als PMML

PMML (Predictive Model Markup Language) ist ein XML-Format zur Beschreibung von Data-Mining-Modellen und statistischen Modellen, einschließlich der Eingaben zu den Modellen, der zur Vorbereitung der Daten für das Data-Mining verwendeten Transformationen sowie der Parameter, die die Modelle selbst definieren. IBM SPSS Modeler kann PMML importieren und exportieren, wodurch es möglich ist, Modelle mit anderen Anwendungen auszutauschen, die dieses Format unterstützen, wie beispielsweise IBM SPSS Statistics.

Weitere Informationen zu PMML finden Sie auf der Website der Data Mining Group (<http://www.dmg.org>).

So exportieren Sie ein Modell:

PMML-Export wird für die meisten der in IBM SPSS Modeler erstellten Modelltypen unterstützt. Weitere Informationen finden Sie im Thema „Modelltypen, die PMML unterstützen“.

1. Klicken Sie mit der rechten Maustaste auf ein Modellnugget in der Modellpalette. (Alternativ können Sie auf ein Modellnugget im Erstellungsbereich klicken und das Menü "Datei" auswählen.)
2. Klicken Sie im Menü auf **PMML exportieren**.
3. Geben Sie im Dialogfeld "Exportieren" (oder "Speichern") ein Zielverzeichnis und einen eindeutigen Namen für das Modell an.

Hinweis: Im Dialogfeld "Benutzeroptionen" können Sie die Optionen für den PMML-Export ändern. Klicken Sie im Hauptmenü auf Folgendes:

Tools > Optionen > Benutzeroptionen

Wählen Sie die Registerkarte "PMML" aus.

Weitere Informationen finden Sie im Thema „Festlegen von PMML-Exportoptionen“ auf Seite 161.

So importieren Sie ein als PMML gespeichertes Modell:

Modelle, die aus IBM SPSS Modeler oder einer anderen Anwendung als PMML exportiert wurden, können in die Modellpalette importiert werden. Weitere Informationen finden Sie im Thema „Modelltypen, die PMML unterstützen“.

1. Klicken Sie in der Modellpalette mit der rechten Maustaste auf die Palette und wählen Sie aus dem Menü die Option **PMML importieren** aus.
2. Wählen Sie die zu importierende Datei aus und geben Sie nach Bedarf Optionen für Variablenbeschriftungen an.
3. Klicken Sie auf **Öffnen**.

Verwenden Sie Variablenbeschriftungen, falls im Modell vorhanden. Im PMML-Code können sowohl die Variablennamen als auch die Variablenbeschriftungen (beispielsweise "Referrer ID" für *RefID*) für Variablen im Datenwörterbuch gefunden. Wählen Sie diese Option aus, um Variablenbeschriftungen zu verwenden, wenn diese im ursprünglich exportierten PMML-Code vorhanden sind.

Wenn Sie die Optionen zur Variablenbeschriftung ausgewählt haben, im PMML-Code jedoch keine Variablenbeschriftungen vorhanden sind, werden die Variablennamen wie üblich verwendet.

Modelltypen, die PMML unterstützen

PMML-Export

IBM SPSS Modeler-Modelle. Die folgenden in IBM SPSS Modeler erstellten Modelle können als PMML 4.0 exportiert werden:

- C&R-Baum
- QUEST
- CHAID
- Lineare Regression
- Netz
- C5.0
- Logistische Regression
- Genlin
- SVM
- Apriori
- Carma
- K-Means
- Kohonen
- Two Step
- GLMM (nur GLMM-Modelle mit festem Effekt werden unterstützt)
- Entscheidungsliste
- Cox
- Sequenz (Scoring für PMML-Sequenzmodelle wird nicht unterstützt)
- Statistics-Modell

Datenbankeigene Modelle. Bei Modellen, die mithilfe von datenbankeigenen Algorithmen generiert wurden, steht der PMML-Export nur bei IBM InfoSphere Warehouse-Modellen zur Verfügung. Modelle, die mithilfe von Analysis Services von Microsoft oder mit Oracle Data Miner erstellt wurden, können nicht exportiert werden. Beachten Sie außerdem, dass als PMML exportierte IBM Modelle danach nicht wieder in IBM SPSS Modeler importiert werden können.

PMML-Import

IBM SPSS Modeler kann PMML-Modelle importieren und scoren, die von aktuellen Versionen aller IBM SPSS Statistics-Produkte erstellt wurden, darunter Modelle, die aus IBM SPSS Modeler exportiert wurden, sowie Modell- bzw. Transformations-PMML, die von IBM SPSS Statistics 17.0 oder höher generiert wurde. Dies gilt also im Grunde für jegliche PMML, die die Scoring-Engine scoren kann - mit folgenden Ausnahmen:

- Apriori- und CARMA-Modelle sowie Anomalieerkennungsmo-
delle und Sequenzmodelle können nicht importiert werden.
- PMML-Modelle können nach dem Import in IBM SPSS Modeler nicht durchsucht werden, obwohl sie für das Scoring verwendet werden können. (Dies gilt auch für Modelle, die ursprünglich aus IBM SPSS Modeler exportiert wurden. Um diese Einschränkung zu vermeiden, sollten Sie das betreffende Modell als generierte Modelldatei (*.gm) und nicht als PMML exportieren.
- Als PMML exportierte IBM InfoSphere Warehouse-Modelle können nicht importiert werden.
- Eine eingeschränkte Validierung findet beim Import statt, aber die vollständige Validierung erfolgt beim Versuch, das Modell zu scoren. Daher kann der Import erfolgreich durchgeführt werden, das Scoring aber fehlschlagen oder falsche Ergebnisse erzeugen.

Kapitel 11. Projekte und Berichte

Einführung in Projekte

Ein **Projekt** ist eine Gruppe von Dateien, die in Bezug zu einer Data-Mining-Aufgabe stehen. Zu Projekten gehören Datenstreams, Diagramme, generierte Modelle, Berichte und alles andere, was Sie in IBM SPSS Modeler erstellt haben. Auf den ersten Blick kann der Eindruck entstehen, IBM SPSS Modeler-Projekte seien einfach eine Methode zur Organisation von Ausgaben, doch sie haben noch viele weitere Funktionen. Mit Projekten haben Sie folgende Möglichkeiten:

- Erstellen von Anmerkungen für jedes Objekt in der Projektdatei.
- Verwenden der CRISP-DM-Methodologie als Orientierung beim Data-Mining. Projekte können außerdem ein CRISP-DM-Hilfesystem enthalten, das Einzelheiten und reale Beispiele für Data-Mining mit CRISP-DM bietet.
- Hinzufügen von nicht aus IBM SPSS Modeler stammenden Objekten zum Projekt, beispielsweise einer PowerPoint-Präsentation zur Vorstellung Ihrer Data-Mining-Ziele oder White Papers für die Algorithmen, die Sie verwenden möchten.
- Erstellen von umfassenden und gleichzeitig einfachen Aktualisierungsberichten auf der Grundlage Ihrer Anmerkungen. Diese Berichte können in HTML generiert werden, um eine einfache Veröffentlichung im Intranet Ihrer Organisation zu ermöglichen.

Hinweis: Wenn der Projektbereich im IBM SPSS Modeler-Fenster nicht sichtbar ist, klicken Sie auf die Option **Projekt** im Menü "Ansicht".

Sie haben zwei Möglichkeiten, die einem Projekt hinzugefügten Objekte anzuzeigen: in der Ansicht **Klassen** und in der Ansicht **CRISP-DM**. Alles, was Sie zu einem Projekt hinzufügen, wird zu beiden Ansichten hinzugefügt. Sie können zwischen den Ansichten umschalten, um die Organisation zu erstellen, die am besten geeignet ist.

CRISP-DM-Ansicht

Durch Unterstützung von CRISP-DM (Cross-Industry Standard Process for Data Mining) bieten IBM SPSS Modeler-Projekte eine in der Branche bewährte, nicht gesetzlich geschützte Methode zur Organisation der einzelnen Teile Ihrer Data-Mining-Arbeit. CRISP-DM verwendet sechs Phasen zur Beschreibung des Prozesses vom Start (Zusammenstellen der Geschäftsanforderungen) bis zum Ende (Bereitstellen der Ergebnisse). Auch wenn einige Phasen bei IBM SPSS Modeler normalerweise keine Arbeit beinhalten, umfasst der Projektbereich alle sechs Phasen, sodass alle zum Projekt gehörenden Materialien an einer zentralen Stelle gespeichert und verfolgt werden können. Zum Geschäftsverständnis gehören üblicherweise das Sammeln von Anforderungen sowie Besprechungen mit Kollegen zur Bestimmung der Ziele und weniger die Arbeit mit den Daten in IBM SPSS Modeler. Mit dem Projektbereich können Sie Ihre Aufzeichnungen aus solchen Besprechungen im Ordner *Geschäftsverständnis* speichern, um sie später nachlesen oder in Berichte aufnehmen zu können.

Die CRISP-DM-Ansicht im Projektbereich ist mit einem eigenen Hilfesystem ausgestattet, das Sie durch den Data-Mining-Zyklus führt. In IBM SPSS Modeler ist der Zugriff auf diese Hilfe durch Klicken auf **CRISP-DM-Hilfe** im Hilfemenü möglich.

Hinweis: Wenn der Projektbereich nicht im Fenster sichtbar ist, klicken Sie auf die Option **Projekt** im Menü "Ansicht".

Einrichten der Standardprojektphase

Die zu einem Projekt hinzugefügten Objekte werden zu einer Standardphase von CRISP-DM hinzugefügt. Daher müssen Sie die Objekte manuell entsprechend der Data-Mining-Phase organisieren, in der Sie sie verwendet haben. Es ist sinnvoll, den Standardordner auf die Phase einzustellen, in der Sie gerade arbeiten.

So legen Sie fest, welche Phase als Standard verwendet wird:

1. Klicken Sie in der CRISP-DM-Ansicht mit der rechten Maustaste auf den Ordner für die Phase, die als Standard festgelegt werden soll.
2. Klicken Sie im Menü auf **Als Standard festlegen**.

Der Standardordner wird fett gedruckt angezeigt.

Klassenansicht

Mit der Ansicht "Klassen" im Projektbereich können Sie Ihre Arbeit in IBM SPSS Modeler nach den erstellten Objekttypen in Kategorien organisieren. Die gespeicherten Objekte können zu einer der folgenden Kategorien hinzugefügt werden:

- Streams
- Knoten
- Modelle
- Tabellen, Diagramme, Berichte
- Andere (Nicht-IBM SPSS Modeler-Dateien, wie beispielsweise Präsentationen oder White Papers, die für Ihre Data-Mining-Arbeit relevant sind)

Wenn Objekte zur Klassenansicht hinzugefügt werden, werden sie auch zum Standardphasenordner in der CRISP-DM-Ansicht hinzugefügt.

Hinweis: Wenn der Projektbereich nicht im Fenster sichtbar ist, klicken Sie auf die Option **Projekt** im Menü "Ansicht".

Erstellen von Projekten

Bei einem Projekt handelt es sich im Wesentlichen um eine Datei mit Verweisen auf alle Dateien, die Sie dem Projekt zuordnen. Projektelemente werden also sowohl einzeln als auch als Verweis in der Projektdatei (.cpj) gespeichert. Aufgrund dieser referenziellen Struktur sollten Sie auf folgende Punkte achten:

- Projektelemente müssen zunächst einzeln gespeichert werden, bevor sie zu einem Projekt hinzugefügt werden. Wenn ein Element nicht gespeichert ist, werden Sie aufgefordert, es zu speichern, bevor Sie es zum aktuellen Projekt hinzufügen.
- Einzeln aktualisierte Objekte, beispielsweise Streams, werden auch in der Projektdatei aktualisiert.
- Beim manuellen Verschieben oder Löschen von Objekten (beispielsweise Streams, Knoten und Ausgabeobjekte) aus dem Dateisystem werden die Verknüpfungen in der Projektdatei ungültig.

Erstellen eines neuen Projekts

Neue Projekte können problemlos im IBM SPSS Modeler-Fenster erstellt werden. Sie können entweder mit dem Erstellen eines Projekts beginnen, wenn noch keines geöffnet ist, oder Sie können ein bestehendes Objekt schließen und von vorne anfangen.

Klicken Sie im Hauptmenü auf Folgendes:

Datei > Projekt > Neues Projekt...

Hinzufügen zu einem Projekt

Nach dem Erstellen oder Öffnen von Projekten stehen Ihnen mehrere Methoden zur Verfügung, um Objekte, wie beispielsweise Datenstreams, Knoten und Berichte, hinzuzufügen.

Hinzufügen von Objekten aus den Managern

Mithilfe der Manager in der rechten oberen Ecke des IBM SPSS Modeler-Fensters können Sie Streams oder Ausgaben hinzufügen.

1. Wählen Sie ein Objekt, beispielsweise eine Tabelle oder einen Stream aus einer der Manager-Registerkarten aus.
2. Klicken Sie mit der rechten Maustaste und klicken Sie auf die Option **Zu Projekt hinzufügen**.
Wenn das Objekt zuvor gespeichert wurde, wird es automatisch zum entsprechenden Objektordner (in der Klassenansicht) oder zum Ordner der Standardphase (in der CRISP-DM-Ansicht) hinzugefügt.
3. Alternativ können Sie durch Ziehen und Ablegen Objekte aus den Managern in den Projektbereich verschieben.

Hinweis: Sie werden möglicherweise dazu aufgefordert, das Objekt zuerst zu speichern. Beim Speichern muss **Datei zu Projekt hinzufügen** im Dialogfeld "Speichern" ausgewählt sein. Dadurch wird das Objekt nach dem Speichern automatisch zum Projekt hinzugefügt.

Hinzufügen von Knoten aus dem Erstellungsbereich

Mit dem Dialogfeld "Speichern" können Sie einzelne Knoten aus dem Streamerstellungsbereich hinzufügen.

1. Wählen Sie einen Knoten im Erstellungsbereich aus.
2. Klicken Sie mit der rechten Maustaste und klicken Sie auf die Option **Knoten speichern**. Klicken Sie alternativ im Hauptmenü auf Folgendes:
Bearbeiten > Knoten > Knoten speichern...
3. Wählen Sie im Dialogfeld "Speichern" die Option **Datei zum Projekt hinzufügen** aus.
4. Erstellen Sie einen Namen für den Knoten und klicken Sie auf **Speichern**.

Dadurch wird die Datei gespeichert und zum Projekt hinzugefügt. Knoten werden in der Klassenansicht zum Ordner *Knoten* und in der CRISP-DM-Ansicht zum Ordner der Standardphase hinzugefügt.

Hinzufügen externer Dateien

Sie können eine große Vielzahl von Nicht-IBM SPSS Modeler-Objekten zu einem Projekt hinzufügen. Dies ist sinnvoll, wenn der gesamte Data-Mining-Vorgang innerhalb von IBM SPSS Modeler verwaltet werden soll. Sie können beispielsweise Verknüpfungen mit Daten, Notizen, Präsentationen und Grafiken in einem Objekt speichern. In der CRISP-DM-Ansicht können externe Dateien zum Ordner Ihrer Wahl hinzugefügt werden. In der Klassenansicht können externe Dateien nur im Ordner *Andere* gespeichert werden.

So fügen Sie externe Dateien zu einem Projekt hinzu:

1. Ziehen Sie die Dateien vom Desktop auf das Projekt.
ODER
2. Klicken Sie mit der rechten Maustaste auf den Zielordner in der CRISP-DM- oder Klassen-Ansicht.
3. Klicken Sie im Menü auf die Option **Zu Ordner hinzufügen**.
4. Wählen Sie eine Datei im Dialogfeld aus und klicken Sie auf **Öffnen**.

Dadurch wird ein Verweis zum ausgewählten Objekt innerhalb der IBM SPSS Modeler-Projekte hergestellt.

Übertragen von Projekten an das IBM SPSS Collaboration and Deployment Services Repository

Sie können ein gesamtes Projekt, einschließlich aller Komponentendateien, in einem Schritt an das IBM SPSS Collaboration and Deployment Services Repository übertragen. Alle Objekte, die sich bereits am Zielspeicherort befinden, werden nicht verschoben. Die Funktion lässt sich auch umgekehrt nutzen: Sie können gesamte Projekte vom IBM SPSS Collaboration and Deployment Services Repository in Ihr lokales Dateisystem übertragen.

Hinweis: Für den Zugriff auf ein Repository von IBM SPSS Collaboration and Deployment Services ist eine separate Lizenz erforderlich. Weitere Informationen finden Sie unter <http://www.ibm.com/software/analytics/spss/products/deployment/cds/>.

Übertragen von Projekten

Vergewissern Sie sich, dass das zu übertragende Projekt im Projektbereich geöffnet ist.

So übertragen Sie ein Projekt:

1. Klicken Sie mit der rechten Maustaste auf den Projektordner und klicken Sie auf die Option **Projekt übertragen**.
2. Wenn Sie dazu aufgefordert werden, melden Sie sich bei IBM SPSS Collaboration and Deployment Services Repository an.
3. Geben Sie den neuen Speicherort für das Projekt an und klicken Sie auf **OK**.

Festlegen der Projekteigenschaften

Sie können den Inhalt und die Dokumentation eines Projekts mithilfe des Dialogfelds "Projekteigenschaften" anpassen. So können Sie auf die Projekteigenschaften zugreifen:

1. Klicken Sie mit der rechten Maustaste auf ein Objekt oder einen Ordner im Projektbereich und klicken Sie auf die Option **Projekteigenschaften**.
2. Klicken Sie auf die Registerkarte **Projekt**, um grundlegende Projektinformationen anzugeben.

Erstellt. Zeigt das Erstellungsdatum des Projekts an (nicht bearbeitbar).

Zusammenfassung. Sie können eine Übersicht für Ihr Data-Mining-Projekt eingeben, die im Projektbericht angezeigt werden soll.

Inhalt. Listet Typ und Anzahl der von der Projektdatei referenzierten Komponenten auf (nicht bearbeitbar).

Ungespeicherte Objekte speichern als. Gibt an, ob ungespeicherte Objekte im lokalen Dateisystem oder im Repository gespeichert werden sollen. Weitere Informationen finden Sie im Thema „Informationen zu IBM SPSS Collaboration and Deployment Services Repository“ auf Seite 125.

Objektverweise beim Laden des Projekts aktualisieren. Wählen Sie diese Option aus, um die Verweise des Projekts auf seine Komponenten zu aktualisieren. *Hinweis:* Die zu einem Projekt hinzugefügten Dateien werden nicht in der Projektdatei selbst gespeichert. Stattdessen wird ein Verweis auf die Dateien im Projekt gespeichert. Das bedeutet, dass durch Verschieben oder Löschen einer Datei das betreffende Objekt aus dem Projekt entfernt wird.

Erstellen von Anmerkungen für ein Projekt

Der Projektbereich bietet eine Reihe von Möglichkeiten, um Anmerkungen für Ihre Data-Mining-Arbeit zu erstellen. Anmerkungen auf der Projektebene werden häufig verwendet, um weiter reichende Ziele und Entscheidungen aufzuzeichnen, während die Anmerkungen auf der Ordner- oder Knotenebene zusätzliche Details bieten. Die Registerkarte "Anmerkungen" bietet genügend Raum zum Dokumentieren

von Informationen für die Projektebene, wie beispielsweise den Ausschluss von Daten mit nicht wiederherstellbaren fehlenden Daten oder vielversprechenden Hypothesen, die während der Datenexploration aufgestellt wurden.

So erstellen Sie Anmerkungen zu einem Projekt:

1. Wählen Sie den Projektordner entweder in der CRISP-DM- oder in der Klassenansicht aus.
2. Klicken Sie mit der rechten Maustaste auf den Ordner und klicken Sie auf die Option **Projekteigenschaften**.
3. Klicken Sie auf die Registerkarte **Anmerkungen**.
4. Geben Sie Schlüsselwörter und Text zur Beschreibung des Projekts ein.

Ordneigenschaften und Anmerkungen

Einzelne Projektordner (in der CRISP-DM-Ansicht und der Klassenansicht) können mit Anmerkungen versehen werden. In der CRISP-DM-Ansicht kann dies eine sehr effektive Methode sein, um die Ziele Ihrer Organisation für die einzelnen Phasen des Data-Mining zu dokumentieren. Beispielsweise können Sie mit dem Anmerkungsstool im Ordner *Geschäftsverständnis* Hinweise aufnehmen wie "Das Geschäftsziel für diese Studie besteht darin, die Abwanderung hochwertiger Kunden zu reduzieren". Dieser Text könnte dann durch Auswahl der Option **In Bericht einschließen** automatisch in den Projektbericht aufgenommen werden.

So erstellen Sie Anmerkungen zu einem Ordner:

1. Wählen Sie einen Ordner im Projektbereich aus.
2. Klicken Sie mit der rechten Maustaste auf den Ordner und klicken Sie auf die Option **Ordneigenschaften**.

In der CRISP-DM-Ansicht bestehen die Anmerkungen für die Ordner aus einer Übersicht über den Zweck der einzelnen Phasen sowie in Hinweisen zur Durchführung der relevanten Data-Mining-Aufgaben. Sie können diese Anmerkungen nach Belieben entfernen oder bearbeiten.

Name. In diesem Bereich wird der Name des ausgewählten Felds angezeigt.

QuickInfo-Text. Dient zum Erstellen benutzerdefinierter QuickInfos, die angezeigt werden, wenn der Mauszeiger über einen Projektordner geführt wird. Dies ist vor allem in der CRISP-DM-Ansicht sinnvoll, beispielsweise um einen schnellen Überblick über die Ziele der einzelnen Phasen zu bieten oder den Status einer Phase (z. B. "In Arbeit" oder "Abgeschlossen") anzugeben.

Anmerkungsfeld. Dieses Feld ist für längere Anmerkungen gedacht, die im Projektbericht zusammengestellt werden können. Die CRISP-DM-Ansicht enthält eine Beschreibung der einzelnen Data-Mining-Phasen in der Anmerkung, Sie können diese jedoch für Ihr Projekt nach Belieben anpassen.

In Bericht einschließen. Um die Anmerkung in die Berichte einzuschließen, wählen Sie **In Bericht einschließen** aus.

Objekteigenschaften

Sie können Objekteigenschaften anzeigen und auswählen, ob einzelne Objekte in den Projektbericht aufgenommen werden sollen. So können Sie auf die Objekteigenschaften zugreifen:

1. Klicken Sie mit der rechten Maustaste auf ein Objekt im Projektbereich.
2. Klicken Sie im Menü auf die Option **Objekteigenschaften**.

Name. In diesem Bereich wird der Name des gespeicherten Objekts aufgeführt.

Pfad. In diesem Bereich wird der Speicherort des gespeicherten Objekts aufgeführt.

In Bericht einschließen. Wählen Sie diese Option aus, um die Objektdetails in einem generierten Bericht einzuschließen.

Schließen von Projekten

Wenn Sie IBM SPSS Modeler beenden oder ein neues Projekt öffnen, wird die vorhandene Projektdatei (.cpj) geschlossen.

Einige Dateien, die mit dem Projekt assoziiert sind (z. B. Streams, Knoten oder Grafiken), sind möglicherweise noch geöffnet. Wenn diese Dateien geöffnet bleiben sollen, antworten Sie mit **Nein** auf die Nachricht ... **Sollen diese Dateien gespeichert und geschlossen werden?**

Wenn Sie zugehörige Dateien nach dem Schließen eines Projekts bearbeiten und speichern, werden diese aktualisierten Versionen beim nächsten Öffnen des Projekts in das Projekt aufgenommen. Um dies zu verhindern, müssen Sie die betreffende Datei aus dem Projekt entfernen oder unter einem anderen Dateinamen speichern.

Generieren von Berichten

Eine der nützlichsten Funktionen von Projekten ist die Möglichkeit, Berichte auf der Grundlage der Projektelemente und der Anmerkungen zu erstellen. Dies ist eine äußerst wichtige Komponente des effektiven Data-Mining, wie in der gesamten CRISP-DM-Methodologie. Sie können einen Bericht direkt in einem von mehreren Dateitypen erstellen oder in einem Ausgabefenster auf dem Bildschirm, um ihn sofort betrachten zu können. Von dort können Sie den Bericht drucken, speichern oder in einem Web-Browser anzeigen. Sie können gespeicherte Berichte an andere Personen in Ihrer Organisation verteilen.

Berichte werden häufig mehrmals während des Data-Mining-Prozesses aus den Berichtdateien generiert, um sie an die am Projekt beteiligten Personen weiterzugeben. Der Bericht entnimmt Informationen zu den referenzierten Objekten aus der Projektdatei sowie aus den erstellten Anmerkungen (sofern vorhanden). Sie können die Berichte entweder auf der Grundlage der Klassenansicht oder der CRISP-DM-Ansicht erstellen.

So generieren Sie einen Bericht:

1. Wählen Sie den Projektordner entweder in der CRISP-DM- oder in der Klassenansicht aus.
2. Klicken Sie mit der rechten Maustaste auf den Ordner und klicken Sie auf die Option **Projektbericht**.
3. Geben Sie die Berichtsoptionen an und klicken Sie auf **Bericht erstellen**.

Die Optionen im Berichtdialogfeld bieten mehrere Möglichkeiten zur Generierung des erforderlichen Berichtstyps:

Ausgabename. Geben Sie den Namen des Ausgabefensters an, wenn die Ausgabe des Berichts an den Bildschirm gesendet werden soll. Sie können einen benutzerdefinierten Namen angeben oder IBM SPSS Modeler das Fenster automatisch benennen lassen.

Ausgabe auf Bildschirm. Wählen Sie diese Option aus, um den Bericht zu generieren und in einem Ausgabefenster anzuzeigen. Beachten Sie, dass Sie die Möglichkeit haben, den Bericht aus dem Ausgabefenster in verschiedene Dateitypen zu exportieren.

Ausgabe in Datei. Wählen Sie diese Option aus, um den Bericht zu erstellen und als Datei mit dem in der Liste "Dateityp" angegebenen Typ zu speichern.

Dateiname. Geben Sie einen Dateinamen für den generierten Bericht an. Die Dateien werden standardmäßig im Verzeichnis `\bin` von IBM SPSS Modeler gespeichert. Mit der Auslassungsschaltfläche (...) können Sie einen anderen Speicherort angeben.

Dateityp. Die folgenden Dateitypen stehen zur Auswahl:

- **HTML-Dokument.** Der Bericht wird als einzelne HTML-Datei gespeichert. Wenn Ihr Bericht Diagramme enthält, werden diese als PNG-Dateien gespeichert und von der HTML-Datei referenziert. Wenn Sie Ihren Bericht im Internet veröffentlichen, müssen Sie die HTML-Datei und alle von ihr referenzierten Bilder hochladen.
- **Textdokument.** Der Bericht wird als einzelne Textdatei gespeichert. Wenn der Bericht Diagramme enthält werden nur die Referenzen auf Dateinamen und Pfad in den Bericht aufgenommen.
- **Microsoft Word-Dokument.** Der Bericht wird als Einzeldokument gespeichert. Alle Diagramme sind direkt in das Dokument eingebettet.
- **Microsoft Excel-Dokument.** Der Bericht wird als einzelnes Arbeitsblatt gespeichert. Alle Diagramme sind direkt in das Arbeitsblatt eingebettet.
- **Microsoft PowerPoint-Dokument.** Jede Phase wird auf einer Folie angezeigt. Alle Diagramme sind direkt in die PowerPoint-Folien eingebettet.
- **Ausgabeobjekt.** Wenn diese Datei (.cou) in IBM SPSS Modeler geöffnet wird, stimmt sie mit der oben genannten Option **Ausgabe auf Bildschirm** in der Gruppe **Berichtformat** überein.

Hinweis: Zum Exportieren einer Microsoft Office-Datei muss die entsprechende Anwendung installiert sein.

Titel. Geben Sie einen Titel für den Bericht an.

Berichtstruktur. Wählen Sie **CRISP-DM** oder **Klassen** aus. Die CRISP-DM-Ansicht bietet einen Statusbericht mit überblickartigen Synapsen sowie Einzelheiten zu den einzelnen Data-Mining-Phasen. Bei der Ansicht "Klassen" handelt es sich um eine objektbasierte Ansicht, die besser für die interne Verfolgung von Daten und Streams geeignet ist.

Autor. Der Standardbenutzername wird angezeigt. Sie können ihn jedoch ändern.

Bericht enthält. Wählen Sie eine Methode für die Aufnahme von Objekten in den Bericht aus. Wählen Sie **alle Ordner und Objekte** aus, um alle zur Projektdatei hinzugefügten Elemente einzuschließen. Sie können auch nur die Elemente einschließen lassen, bei denen in den Objekteigenschaften **In Bericht einschließen** ausgewählt ist. Alternativ können Sie, um ungemeldete Elemente zu überprüfen, festlegen, dass nur Elemente aufgenommen werden sollen, die für den Ausschluss markiert sind (bei denen **In Bericht einschließen** inaktiviert ist).

Auswählen. Mit dieser Option können Sie Projektaktualisierungen bereitstellen, indem Sie nur Elemente der Kategorie **zuletzt verwendete Elemente** in den Bericht aufnehmen. Alternativ können Sie ältere und vielleicht ungelöste Probleme nachverfolgen, indem Sie die Parameter für **alte Elemente** festlegen. Wählen Sie **alle Elemente** aus, um die Zeit als Berichtsparameter zu ignorieren.

Sortieren nach. Sie können eine Kombination der folgenden Objektmerkmale auswählen, um sie innerhalb eines Ordners zu sortieren:

- **Typ.** Objekte werden nach Typ gruppiert.
- **Name.** Objekte werden alphabetisch angeordnet.
- **Einfügedatum.** Objekte werden nach dem Datum ihrer Hinzufügung zum Projekt sortiert.

Speichern und Exportieren von generierten Berichten

Ein auf dem Bildschirm generierter Bericht wird in einem neuen Ausgabefenster angezeigt. Alle im Bericht enthaltenen Diagramme werden als Inlinebilder angezeigt.

Berichtsterminologie

Im Bericht wird die Gesamtzahl der Knoten in den einzelnen Streams aufgeführt. Die Zahlen werden unter folgenden Überschriften angegeben, wobei IBM SPSS Modeler-Terminologie, nicht CRISP-DM-Terminologie, verwendet wird:

- **Datenleser.** Quellenknoten.
- **Datenschreiber.** Exportknoten.
- **Modellersteller.** Erstellungs- bzw. Modellierungsknoten.
- **Modellanwender.** Generierte Modelle, auch als "Nuggets" bezeichnet.
- **Ausgabeersteller.** Diagramm- oder Ausgabeknoten.
- **Andere.** Alle anderen Knoten, die zum Projekt gehören. Dazu gehören beispielsweise die Knoten, die auf der Registerkarte "Feldoperationen" oder "Datensatzoperationen" auf der Knotenpalette verfügbar sind.

So speichern Sie einen Bericht:

1. Klicken Sie im Menü "Datei" auf die Option **Speichern**.
2. Geben Sie einen Dateinamen an.
Der Bericht wird als Ausgabeobjekt gespeichert.

So exportieren Sie einen Bericht:

3. Klicken Sie im Menü "Datei" auf die Option **Exportieren** und wählen Sie den zu exportierenden Dateityp aus.
4. Geben Sie einen Dateinamen an.

Der Bericht wird im gewählten Format gespeichert.

Ein Export ist in folgende Dateitypen möglich:

- HTML
- Text
- Microsoft Word
- Microsoft Excel
- Microsoft PowerPoint

Hinweis: Zum Exportieren einer Microsoft Office-Datei muss die entsprechende Anwendung installiert sein.

Mit den Schaltflächen oben im Fenster haben Sie folgende Möglichkeiten:

- Drucken Sie den Bericht.
- Zeigen Sie den Bericht in einem externen Web-Browser als HTML an.

Kapitel 12. Anpassen von IBM SPSS Modeler

Anpassen von IBM SPSS Modeler-Optionen

Es gibt eine Reihe von Operationen, mit denen Sie IBM SPSS Modeler an Ihre Bedürfnisse anpassen können. Diese Anpassung besteht hauptsächlich aus der Festlegung bestimmter Benutzeroptionen, wie Arbeitsspeicherzuordnung, Standardverzeichnisse und die Verwendung von Klängen und Farben. Außerdem können Sie die Knotenpalette anpassen, die sich unten im IBM SPSS Modeler-Fenster befindet.

Festlegen von IBM SPSS Modeler-Optionen

Es gibt verschiedene Möglichkeiten, Optionen für IBM SPSS Modeler anzupassen und festzulegen:

- Legen Sie Systemoptionen, wie Speicherbelegung und Ländereinstellung fest, durch Klicken auf **Systemoptionen** im Menü **Extras > Optionen**.
- Legen Sie Benutzeroptionen, wie Schriftarten und Farben für die Anzeige fest, durch Klicken auf **Benutzeroptionen** im Menü **Extras > Optionen**.
- Legen Sie das Verzeichnis für die Anwendungen fest, die mit IBM SPSS Modeler verwendet werden können, durch Klicken auf **Hilfsanwendungen** im Menü **Extras > Optionen**.
- Legen Sie die in IBM SPSS Modeler verwendeten Standardverzeichnisse fest durch Klicken auf **Verzeichnis festlegen** oder **Serververzeichnis festlegen** im Menü "Datei".

Sie können auch Optionen festlegen, die für einige Ihrer Streams oder für alle Streams gelten. Weitere Informationen finden Sie im Thema „Festlegen von Optionen für Streams“ auf Seite 44.

Systemoptionen

Sie können die bevorzugte Sprache oder Ländereinstellung für IBM SPSS Modeler festlegen, indem Sie im Menü **Extras > Optionen** auf **Systemoptionen** klicken. Hier können Sie auch die maximale Speicherbelegung für IBM SPSS Modeler angeben. Beachten Sie, dass die in diesem Dialogfeld vorgenommenen Änderungen so lange unwirksam sind, bis Sie IBM SPSS Modeler erneut starten.

Maximale Speichergröße. Wählen Sie diese Option aus, um einen Grenzwert in Megabyte für die Speicherbelegung von IBM SPSS Modeler festzulegen. Auf einigen Plattformen beschränkt IBM SPSS Modeler die Prozessgröße, um die Belastung der Computer mit begrenzten Ressourcen oder hoher Auslastung zu reduzieren. Wenn Sie große Datenmengen bearbeiten, führt dies eventuell zu einer Fehlermeldung, dass nicht genügend Speicher vorhanden ist. Sie können die Speicherauslastung verbessern, indem Sie einen neuen Schwellenwert angeben.

Ländereinstellungen aus System verwenden. Diese Option wird standardmäßig ausgewählt und auf "Englisch" (USA) gesetzt. Inaktivieren Sie diese Option, um eine andere Sprache aus der Liste mit verfügbaren Sprachen und Ländereinstellungen auszuwählen.

Speicherverwaltung

Zusätzlich zur Einstellung **Maximale Speichergröße**, die Sie im Dialogfeld "Systemoptionen" angegeben haben, gibt es mehrere Möglichkeiten zur Optimierung der Speicherbelegung:

- Ändern Sie die Option **Maximale Elemente für nominale Felder** im Dialogfeld für die Streameigenschaften. Diese Option legt eine maximale Anzahl Mitglieder für nominale Felder fest, nach denen das Messniveau des Felds zu *ohne Typ* wird. Weitere Informationen finden Sie im Thema „Festlegen allgemeiner Optionen für Streams“ auf Seite 44.
- Zwingen Sie IBM SPSS Modeler, Speicher freizugeben, indem Sie in die rechte untere Ecke des Fensters klicken, in der der von IBM SPSS Modeler belegte Speicher und die zugeordnete Speichergröße angezeigt werden (*xx MB/xx MB*). Wenn Sie in diesen Bereich klicken, wird dieser dunkler und die Zah-

len für die Speicherzuordnung werden niedriger. Sobald der Bereich erneut die reguläre Farbe annimmt, hat IBM SPSS Modeler den gesamten möglichen Speicher freigegeben.

Festlegen von Standardverzeichnissen

Sie können das für Dateibrowser und die Ausgabe verwendete Standardverzeichnis angeben, indem Sie im Menü "Datei" **Verzeichnis festlegen** oder **Serververzeichnis festlegen** auswählen.

- **Verzeichnis festlegen.** Sie können diese Option verwenden, um das Arbeitsverzeichnis festzulegen. Das Standardarbeitsverzeichnis basiert auf dem Installationspfad Ihrer IBM SPSS Modeler-Version oder auf dem Befehlszeilenpfad, der zum Aufrufen von IBM SPSS Modeler verwendet wird. Im lokalen Modus ist das Arbeitsverzeichnis der Pfad, der für alle clientseitigen Operationen und Ausgabedateien verwendet wird (falls diese mit relativen Pfaden angegeben werden).
- **Serververzeichnis festlegen.** Die Option "Serververzeichnis festlegen" im Menü "Datei" wird aktiviert, wenn eine Remote-Serververbindung hergestellt wird. Verwenden Sie diese Option, um das Standardverzeichnis für alle Server- und Datendateien festzulegen, die für die Ein- und Ausgabe angegeben sind. Das Serverstandardverzeichnis lautet *\$CLEO/data*, wobei *\$CLEO* das Verzeichnis ist, in dem die Serverversion von IBM SPSS Modeler installiert ist. Mithilfe der Befehlszeile können Sie diesen Standardwert auch überschreiben, indem Sie das Flag `-server_directory` mit dem Befehlszeilenargument `modelerclient` verwenden.

Festlegen von Benutzeroptionen

Sie können allgemeine Optionen für IBM SPSS Modeler festlegen, indem Sie im Menü **Extras > Optionen** den Eintrag **Benutzeroptionen** auswählen. Diese Optionen gelten für alle in IBM SPSS Modeler verwendeten Streams.

Die folgenden Optionen können festgelegt werden, indem Sie auf die entsprechende Registerkarte klicken:

- Benachrichtigungsoptionen, wie Modellüberschreibung und Fehlernachrichten.
- Anzeigeoptionen, wie Diagramm- und Hintergrundfarben.
- Syntaxfarbanzeigeoptionen.
- PMML-Exportoptionen, die beim Export von Modellen in Predictive Model Markup Language (PMML) verwendet werden.
- Informationen zum Benutzer bzw. Autor, wie Name, Initialen und E-Mail-Adresse. Diese Informationen können auf der Registerkarte "Anmerkungen" für Knoten und andere von Ihnen erstellte Objekte angezeigt werden.

Um streamspezifische Optionen, wie Dezimaltrennzeichen, Uhrzeit- und Datumsformate, Optimierung, Stream-Layout sowie Stream-Scripts, festzulegen, verwenden Sie das Dialogfeld "Streameigenschaften", das über die Menüs "Datei" und "Extras" aufgerufen werden kann.

Festlegen von Benachrichtigungsoptionen

Mithilfe der Registerkarte "Benachrichtigungen" im Dialogfeld "Benutzeroptionen" können Sie verschiedene Optionen bzgl. des Auftretens und des Typs von Warnungen und Bestätigungsfenstern in IBM SPSS Modeler festlegen. Sie können auch das Verhalten der Registerkarten "Ausgaben" und "Modelle" im Managerbereich festlegen, wenn eine neue Ausgabe und neue Modelle generiert werden.

Feedback-Dialogfeld für Streamausführung anzeigen. Wählen Sie diese Option aus, um ein Dialogfeld mit Fortschrittsanzeige einzublenden, wenn ein Stream seit drei Sekunden ausgeführt wird. Das Dialogfeld enthält auch Einzelheiten zu den Ausgabeobjekten, die vom Stream erstellt werden.

- **Dialogfeld nach Abschluss schließen.** Das Dialogfeld wird standardmäßig am Ende der Streamausführung geschlossen. Inaktivieren Sie dieses Kontrollkästchen, wenn das Dialogfeld sichtbar bleiben soll, nachdem der Stream beendet ist.

Warnen, wenn eine Datei durch einen Knoten überschrieben wird. Wählen Sie diese Option aus, wenn eine Fehlernachricht angezeigt werden soll, wenn eine bereits vorhandene Datei durch Knotenoperationen überschrieben wird.

Warnen, wenn eine Datenbanktabelle durch einen Knoten überschrieben wird. Wählen Sie diese Option aus, wenn eine Fehlernachricht angezeigt werden soll, wenn eine bereits vorhandene Datenbanktabelle durch Knotenoperationen überschrieben wird.

Akustische Benachrichtigungen

Verwenden Sie die Liste, um festzulegen, ob Sie bei Ereignissen und Fehlern durch ein Tonsignal benachrichtigt werden möchten. Es stehen mehrere Klänge zur Verfügung. Mit der Wiedergabeschaltfläche (Lautsprecher) geben Sie einen ausgewählten Klang wieder. Mit der Schaltfläche mit den Auslassungspunkten (...) wählen Sie einen Klang aus.

Hinweis: Die *.wav*-Dateien, mit denen in IBM SPSS Modeler Klänge erstellt werden, sind im Verzeichnis */media/sounds* Ihrer Installation gespeichert.

- **Alle Klänge stummschalten.** Wählen Sie diese Option aus, um die akustische Benachrichtigung für alle Ereignisse zu inaktivieren.

Visuelle Benachrichtigungen

Die Optionen in dieser Gruppe werden verwendet, um das Verhalten der Manager-Registerkarten "Ausgaben" und "Modelle" oben rechts beim Generieren neuer Elemente festzulegen. Wählen Sie **Neues Modell** oder **Neue Ausgabe** in der Liste aus, um das Verhalten der entsprechenden Registerkarte festzulegen.

Die folgenden Optionen sind für **Neues Modell** verfügbar:

Modell zu Stream hinzufügen. Wenn ausgewählt (Standard), wird dem Stream sowie der Registerkarte "Modelle" ein neues Modell hinzugefügt, sobald das Modell erstellt wird. Im Stream wird das Modell mit einer Verknüpfung zu dem Modellierungsknoten gezeigt, über den das Modell erstellt wurde. Wenn Sie dieses Kontrollkästchen inaktivieren, wird das Modell nur der Registerkarte "Modelle" hinzugefügt.

Bisheriges Modell ersetzen. Wenn ausgewählt (Standard), wird ein bestehendes Modell aus diesem Stream auf der Registerkarte "Modelle" und im Streamerstellungsbereich überschrieben. Wenn dieses Kontrollkästchen inaktiviert ist, wird das Modell den vorhandenen Modellen auf der Registerkarte und im Erstellungsbereich hinzugefügt. Beachten Sie, dass diese Einstellung durch die Modellersetzungseinstellung bei einer Modellverknüpfung überschrieben wird.

Die folgenden Optionen sind für **Neue Ausgabe** verfügbar:

Warnung, wenn Ausgabe größer als [n]. Wählen Sie aus, ob eine Warnung angezeigt werden soll, wenn die Anzahl der Elemente auf der Registerkarte "Ausgaben" einen zuvor angegebenen Wert übersteigt. Der Standardwert ist 20; Sie können diesen Wert jedoch bei Bedarf ändern.

Die folgenden Optionen sind in allen Fällen verfügbar:

Tabulator auswählen. Wählen Sie aus, ob Sie zur Registerkarte "Ausgaben" oder "Modelle" wechseln möchten, wenn das entsprechende Objekt während der Streamausführung generiert wird.

- Wählen Sie **Immer** aus, um zur entsprechenden Registerkarte im Managerbereich zu wechseln.
- Wählen Sie **Wenn durch aktuellen Stream generiert** aus, um nur bei Objekten, die von dem Stream generiert wurden, der derzeit im Erstellungsbereich sichtbar ist, zur entsprechenden Registerkarte zu wechseln.

- Wählen Sie **Nie** aus, um die Software so zu beschränken, dass nicht zu der entsprechenden Registerkarte gewechselt wird, um Sie über generierte Ausgaben oder Modelle zu benachrichtigen.

Tabulator blinken lassen. Legen Sie hier fest, ob die Registerkarte "Ausgaben" oder "Modelle" im Managerbereich blinken soll, wenn neue Ausgaben bzw. Modelle generiert wurden.

- Wählen Sie **Wenn nicht ausgewählt** aus, damit die entsprechende Registerkarte blinkt (falls nicht bereits ausgewählt), sobald neue Objekte im Managerbereich generiert wurden.
- Wählen Sie **Nie** aus, um die Software so zu beschränken, dass die entsprechende Registerkarte nicht blinkt, um Sie über generierte Objekte zu benachrichtigen.

Sichtbarmachen durch Scrollen in Palette (Nur neue Modelle). Legen Sie fest, ob Sie durch die Registerkarte "Modelle" im Managerbereich automatisch blättern möchten, damit das aktuellste Modell sichtbar wird.

- Wählen Sie **Immer** aus, um das Scrollen zu aktivieren.
- Wählen Sie **Wenn durch aktuellen Stream generiert** aus, um nur nach Objekten zu suchen, die von dem Stream generiert wurden, der derzeit im Erstellungsbereich sichtbar ist.
- Wählen Sie **Nie** aus, um die Software so zu beschränken, dass durch die Registerkarte "Modelle" nicht automatisch geblättert wird.

Fenster öffnen (Nur Neue Ausgabe). Legen Sie fest, ob nach der Generierung automatisch ein Ausgabe-fenster geöffnet werden soll.

- Wählen Sie **Immer** aus, um immer ein neues Ausgabefenster zu öffnen.
- Wählen Sie **Wenn durch aktuellen Stream generiert** aus, um ein neues Fenster für die Ausgabe zu öffnen, die von dem Stream generiert wurde, der derzeit im Erstellungsbereich sichtbar ist.
- Wählen Sie **Nie** aus, um die Software so zu beschränken, dass keine neuen Fenster für die generierte Ausgabe automatisch geöffnet werden.

Klicken Sie auf **Standardwerte**, um zu den Systemstandardeinstellungen für diese Registerkarte zurück-zukehren.

Festlegen von Anzeigoptionen

Auf der Registerkarte "Anzeige" im Dialogfeld "Benutzeroptionen" können Sie die Optionen für die Anzeige von Schriftarten und Farben in IBM SPSS Modeler festlegen.

Begrüßungsdialogfeld beim Start anzeigen. Wählen Sie diese Option aus, um das Begrüßungsdialogfeld beim Start anzuzeigen. Das Begrüßungsdialogfeld enthält Optionen, mit denen das Lernprogramm mit Anwendungsbeispielen gestartet, ein Demostream bzw. ein bestehender Stream oder ein bestehendes Projekt geöffnet oder ein neuer Stream erstellt werden kann.

Stream- und Superknoten-Markups anzeigen. Wählen Sie diese Option aus, damit vorhandene Markups zu Streams und Superknoten standardmäßig angezeigt werden. Markup umfasst Streamkommentare, Modellverknüpfungen und die Hervorhebung von Scoring-Verzweigungen.

Standardschriftarten und -farben (Umstellung beim Neustart). Die Optionen in diesem Kontrollfeld werden verwendet, um das Bildschirmdesign und das Farbschema von IBM SPSS Modeler sowie die Größe der angezeigten Schriftarten festzulegen. Die hier ausgewählten Optionen werden erst wirksam, wenn Sie IBM SPSS Modeler schließen und neu starten.

- **Erscheinungsbild.** Ermöglicht es Ihnen, ein Standardfarbschema und Bildschirmdesign zu wählen. Sie haben folgende Möglichkeiten:
 - SPSS Standard** (Standard), ein gemeinsames Design für alle Produkte von IBM SPSS.
 - SPSS Classic**, ein Design, mit dem Benutzer früherer IBM SPSS Modeler-Versionen vertraut sind.
 - Windows**, ein Windows-Design, das u. U. hilfreich für einen höheren Kontrast im Streamerstellungsbereich und in den Paletten ist.

- **Standardschriftgröße für Knoten.** Geben Sie eine Schriftartengröße an, die in den Knotenpaletten verwendet werden kann, sowie für Knoten, die im Streamerstellungsbereich angezeigt werden.

Hinweis: Sie können die Größe der Knotensymbole für einen Stream im Fenster "Layout" der Registerkarte "Optionen" des Dialogfelds mit den Streameigenschaften festlegen. Wählen Sie im Hauptmenü Folgendes:

Tools > Streameigenschaften > Optionen > Layout.

Benutzerdefinierte Farben. In dieser Tabelle werden die aktuell ausgewählten Farben aufgelistet, die für verschiedene Anzeigeelemente verwendet werden. Sie können für jedes in der Tabelle aufgeführte Element die aktuelle Farbe ändern, indem Sie in die entsprechende Zeile in der Spalte **Farbe** doppelklicken und eine Farbe aus der Liste auswählen. Um eine benutzerdefinierte Farbe festzulegen, blättern Sie in der Liste nach unten und klicken Sie auf den Eintrag **Farbe...**

Reihenfolge für Diagrammkategorienfarbe. In dieser Tabelle werden die aktuell ausgewählten Farben aufgelistet, die in den neu erstellten Diagrammen verwendet werden. Die Reihenfolge der Farben spiegelt die Reihenfolge wider, in der Sie im Diagramm verwendet werden. Wenn z. B. ein als eine Farbüberlagerung verwendetes nominales Feld vier eindeutige Werte enthält, werden lediglich die ersten vier der hier aufgeführten Farben verwendet. Sie können für jedes in der Tabelle aufgeführte Element die aktuelle Farbe ändern, indem Sie in die entsprechende Zeile in der Spalte **Farbe** doppelklicken und eine Farbe aus der Liste auswählen. Um eine benutzerdefinierte Farbe festzulegen, blättern Sie in der Liste nach unten und klicken Sie auf den Eintrag **Farbe...** Hier vorgenommene Änderungen wirken sich nicht auf zuvor erstellte Diagramme aus.

Klicken Sie auf **Standardwerte**, um zu den Systemstandardeinstellungen für diese Registerkarte zurückzukehren.

Festlegen von Optionen für die Syntaxanzeige

Auf der Registerkarte "Syntax" im Dialogfeld "Benutzeroptionen" können Sie die Optionen für die Schriftartattribute und Anzeigefarben in Scripts festlegen, die Sie in IBM SPSS Modeler erstellen.

Syntaxhervorhebung. In dieser Tabelle werden die aktuell ausgewählten Farben aufgelistet, die für verschiedene Syntaxelemente verwendet werden, einschließlich der Schriftart und des Fensters, in dem sie angezeigt wird. Sie können für jedes in der Tabelle aufgeführte Element die Farbe ändern, indem Sie auf die entsprechende Dropdown-Liste in der Zeile klicken und eine Farbe aus der Liste auswählen. Darüber hinaus können Sie für Schriftartelemente eine Hervorhebung durch Fettdruck oder Kursivdruck auswählen.

Vorschau. Diese Tabelle enthält ein Beispiel einer Syntaxanzeige, die die Farben und Schriftartattribute verwendet, die Sie in der Tabelle **Syntaxhervorhebung** auswählen. Diese Vorschau wird aktualisiert, sobald Sie eine Auswahl ändern.

Klicken Sie auf **Standardwerte**, um zu den Systemstandardeinstellungen für diese Registerkarte zurückzukehren.

Festlegen von PMML-Exportoptionen

Auf der Registerkarte "PMML" können Sie steuern, wie IBM SPSS Modeler Modelle in die Predictive Model Markup Language (PMML) exportiert. Weitere Informationen finden Sie im Thema „Importieren und Exportieren von Modellen als PMML“ auf Seite 146.

PMML exportieren. Hier können Sie Variationen von PMML konfigurieren, die am besten mit Ihrer Zielanwendung funktionieren.

- Wählen Sie **mit Erweiterungen** aus, um PMML-Erweiterungen für Sonderfälle zuzulassen, für die es keine Standard-PMML-Entsprechung gibt. Beachten Sie, dass dies in den meisten Fällen zu demselben Ergebnis führt wie Standard-PMML.

- Wählen Sie **Als Standard-PMML...** aus, um PMML zu exportieren, die so nah wie möglich am PMML-Standard bleibt.

PMML-Standardoptionen. Wenn oben **Als Standard-PMML...** ausgewählt ist, können Sie eine von zwei gültigen Möglichkeiten auswählen, um lineare und logistische Regressionsmodelle zu exportieren:

- Als **PMML <GeneralRegression>-Modelle**
- Als **PMML <Regression>-Modelle**

Weitere Informationen zu PMML finden Sie auf der Website der Data Mining Group unter <http://www.dmg.org/>

Festlegen von Benutzerinformationen

Benutzer-/Autoreninformationen. Hier eingegebene Informationen können auf der Registerkarte "Anmerkungen" für Knoten und andere von Ihnen erstellte Objekte angezeigt werden.

Anpassen der Knotenpalette

Streams werden mithilfe von Knoten erstellt. Die Knotenpalette unten im IBM SPSS Modeler-Fenster enthält alle Knoten, die bei der Streamerstellung verwendet werden können. Weitere Informationen finden Sie im Thema „Knotenpalette“ auf Seite 16.

Sie haben zwei Möglichkeiten für die Neuordnung der Knotenpalette:

- Anpassen des Palettenmanagers. Weitere Informationen finden Sie im Thema „Anpassen des Palettenmanagers“.
- Ändern der Art und Weise, wie Palettenregisterkarten, die Unterpaletten enthalten, auf der Knotenpalette angezeigt werden. Weitere Informationen finden Sie im Thema „Erstellen einer Unterpalette“ auf Seite 164.

Anpassen des Palettenmanagers

Der Palettenmanager kann entsprechend Ihrer Nutzung von IBM SPSS Modeler angepasst werden. Wenn Sie z. B. häufig Zeitreihendaten aus einer Datenbank analysieren, möchten Sie u. U. sicherstellen, dass der Datenbankquellenknoten, der Zeitintervallknoten, der Zeitreihenknoten und der Zeitdiagrammknoten gemeinsam auf einer speziellen Registerkarte der Palette zur Verfügung stehen. Im Palettenmanager können Sie diese Anpassungen problemlos vornehmen, indem Sie auf der Knotenpalette benutzerdefinierte Palettenregisterkarten erstellen.

Mit dem Palettenmanager können Sie verschiedene Aufgaben durchführen:

- Festlegen, welche Palettenregisterkarten auf der Knotenpalette unterhalb des Streamerstellungsbereichs angezeigt werden sollen.
- Ändern der Reihenfolge, in der die Palettenregisterkarten auf der Knotenpalette angezeigt werden.
- Erstellen und Bearbeiten eigener Palettenregisterkarten und gegebenenfalls zugehöriger Unterpaletten.
- Bearbeiten der Standardknotenauswahl auf der Registerkarte "Favoriten".

Um auf den Palettenmanager zuzugreifen, klicken Sie im Menü "Tools" auf **Paletten verwalten**.

Palettenname. Jede verfügbare Palettenregisterkarte wird aufgeführt, unabhängig davon, ob sie auf der Knotenpalette angezeigt wird oder nicht. Dies umfasst gegebenenfalls auch alle von Ihnen selbst erstellten Palettenregisterkarten. Weitere Informationen finden Sie im Thema „Erstellen einer Palettenregisterkarte“ auf Seite 163.

Anzahl der Knoten. Die Anzahl der auf den einzelnen Palettenregisterkarten angezeigten Knoten. Ein hoher Wert bedeutet, dass es praktischer sein kann, Unterpaletten zu erstellen, um die Knoten auf der Registerkarte zu unterteilen. Weitere Informationen finden Sie im Thema „Erstellen einer Unterpalette“ auf Seite 164.

Angezeigt? Wählen Sie dieses Feld aus, um die Palettenregisterkarte auf der Knotenpalette anzuzeigen. Weitere Informationen finden Sie im Thema „Anzeigen von Palettenregisterkarten auf der Knotenpalette“.

Unterpaletten. Um Unterpaletten zur Anzeige auf einer Palettenregisterkarte auszuwählen, markieren Sie den entsprechenden **Palettennamen** und klicken Sie auf diese Schaltfläche, um die Registerkarte "Unterpaletten" anzuzeigen. Weitere Informationen finden Sie im Thema „Erstellen einer Unterpalette“ auf Seite 164.

Standardeinstellungen wiederherstellen. Durch Klicken auf diese Schaltfläche können Sie alle Änderungen und Zusätze, die Sie an den Paletten und Unterpaletten vorgenommen haben, vollständig entfernen und zu den Standardeinstellungen für die Palette zurückkehren.

Erstellen einer Palettenregisterkarte

So erstellen Sie eine benutzerdefinierte Palettenregisterkarte:

1. Öffnen Sie im Menü "Extras" die Option "Palettenmanager".
2. Klicken Sie rechts neben der Spalte *Angezeigt?* auf die Schaltfläche zum Hinzufügen von Paletten; das Dialogfeld "Palette erstellen/bearbeiten" wird angezeigt.
3. Geben Sie einen eindeutigen **Palettennamen** ein.
4. Wählen Sie im Bereich **Verfügbare Knoten** den Knoten aus, der zur Palettenregisterkarte hinzugefügt werden soll.
5. Klicken Sie unter "Knoten hinzufügen" auf die Schaltfläche mit Rechtspfeil, um den hervorgehobenen Knoten zum Bereich **Ausgewählte Knoten** hinzuzufügen. Wiederholen Sie den Vorgang, bis Sie alle gewünschten Knoten hinzugefügt haben.

Nachdem alle erforderlichen Knoten hinzugefügt wurden, können Sie die Reihenfolge ändern, in der sie auf der Palettenregisterkarte angezeigt werden:

6. Mithilfe der einfachen Pfeilschaltflächen können Sie einen Knoten eine Zeile nach oben bzw. unten verschieben.
7. Mithilfe der Schaltflächen mit Pfeil und Balken können Sie einen Knoten ganz nach unten oder oben in der Liste verschieben.
8. Um einen Knoten aus einer Palette zu entfernen, markieren Sie den Knoten und klicken Sie auf die Schaltfläche "Löschen" rechts neben dem Bereich **Ausgewählte Knoten**.

Anzeigen von Palettenregisterkarten auf der Knotenpalette

Es kann in IBM SPSS Modeler Optionen geben, die Sie nie verwenden. In diesem Fall können Sie die Registerkarten mit den entsprechenden Knoten mithilfe des Palettenmanagers ausblenden.

So können Sie auswählen, welche Registerkarten auf der Knotenpalette angezeigt werden sollen:

1. Öffnen Sie im Menü "Extras" die Option "Palettenmanager".
2. Mithilfe der Kontrollkästchen in der Spalte *Angezeigt?* können Sie auswählen, ob die einzelnen Palettenregisterkarten ein- oder ausgeblendet werden sollen.

Um eine Palettenregisterkarte endgültig aus der Knotenpalette zu entfernen, markieren Sie den Knoten und klicken Sie auf die Schaltfläche "Löschen" rechts neben der Spalte *Angezeigt?* Gelöschte Palettenregisterkarten können nicht wiederhergestellt werden.

Hinweis: Mit Ausnahme der Registerkarte "Favoriten" können die im Lieferumfang von IBM SPSS Modeler enthaltenen Standardpalettenregisterkarten nicht gelöscht werden.

Ändern der Anzeigereihenfolge auf der Knotenpalette

Nachdem Sie ausgewählt haben, welche Palettenregisterkarten angezeigt werden sollen, können Sie die Reihenfolge ändern, in der sie auf der Knotenpalette angezeigt werden:

1. Mithilfe der einfachen Pfeilschaltflächen können Sie eine Palettenregisterkarte eine Zeile nach oben bzw. unten verschieben. Beim Verschieben nach oben werden sie links neben die Knotenpalette verschoben und umgekehrt.
2. Mithilfe der Schaltflächen mit Pfeil und Balken können Sie eine Palettenregisterkarte ganz nach unten oder oben in der Liste verschieben. Die Elemente oben in der Liste werden links neben der Knotenpalette angezeigt.

Anzeigen von Unterpaletten auf einer Palettenregisterkarte

So wie Sie festlegen können, welche Palettenregisterkarten auf der Knotenpalette angezeigt werden sollen, so können Sie auch festlegen, welche Unterpaletten auf der zugehörigen übergeordneten Palettenregisterkarte verfügbar sein sollen.

So wählen Sie die Unterpaletten aus, die auf einer Palettenregisterkarte angezeigt werden sollen:

1. Öffnen Sie im Menü "Extras" die Option "Palettenmanager".
2. Wählen Sie die erforderliche Palette aus.
3. Klicken Sie auf die Schaltfläche "Unterpaletten". Das Dialogfeld "Unterpaletten" wird angezeigt.
4. Mithilfe der Kontrollkästchen in der Spalte *Angezeigt?* können Sie auswählen, ob die einzelnen Unterpaletten auf der Palettenregisterkarte erscheinen sollen. Die Unterpalette **Alle** wird immer angezeigt und kann nicht gelöscht werden.
5. Um eine Unterpalette endgültig aus der Palettenregisterkarte zu entfernen, markieren Sie die Unterpalette und klicken Sie auf die Schaltfläche "Löschen" rechts neben der Spalte *Angezeigt?*

Hinweis: Die im Lieferumfang der Palettenregisterkarte "Modellierung" enthaltenen Standardunterpaletten können nicht gelöscht werden.

Ändern der Anzeigereihenfolge auf der Palettenregisterkarte

Nachdem Sie ausgewählt haben, welche Unterpaletten angezeigt werden sollen, können Sie die Reihenfolge ändern, in der sie auf der übergeordneten Palettenregisterkarte angezeigt werden:

1. Mithilfe der einfachen Pfeilschaltflächen können Sie eine Unterpalette eine Zeile nach oben bzw. unten verschieben.
2. Mithilfe der Schaltflächen mit Pfeil und Balken können Sie eine Unterpalette ganz nach unten oder oben in der Liste verschieben.

Die von Ihnen erstellten Unterpaletten werden auf der Knotenpalette angezeigt, wenn Sie die zugehörige übergeordnete Palettenregisterkarte auswählen. Weitere Informationen finden Sie im Thema „Ändern der Ansicht einer Palettenregisterkarte“ auf Seite 165.

Erstellen einer Unterpalette

Da Sie in die von Ihnen erstellten benutzerdefinierten Palettenregisterkarten jeden beliebigen bestehenden Knoten aufnehmen können, ist es möglich, dass Sie mehr Knoten auswählen, als ohne Bildlauf auf dem Bildschirm angezeigt werden können. Um die Notwendigkeit eines Bildlaufs zu vermeiden, können Sie Unterpaletten erstellen, in die Sie die Knoten aufnehmen können, die Sie für die Palettenregisterkarte ausgewählt haben. Wenn Sie beispielsweise eine Palettenregisterkarte erstellt haben, die die Knoten enthält, die Sie am häufigsten zum Erstellen von Streams verwenden, können Sie vier Unterpaletten erstellen, die die Auswahl nach Quellenknoten, Feldoperationen, Modellierung und Ausgabe aufgliedern.

Hinweis: Als Knoten für die Unterpaletten stehen nur diejenigen Knoten zur Auswahl, die zur übergeordneten Palettenregisterkarte hinzugefügt wurden.

So erstellen Sie eine Unterpalette:

1. Öffnen Sie im Menü "Extras" die Option "Palettenmanager".
2. Wählen Sie die Palette aus, zu der Sie Unterpaletten hinzufügen möchten.
3. Klicken Sie auf die Schaltfläche "Unterpaletten". Das Dialogfeld "Unterpaletten" wird angezeigt.
4. Klicken Sie rechts neben der Spalte *Angezeigt?* auf die Schaltfläche zum Hinzufügen von Unterpaletten. Das Dialogfeld "Unterpalette anzeigen/bearbeiten" wird angezeigt.
5. Geben Sie einen eindeutigen **Unterpalettennamen** ein.
6. Wählen Sie im Bereich **Verfügbare Knoten** den Knoten aus, der zur Unterpalette hinzugefügt werden soll.
7. Klicken Sie unter "Knoten hinzufügen" auf den nach rechts weisenden Pfeil, um den ausgewählten Knoten zum Bereich **Ausgewählte Knoten** hinzuzufügen.
8. Klicken Sie nach dem Hinzufügen der erforderlichen Knoten auf **OK**, um zum Dialogfeld "Unterpaletten" zurückzukehren.

Die von Ihnen erstellten Unterpaletten werden auf der Knotenpalette angezeigt, wenn Sie die zugehörige übergeordnete Palettenregisterkarte auswählen. Weitere Informationen finden Sie im Thema „Ändern der Ansicht einer Palettenregisterkarte“.

Ändern der Ansicht einer Palettenregisterkarte

Aufgrund der großen Anzahl an Knoten, die in IBM SPSS Modeler verfügbar sind, sind möglicherweise auf kleinen Bildschirmen nicht alle Knoten sichtbar, ohne dass auf der Knotenpalette ein Bildlauf nach links oder rechts durchgeführt wird. Dieses Problem tritt verstärkt auf der Palettenregisterkarte "Modellierung" auf. Um die Notwendigkeit eines Bildlaufs zu verringern, können Sie auswählen, dass nur die auf einer Unterpalette (sofern verfügbar) enthaltenen Knoten angezeigt werden sollen. Weitere Informationen finden Sie im Thema „Erstellen einer Unterpalette“ auf Seite 164.

Um die auf einer Palettenregisterkarte angezeigten Knoten zu ändern, wählen Sie die Palettenregisterkarte aus und legen dann im Menü auf der linken Seite fest, ob alle Knoten angezeigt werden sollen oder nur diejenigen, die sich auf einer bestimmten Unterpalette befinden.

CEMI-Knotenverwaltung

CEMI wird nun nicht mehr unterstützt und wurde durch CLEF ersetzt, das einen flexibleren und bedienerfreundlichen Funktionssatz bietet. Weitere Informationen finden Sie im *IBM SPSS Modeler 16 CLEF-Entwicklerhandbuch*, das im Lieferumfang dieses Release enthalten ist.

Kapitel 13. Erläuterungen zur Leistung für Streams und Knoten

Sie können Ihre Streams so entwerfen, dass die Leistungsfähigkeit maximiert wird, indem Sie die Knoten in der effizientesten Konfiguration anordnen, an den entsprechenden Stellen Knotencaches aktivieren und andere Punkte beachten, die in diesem Abschnitt ausgeführt sind.

Neben den hier erörterten Punkten können weitere und deutlichere Leistungsverbesserungen normalerweise durch die effektive Nutzung der Datenbank, insbesondere durch SQL-Optimierung, erzielt werden.

Knotenreihenfolge

Selbst wenn Sie keine SQL-Optimierung verwenden, kann sich die Reihenfolge der Knoten in einem Stream auf die Leistung auswirken. Das allgemeine Ziel besteht in der Minimierung des Verarbeitungsaufwands weiter unten im Stream. Daher sollten Knoten, die die Datenmenge verringern, sofern vorhanden, relativ an den Anfang des Streams gestellt werden. IBM SPSS Modeler Server kann während der Kompilierung automatisch einige Neuordnungsregeln anwenden, um bestimmte Knoten an eine Stelle weiter oben im Stream zu rücken, wenn nachgewiesen werden kann, dass dieser Vorgang keine Probleme mit sich bringt. (Diese Funktion ist standardmäßig aktiviert. Fragen Sie bei Ihrem Systemadministrator nach, ob diese Funktion in Ihrer Installation tatsächlich aktiviert ist.)

Durch die Verwendung der SQL-Optimierung sollen Verfügbarkeit und Effizienz maximiert werden. Da die Optimierung anhält, wenn ein Stream eine Operation enthält, die nicht in der Datenbank durchgeführt werden kann, sollten SQL-optimierte Operationen zusammen am Anfang des Streams gruppiert werden. Diese Strategie sorgt dafür, dass ein größerer Teil der Verarbeitung in der Datenbank erfolgt und somit weniger Daten in IBM SPSS Modeler eingebracht werden.

Die folgenden Operationen können in den meisten Datenbanken ausgeführt werden. Versuchen Sie, sie an den *Anfang* des Streams zu stellen:

- Merge nach Schlüssel (join)
- Auswählen
- Aggregieren
- Sortieren
- Muster
- Anhängen
- Verschiedene Vorgänge im Modus *Einschließen*, in dem alle Felder ausgewählt werden
- Füller-Operationen
- Grundlegende Ableitungsoperationen unter Verwendung von Standardarithmetik oder Zeichenfolgenrebearbeitung (je nachdem, welche Operationen von der Datenbank unterstützt werden)
- Dichotomie

Die folgenden Operationen können in den meisten Datenbanken nicht ausgeführt werden. Sie sollten im Stream *hinter* den Operationen in der vorstehenden Liste platziert werden:

- Operationen für andere Daten als Datenbankdaten, beispielsweise für Flatfiles
- Merge nach Reihenfolge
- Balance
- Bestimmte Operationen im Modus *Verwerfen* oder wenn nur ein Subset von Feldern als unterschiedlich ausgewählt wurde

- Alle Operationen, bei denen auf Daten zugegriffen werden muss, die sich in einem anderen Datensatz als dem gerade verarbeiteten befinden.
 - Ableitungen von Status- und Anzahlfeldern
 - Verlaufsknotenoperationen
 - Operationen, die Zeitreihenfunktionen (mit "@" beinhalten)
 - Die Typprüfungsmodi *Warnen* und *Abbrechen*
 - Modellerstellung, -anwendung und -analyse.
- Hinweis:* Entscheidungsbäume, Regelsets, lineare Regression und faktorgenerierte Modelle können SQL erzeugen und daher per Pushback an die Datenbank zurückübertragen werden.
- Datenausgabe an einen anderen Ort als die Datenbank, die die Daten verarbeitet

Knotencaches

Um die Streamausführung zu optimieren, können Sie einen **Cache** für Knoten einrichten, die keine Endknoten sind. Wenn Sie einen Cache für einen Knoten einrichten, wird der Cache mit den Daten gefüllt, die bei der nächsten Ausführung des Datenstreams durch den Knoten laufen. Ab diesem Zeitpunkt werden die Daten aus dem Cache (der sich in einem temporären Verzeichnis auf dem Datenträger befindet) und nicht mehr aus der Datenquelle eingelesen.

Zwischenspeichern ist am nützlichsten im Anschluss an zeitaufwendige Operationen, wie Sortieren, Zusammenführen oder Aggregation. Angenommen, es ist beispielsweise ein Quellenknoten vorhanden, der zum Lesen von Umsatzdaten aus einer Datenbank eingerichtet ist, und ein Aggregatknoten, der den Umsatz nach Standort zusammenfasst. Sie können einen Cache an dem Aggregatknoten einrichten anstatt an dem Quellenknoten, weil Sie möchten, dass der Cache die aggregierten Daten und nicht das gesamte Datenset speichert.

Hinweis: Ein Caching bei Quellenknoten, bei denen einfach eine Kopie der ursprünglichen Dateien, die in IBM SPSS Modeler eingelesen wurden, gespeichert wird, bewirkt meist keine Leistungsverbesserung.

Knoten mit aktiviertem Caching werden mit einem kleinen Dokumentsymbol in der oberen rechten Ecke angezeigt. Wenn die Daten am Knoten zwischengespeichert werden, ist das Dokumentsymbol grün.

So aktivieren Sie einen Cache:

1. Klicken Sie im Streamerstellungsbereich mit der rechten Maustaste auf den Knoten und klicken Sie im Menü auf die Option **Cache**.
2. Klicken Sie im Untermenü auf die Option **Aktivieren**.
3. Sie können den Cache inaktivieren, indem Sie mit der rechten Maustaste auf den Knoten klicken und im Untermenü auf **Inaktivieren** klicken.

Zwischenspeichern von Knoten in einer Datenbank

Bei in einer Datenbank ausgeführten Streams muss nicht das Dateisystem als Cache verwendet werden, sondern die Daten können mitten im Stream in einer temporären Tabelle in der Datenbank zwischengespeichert werden. Bei einer Kombination mit SQL-Optimierung kann dies zu einer erheblichen Leistungssteigerung führen. So kann beispielsweise die Ausgabe aus einem Stream, der mehrere Tabellen zusammenfasst, um eine Data-Mining-Ansicht zu erstellen, im Cache gespeichert und bei Bedarf wieder verwendet werden. Durch die automatische Erzeugung von SQL für alle nachgeordneten Knoten lässt sich die Leistung weiter erhöhen.

Bei Verwendung von Datenbankcaching mit Zeichenfolgen, die mehr als 255 Zeichen umfassen, müssen Sie entweder sicherstellen, dass sich ein Typknoten im Stream oberhalb des Caching-Knotens befindet und dass die Feldwerte gelesen werden, oder Sie müssen die Zeichenfolgelänge mithilfe des Parameters

default_sql_string_length in der Datei *options.cfg* festlegen. Dies stellt sicher, dass die entsprechende Spalte in der temporären Tabelle auf die korrekte Breite eingestellt wird, um die Zeichenfolgen anzuzeigen.

Um das Caching von Datenbanken nutzen zu können, müssen sowohl SQL-Optimierung als auch Datenbankcaching aktiviert sein. Die Optimierungseinstellungen auf dem Server haben Vorrang vor jenen auf dem Client. Weitere Informationen finden Sie im Thema „Festlegen von Optimierungsoptionen für Streams“ auf Seite 47.

Klicken Sie bei aktiviertem Datenbankcaching einfach mit der rechten Maustaste auf einen Knoten (keinen Endknoten), um die Daten an dieser Stelle im Cache zu speichern, und der Cache wird bei der nächsten Ausführung des Streams automatisch unmittelbar in der Datenbank erstellt. Wenn Datenbankcaching oder SQL-Optimierung nicht aktiviert ist, wird der Cache stattdessen in das Dateisystem geschrieben.

Hinweis: Die folgenden Datenbanken unterstützen temporäre Tabellen zu Caching-Zwecken: DB2, Netezza, Oracle, SQL Server und Teradata. Andere Datenbanken verwenden für das Datenbankcaching eine normale Tabelle. Der SQL-Code kann für bestimmte Datenbanken angepasst werden. Wenden Sie sich an den Technical Support, um Unterstützung anzufordern.

Leistung: Prozessknoten

Sortieren. Der Sortierknoten muss das gesamte Eingabedataset lesen, bevor es sortiert werden kann. Die Daten werden bis zu einer gewissen Obergrenze im Arbeitsspeicher gespeichert, der Rest wird auf den Datenträger geschrieben. Der Sortieralgorithmus ist ein Kombinationsalgorithmus: Daten werden bis zur angegebenen Grenze in den Arbeitsspeicher eingelesen und mithilfe eines schnellen hybriden Quick-Sort-Algorithmus sortiert. Wenn alle Daten in den Arbeitsspeicher passen, ist der Sortiervorgang abgeschlossen. Andernfalls wird der Merge-Sort-Algorithmus angewendet. Die sortierten Daten werden in eine Datei geschrieben und die nächste Datenmenge wird in den Arbeitsspeicher eingelesen, sortiert und auf den Datenträger geschrieben. Dieser Vorgang wird wiederholt, bis alle Daten gelesen wurden; anschließend werden die sortierten Datenmengen zusammengeführt (Merge). Für die Zusammenführung können mehrere Durchläufe über die auf dem Datenträger gespeicherten Daten erforderlich sein. Bei Spitzennutzung hat der Sortierknoten zwei vollständige Kopien des Datasets auf dem Datenträger gespeichert, eine sortierte und eine unsortierte.

Die Gesamtlaufzeit des Algorithmus liegt in der Größenordnung von $N \cdot \log(N)$. Dabei ist N die Anzahl der Datensätze. Die Sortierung im Arbeitsspeicher ist schneller als die Zusammenführung über den Datenträger, sodass sich die tatsächliche Laufzeit durch die Zuordnung von mehr Arbeitsspeicher zum Sortiervorgang reduzieren lässt. Der Algorithmus ordnet sich selbst einen Teil des physischen RAM zu, der von der IBM SPSS Modeler Server-Konfigurationsoption *Multiplikator für Arbeitsspeicherverwendung* gesteuert wird. Um den für die Sortierung verwendeten Arbeitsspeicher zu erhöhen, müssen Sie mehr physischen RAM bereitstellen oder diesen Wert erhöhen. Beachten Sie: Wenn der Anteil des verwendeten Arbeitsspeichers die Arbeitsmenge des Prozesses überschreitet, sodass ein Teil des Arbeitsspeichers auf den Datenträger ausgelagert wird, lässt die Leistung nach, da das Muster für den Arbeitsspeicherzugriff des arbeitsspeicherinternen Sortieralgorithmus zufällig ist und zu übermäßigem Paging führen kann. Der Sortieralgorithmus wird außer vom Sortierknoten noch von mehreren anderen Knoten verwendet; es gelten jedoch dieselben Leistungsregeln.

Klassieren. Der Klassierknoten liest das gesamte Eingabedataset ein, um die Klassengrenzen zu berechnen, bevor er Datensätze zu Klassen zuordnet. Das Dataset wird während der Berechnung der Grenzen im Cache gespeichert; anschließend wird es für die Zuordnung erneut überprüft. Bei der Klassiermethode *Feste Breite* oder *Mittelwert/Standardabweichung* wird das Dataset direkt auf dem Datenträger im Cache gespeichert. Diese Methoden weisen eine lineare Laufzeit auf und erfordern so viel Speicherplatz auf dem Datenträger, dass das gesamte Dataset gespeichert werden kann. Bei der Klassiermethode *Ränge* oder *n-tile* wird das Dataset mithilfe des zuvor beschriebenen Sortieralgorithmus sortiert und das sortierte Dataset wird als Cache verwendet. Die Sortierung führt bei diesen Methoden zu einer Laufzeit von

$M*N*\log(N)$. Dabei ist M die Anzahl der klassierten Felder und N ist die Anzahl der Datensätze. Auf dem Datenträger muss das Doppelte der Datasetgröße an Speicherplatz verfügbar sein.

Das Generieren eines Ableitungsknotens auf der Grundlage von generierten Klassen verbessert die Leistung in den nachfolgenden Durchgängen. Ableitungsoperationen sind wesentlich schneller als das Klassieren.

Merge nach Schlüssel (join). Bei der Merge-Methode *Schlüssel* (entspricht einem Datenbankjoin) sortiert der Zusammenführungsknoten alle Eingabedatasets anhand der Schlüsselfelder. Dieser Teil des Verfahrens weist eine Laufzeit von $M*N*\log(N)$ auf. Dabei ist M die Anzahl der Eingaben und N die Anzahl der Datensätze in der größten Eingabe: Es erfordert genügend Datenträgerspeicher, um alle Eingabedatasets sowie eine zweite Kopie des größten Datensets zu speichern. Die Laufzeit des Merging selbst ist proportional zur Größe des Ausgabedatasets, die von der Häufigkeit übereinstimmender Schlüssel abhängt. Im schlechtesten Fall, bei dem die Ausgabe das kartesische Produkt der Eingaben ist, kann sich die Laufzeit dem Wert NM nähern. Dies kommt jedoch nur selten vor. Die meisten Joins weisen deutlich weniger übereinstimmende Schlüssel auf. Wenn ein Dataset größer ist als die anderen oder wenn die eingehenden Daten bereits durch ein Schlüsselfeld sortiert wurden, können Sie mithilfe der Registerkarte "Optimierung" die Leistung dieses Knotens verbessern.

Aggregieren. Wenn die Option *Schlüssel sind zusammenhängend* nicht ausgewählt ist, liest dieser Knoten das gesamte Eingabedataset ein (speichert es jedoch nicht), bevor eine aggregierte Ausgabe erstellt wird. In extremeren Situationen, wenn die Größe der aggregierten Daten einen Grenzwert erreicht (von der Konfigurationsoption *Multiplikator für Arbeitsspeicherverwendung* von IBM SPSS Modeler Server festgelegt), wird der Rest des Datensets so sortiert und verarbeitet, als wäre die Option *Schlüssel sind zusammenhängend* ausgewählt worden. Wenn diese Option nicht ausgewählt wurde, werden keine Daten gespeichert, weil die aggregierten Ausgabedatensätze beim Lesen der Eingabedaten produziert werden.

Duplikat. Der Duplikatknoten speichert alle eindeutigen Schlüsselfelder im Eingabedataset. In Fällen, in denen alle Felder Schlüsselfelder und alle Datensätze eindeutig sind, wird das gesamte Dataset gespeichert. Standardmäßig sortiert der Duplikatknoten die Daten in den Schlüsselfeldern und wählt (oder verwirft) dann den ersten eindeutigen Datensatz aus jeder Gruppe. Für kleinere Datensets mit einer geringen Anzahl eindeutiger Schlüssel oder für diejenigen, die vorsortiert wurden, können Sie Optionen wählen, um Geschwindigkeit und Effizienz der Verarbeitung zu steigern.

Typ. In einigen Fällen speichert der Typknoten beim Lesen der Werte die Eingabedaten im Cache. Der Cache wird für die nachfolgende Verarbeitung verwendet. Für den Cache muss genügend Plattenspeicherplatz zum Speichern des gesamten Datensets verfügbar sein. Dafür wird jedoch die Verarbeitung beschleunigt.

Auswertung. Der Evaluierungsknoten muss die Eingabedaten sortieren, um n -tile zu berechnen. Der Sortiervorgang wird für jedes evaluierte Modell wiederholt, da die Scores und folglich die Datensatzreihenfolge in jedem Fall unterschiedlich sind. Die Laufzeit beträgt $M*N*\log(N)$. Dabei ist M die Anzahl der Modelle und N die Anzahl der Datensätze.

Leistung: Modellierungsknoten

Netz und Kohonen. Trainingsalgorithmen für neuronale Netze (einschließlich des Kohonen-Algorithmus) führen zahlreiche Durchläufe über die Trainingsdaten durch. Die Daten werden bis zu einer gewissen Obergrenze im Arbeitsspeicher gespeichert, der Rest wird auf den Datenträger geschrieben. Der Zugriff auf die Trainingsdaten über den Datenträger ist aufwendig, da es sich um eine Zufallszugriffsmethode handelt, die zu übermäßiger Datenträgeraktivität führen kann. Sie können die Verwendung des Datenträgerspeichers für diese Algorithmen inaktivieren, sodass alle Daten im Arbeitsspeicher gespeichert werden müssen. Wählen Sie dazu im Dialogfeld des Knotens auf der Registerkarte "Modell" die Option **Auf Geschwindigkeit optimieren** aus. Wenn der zum Speichern der Daten erforderliche Arbeitsspeicher größer ist als die Arbeitsmenge des Serverprozesses, wird ein Teil davon auf den Datenträger ausgelagert und die Leistung wird entsprechend beeinträchtigt.

Bei Auswahl von **Speicher optimiert** wird dem Algorithmus ein Prozentsatz des physischen RAM zugeordnet. Dies geschieht gemäß dem Wert in der Konfigurationsoption *Obergrenze für Modellierungsarbeitspeicher (in Prozent)* von IBM SPSS Modeler Server. Um mehr Arbeitsspeicher für das Trainieren neuronaler Netze zu verwenden, stellen Sie entweder mehr RAM zur Verfügung oder erhöhen Sie den Wert für diese Option. Beachten Sie jedoch, dass ein zu hoher Wert zu Paging führt.

Die Laufzeit des Netzalgorithmus hängt von der gewünschten Genauigkeitsstufe ab. Sie können die Laufzeit durch Festlegung einer Stoppbedingung im Dialogfeld des Knotens steuern.

K-Means. Der K-Means-Clusteralgorithmus weist dieselben Optionen für die Steuerung der Arbeitsspeicherverwendung auf wie die Netzalgorithmen. Die Leistung für auf Datenträger gespeicherte Daten ist jedoch besser, weil der Zugriff auf die Daten sequenziell erfolgt.

Leistung: CLEM-Ausdrücke

CLEM-Sequenzfunktionen ("@-Funktionen"), die einen Rückblick auf den Datenstream ermöglichen, müssen so viele Daten speichern, wie für den längsten Rückblick erforderlich sind. Bei Operationen mit unbegrenztem Rückblicksgrad müssen alle Werte im Feld gespeichert werden. Eine unbegrenzte Operation ist eine Operation, bei der es sich beim Offset-Wert nicht um ein ganzzahliges Literal handelt; beispielsweise @OFFSET(Verkäufe, Monat). Der Offset-Wert ist der Feldname *Monat*, dessen Wert bis zur Ausführung unbekannt ist. Der Server muss alle Werte des Felds *Verkäufe* speichern, um genaue Ergebnisse zu gewährleisten. Wenn eine Obergrenze bekannt ist, sollten Sie sie als zusätzliches Argument angeben; Beispiel: @OFFSET(Verkäufe, Monat, 12). Diese Operation weist den Server an, nicht mehr als die 12 aktuellsten Werte von *Verkäufe* zu speichern. Sequenzfunktionen, ob mit Grenzen oder ohne, behindern fast immer die SQL-Generierung.

Kapitel 14. Eingabehilfen in IBM SPSS Modeler

Übersicht über die Eingabehilfen in IBM SPSS Modeler

IBM SPSS Modeler bietet Eingabehilfen für alle Benutzer sowie besondere Unterstützung für Benutzer mit beeinträchtigtem Sehvermögen oder anderen funktionalen Einschränkungen. In diesem Abschnitt werden die Funktionen und Methoden für die Arbeit mit erweiterten Eingabehilfen, wie Sprachausgabeprogramm und Tastenkombinationen, beschrieben.

Arten der Unterstützung von Eingabehilfen

Ob Sie nun unter einer Sehbehinderung leiden oder aus anderen Gründen auf die Tastatur angewiesen sind, es gibt eine Vielzahl alternativer Methoden zur Verwendung dieses Data-Mining-Toolkits. Sie können z. B. Streams erstellen, Optionen angeben und die Ausgabe lesen, ohne die Maus zu verwenden. Die verfügbaren Tastenkombinationen werden in den folgenden Themen beschrieben. Außerdem bietet IBM SPSS Modeler umfangreiche Unterstützung für Sprachausgabeprogramme, wie JAWS für Windows. Sie können auch das Farbschema ändern, um für zusätzlichen Kontrast zu sorgen. Die verschiedenen Möglichkeiten zur Unterstützung werden in den folgenden Themen beschrieben.

Zugriffsmöglichkeiten für Sehbehinderte

In IBM SPSS Modeler gibt es viele Eigenschaften, die Sie individuell festlegen können, um besser mit der Software arbeiten zu können.

Anzeigeoptionen

Sie können zum Anzeigen von Diagrammen verschiedene Farben auswählen. Sie können für die Software auch Ihre individuellen Windows-Einstellungen übernehmen. So wird der visuelle Kontrast unter Umständen erhöht.

1. Um die Anzeigeoptionen einzustellen, klicken Sie im Menü "Extras" auf die Option **Benutzeroptionen**.
2. Klicken Sie auf die Registerkarte **Anzeige**. Die Optionen auf dieser Registerkarte beziehen sich auf das Farbschema der Software, die Farben für Diagramme sowie die Schriftgrößen für Knoten.

Verwenden von Tönen zur Benachrichtigung

Durch das Aktivieren bzw. Inaktivieren von Tönen können Sie steuern, wie Ihnen bestimmte Operationen in der Software gemeldet werden. Zum Beispiel können Sie Töne für Ereignisse aktivieren, wie die Erstellung von Knoten und die Löschung bzw. Generierung neuer Ausgabe bzw. Modelle.

1. Um die Benachrichtigungsoptionen einzustellen, klicken Sie im Menü "Extras" auf die Option **Benutzeroptionen**.
2. Klicken Sie auf die Registerkarte **Benachrichtigungen**.

Steuern des automatischen Aufrufs neuer Fenster

Die Registerkarte "Benachrichtigungen" im Dialogfeld "Benutzeroptionen" wird auch dazu verwendet, zu steuern, ob die neu generierte Ausgabe, wie Tabellen und Diagramme, in einem separaten Fenster aufgerufen wird. Unter Umständen ist es für Sie einfacher, diese Option zu inaktivieren und nur im Bedarfsfall ein Ausgabefenster zu öffnen.

1. Um diese Optionen einzustellen, klicken Sie im Menü "Extras" auf die Option **Benutzeroptionen**.
2. Klicken Sie auf die Registerkarte **Benachrichtigungen**.
3. Wählen Sie im Dialogfeld in der Liste in der Gruppe **Visuelle Benachrichtigungen** die Option **Neue Ausgabe** aus.

4. Wählen Sie unter **Fenster öffnen** die Option **Nie** aus.

Knotengröße

Knoten können entweder in der Standard- oder in einer kleineren Größe angezeigt werden. Sie können die verschiedenen Größeneinstellungen an Ihre Anforderungen anpassen.

1. Um die Größeneinstellungen für die Knoten festzulegen, klicken Sie im Menü "Datei" auf die Option **Streameigenschaften**.
2. Klicken Sie auf die Registerkarte **Layout**.
3. Wählen Sie in der **Symbolgröße** die Option **Standard** aus.

Zugriffsmöglichkeiten für Blinde

Die Unterstützung für blinde Benutzer ist nur möglich, wenn ein Sprachausgabeprogramm verwendet wird wie JAWS für Windows. Um die Verwendung eines Sprachausgabeprogramms in IBM SPSS Modeler zu optimieren, können Sie verschiedene Einstellungen festlegen.

Anzeigeoptionen

Sprachausgabeprogramme funktionieren in der Regel besser, wenn der visuelle Bildschirmkontrast höher ist. Wenn bereits eine Windows-Einstellung mit hohem Kontrast gewählt ist, können Sie diese Windows-Einstellungen für die Software verwenden.

1. Um die Anzeigeoptionen einzustellen, klicken Sie im Menü "Extras" auf die Option **Benutzeroptionen**.
2. Klicken Sie auf die Registerkarte **Anzeige**.

Verwenden von Tönen zur Benachrichtigung

Durch das Aktivieren bzw. Inaktivieren von Tönen können Sie steuern, wie Ihnen bestimmte Operationen in der Software gemeldet werden. Zum Beispiel können Sie Töne für Ereignisse aktivieren, wie die Erstellung von Knoten und die Löschung bzw. Generierung neuer Ausgabe bzw. Modelle.

1. Um die Benachrichtigungsoptionen einzustellen, klicken Sie im Menü "Extras" auf die Option **Benutzeroptionen**.
2. Klicken Sie auf die Registerkarte **Benachrichtigungen**.

Steuern des automatischen Aufrufs neuer Fenster

Die Registerkarte "Benachrichtigungen" im Dialogfeld "Benutzeroptionen" wird auch dazu verwendet, zu steuern, ob die neu generierte Ausgabe in einem separaten Fenster aufgerufen wird. Unter Umständen ist es für Sie einfacher, diese Option zu inaktivieren und nur im Bedarfsfall ein Ausgabefenster zu öffnen.

1. Um diese Optionen einzustellen, klicken Sie im Menü "Extras" auf die Option **Benutzeroptionen**.
2. Klicken Sie auf die Registerkarte **Benachrichtigungen**.
3. Wählen Sie im Dialogfeld in der Liste in der Gruppe **Visuelle Benachrichtigungen** die Option **Neue Ausgabe** aus.
4. Wählen Sie unter **Fenster öffnen** die Option **Nie** aus.

Tastatureingabehilfen

Die Produktfunktionen können über die Tastatur aufgerufen werden. So können Sie ALT und die entsprechende Taste drücken, um Fenstermenüs zu aktivieren (z. B. Alt+D, um das Menü "Datei" aufzurufen), oder die Tabulatortaste drücken, um durch die Dialogfeldsteuerelemente zu blättern. Darüber hinaus gibt es spezielle Zugriffsmöglichkeiten, die nur für die einzelnen Hauptfenster des Produkts gelten, und hilfreiche Hinweise zum Navigieren in Dialogfeldern.

In diesem Abschnitt werden die wichtigsten Tastatureingabehilfen beschrieben, vom Öffnen eines Streams über die Verwendung von Knotendialogfeldern bis hin zur Arbeit mit Ausgaben. Für eine noch wirksamere Navigation stehen Listen mit Tastenkombinationen zur Verfügung.

Tastenkombinationen für die Navigation im Hauptfenster

Die meiste Data-Mining-Arbeit findet im Hauptfenster von IBM SPSS Modeler statt. Der Hauptbereich ist der **Streamerstellungsbereich**, der zum Erstellen und Ausführen von Datenstreams verwendet wird. Der untere Teil des Fensters enthält die **Knotenpaletten** mit allen verfügbaren Knoten. Die Paletten werden je nach Typ der Data-Mining-Operation für jede Knotengruppe auf Registerkarten angeordnet. Knoten, die zum Importieren von Daten in IBM SPSS Modeler verwendet werden, sind auf der Registerkarte "Datenquellen" gruppiert und Knoten, die zum Ableiten, Filtern und Eingeben von Feldern verwendet werden, sind auf der Registerkarte "Feldoperationen" gruppiert.

Die rechte Seite des Fensters enthält mehrere Tools zur Verwaltung von Streams, Ausgaben und Projekten. Die obere Hälfte auf der rechten Seite enthält die **Manager** mit drei Registerkarten, die zum Verwalten von Streams, Ausgaben und generierten Modellen verwendet werden. Sie können diese Objekte auswählen, indem Sie die Registerkarte und ein Objekt aus der Liste auswählen. Die untere Hälfte auf der rechten Seite enthält den **Projektbereich**, mit dem Sie Ihre Arbeit in Projekte aufteilen können. In diesem Bereich gibt es zwei Registerkarten, mit denen zwei unterschiedliche Ansichten eines Projekts möglich sind. Die Ansicht **Klassen** sortiert Projektobjekte nach Typ, während die Ansicht **CRISP-DM** Objekte nach der entsprechenden Data-Mining-Phase sortiert, z. B. Datenvorbereitung oder Modellierung. Diese verschiedenen Aspekte des IBM SPSS Modeler-Fensters werden im gesamten Hilfesystem sowie im Benutzerhandbuch erläutert.

Die folgende Tabelle enthält Tastenkombinationen, mit denen Sie im IBM SPSS Modeler-Hauptfenster navigieren und Streams erstellen können. Die Tastenkombinationen für Dialogfelder und die Ausgabe werden in den folgenden Themen aufgelistet. Beachten Sie, dass diese Tastenkombinationen nur im Hauptfenster zur Verfügung stehen.

Tabelle 32. Tastenkombinationen im Hauptfenster

Tastenkombination	Funktion
Strg+F5	Verschiebt den Fokus zu den Knotenpaletten.
Strg+F6	Verschiebt den Fokus zum Streamerstellungsbereich.
Strg+F7	Verschiebt den Fokus zum Managerbereich.
Strg+F8	Verschiebt den Fokus zum Projektbereich.

Tabelle 33. Knoten- und Streamtastenkombinationen

Tastenkombination	Funktion
Strg+N	Erstellt einen neuen, leeren Streamerstellungsbereich.
Strg+O	Zeigt das Dialogfeld "Öffnen" an, über das Sie bereits vorhandene Streams auswählen und öffnen können.
Strg+Zifferntasten	Verschiebt den Fokus in einem Fenster zur entsprechenden Registerkarte. In einem Fensterbereich oder in einem Fenster mit Registerkarten springt Strg+1 zur ersten Registerkarte von links, Strg+2 zur zweiten usw.
Strg+Abwärtspfeil	Wird in der Knotenpalette zum Verschieben des Fokus von einer Palettenregisterkarte zum ersten Knoten auf dieser Registerkarte verwendet.
Strg+Aufwärtspfeil	Wird in der Knotenpalette zum Verschieben des Fokus von einem Knoten zu dessen Palettenregisterkarte verwendet.

Tabelle 33. Knoten- und Streamtastenkombinationen (Forts.)

Tastenkombination	Funktion
Eingabe	Wenn in der Knotenpalette ein Knoten ausgewählt ist (einschließlich verfeinerter Modelle in der Palette der generierten Modelle), wird der Knoten durch Drücken dieser Taste in den Streamerstellungsbereich eingefügt. Durch Drücken der Eingabetaste wird bei bereits ausgewähltem Knoten im Erstellungsbereich das Dialogfeld für diesen Knoten geöffnet.
Strg+Eingabetaste	Wenn in der Palette ein Knoten ausgewählt ist, wird dieser Knoten durch diese Tastenkombination dem Streamerstellungsbereich hinzugefügt, ohne dass dieser ausgewählt wird. Gleichzeitig wird der Fokus auf den ersten Knoten in der Palette verschoben.
ALT+Eingabetaste	Wenn in der Palette ein Knoten ausgewählt ist, wird dieser Knoten durch diese Tastenkombination dem Streamerstellungsbereich hinzugefügt und ausgewählt. Gleichzeitig wird der Fokus auf den ersten Knoten in der Palette verschoben.
UMSCHALT + LEERTASTE	Wenn ein Knoten in der Palette einen Fokus aufweist, wird hiermit der Knoten oder Kommentar ausgewählt bzw. seine Auswahl aufgehoben. Wenn auch andere Knoten oder Kommentare ausgewählt sind, wird deren Auswahl ebenfalls aufgehoben.
Strg+Umschalt+Leertaste	Wenn ein Knoten oder Kommentar im Stream oder ein Knoten oder Kommentar in der Palette einen Fokus aufweist, wird hiermit der Knoten oder Kommentar ausgewählt bzw. seine Auswahl aufgehoben. Dies beeinflusst keine anderen ausgewählten Knoten oder Kommentare.
LinkspfeilRechtspfeil	Verschiebt den gesamten Stream horizontal, wenn der Fokus auf dem Streamerstellungsbereich ist. Wenn der Fokus auf einer Palettenregisterkarte liegt, wird zwischen den Registerkarten gewechselt. Wenn der Fokus auf einem Palettenknoten liegt, wird zwischen den Knoten in der Palette gewechselt.
AAufwärtspfeil/Abwärtspfeil	Verschiebt den gesamten Stream vertikal, wenn der Fokus auf dem Streamerstellungsbereich ist. Wenn der Fokus auf einem Palettenknoten liegt, wird zwischen den Knoten in der Palette gewechselt. Wenn der Fokus auf einer Unterpalette liegt, wird zwischen anderen Unterpaletten dieser Palettenregisterkarte gewechselt.
Alt+Linkspfeil/Rechtspfeil	Verschiebt ausgewählte Knoten und Kommentare im Streamerstellungsbereich horizontal in Richtung der Pfeiltaste.
Alt+AufwärtspfeilAbwärtspfeil	Verschiebt ausgewählte Knoten und Kommentare im Streamerstellungsbereich vertikal in Richtung der Pfeiltaste.
Strg+A	Wählt alle Knoten in einem Stream aus.
Strg+Q	Mit dieser Tastenkombination werden der Knoten, auf dem der Fokus ist, sowie alle abwärts liegenden Knoten ausgewählt. Die Auswahl aller aufwärts liegenden Knoten wird aufgehoben.
Strg+W	Mit dieser Tastenkombination wird bei einem ausgewählten Knoten, auf dem der Fokus ist, sowie allen abwärts liegenden Knoten die Auswahl aufgehoben.
Strg+Alt+D	Mit dieser Tastenkombination wird ein ausgewählter Knoten dupliziert.
Strg+Alt+L	Wenn im Stream ein Modellnugget ausgewählt ist, wird ein Dialogfeld "Einfügen" geöffnet, über das Sie ein gespeichertes Modell aus einer .nod-Datei in den Stream laden können.
Strg+Alt+R	Zeigt für einen ausgewählten Knoten die Registerkarte "Anmerkungen" an, über die Sie den Knoten umbenennen können.
Strg+Alt+U	Erstellt einen Benutzereingabe-Quellenknoten.

Tabelle 33. Knoten- und Streamtastenkombinationen (Forts.)

Tastenkombination	Funktion
Strg+Alt+C	Mit dieser Tastenkombination wird der Cache für einen Knoten aktiviert/inaktiviert.
Strg+Alt+F	Leert den Cache für einen Knoten.
Tabulator	Auf dem Streamerstellungsbereich wird zwischen allen Quellenknoten und Kommentaren im aktuellen Stream gewechselt. In einer Knotenpalette wird zwischen den Knoten in der Palette gewechselt. In einer ausgewählten Unterpalette wird zum ersten Knoten in der Unterpalette gewechselt.
UMSCHALT + TAB	Führt denselben Befehl wie die Tabulatortaste aus, jedoch in umgekehrter Reihenfolge.
Strg+Tabulatortaste	Wenn sich der Fokus im Managerbereich oder im Projektbereich befindet, wird der Fokus in den Streamerstellungsbereich verschoben. Wenn sich der Fokus in einer Knotenpalette befindet, wird der Fokus zwischen einem Knoten und seiner Palettenregisterkarte gewechselt.
Beliebige Buchstabentaste	Wenn sich der Fokus auf einem Knoten im aktuellen Stream befindet, wird der Fokus auf den nächsten Knoten gesetzt, dessen Name mit der gedrückten Taste beginnt, und springt dorthin.
F1	Startet das Hilfesystem bei einem für den Fokus relevanten Thema.
F2	Startet den Verbindungsprozess für einen im Erstellungsbereich ausgewählten Knoten. Verwenden Sie die Tabulatortaste, um zum gewünschten Knoten im Erstellungsbereich zu gelangen, und drücken Sie Umschalt+Leertaste, um die Verbindung zu beenden.
F3	Löscht alle Verbindungen für den ausgewählten Knoten im Erstellungsbereich.
F6	Verschiebt den Fokus zwischen dem Managerbereich, dem Projektbereich und der Knotenpalette.
F10	Öffnet das Menü "Datei".
Umschalt+F10	Öffnet das Popup-Menü für den Knoten oder Stream.
Löschen	Löscht einen ausgewählten Knoten aus dem Erstellungsbereich.
ESC	Schließt ein Popup-Menü oder Dialogfeld.
Strg+Alt+X	Mit dieser Tastenkombination wird der Superknoten erweitert.
Strg+Alt+Z	Vergrößert einen Superknoten.
Strg+Alt+Umschalt+Z	Verkleinert einen Superknoten.
Strg+E	Wenn sich der Fokus im Streamerstellungsbereich befindet, wird mit dieser Tastenkombination der aktuelle Stream ausgeführt.

In IBM SPSS Modeler werden auch viele der Standardtastenkombinationen verwendet, wie Strg+C zum Kopieren. Weitere Informationen finden Sie im Thema „Verwenden von Tastenkombinationen“ auf Seite 23.

Tastenkombinationen für Dialogfelder und Tabellen

Einige der Tastenkombinationen und das Sprachausgabeprogramm sind hilfreich, wenn Sie mit Dialogfeldern, Tabellen und Tabellen in Dialogfeldern arbeiten. Nachfolgend erhalten Sie eine vollständige Liste mit speziellen Tastenkombinationen und das Sprachausgabeprogramm.

Tabelle 34. Dialogfeld- und Expression Builder-Tastenkombinationen

Tastenkombination	Funktion
Alt+4	Wird zum Verwerfen aller geöffneten Dialogfelder oder Ausgabefenster verwendet. Die Ausgabe kann auf der Registerkarte "Ausgaben" im Managerbereich abgerufen werden.
STRG + ENDE	Wenn sich der Fokus auf einem beliebigen Steuerelement im Expression Builder befindet, wird die Einfügemarke mit dieser Tastenkombination an das Ende des Ausdrucks verschoben.
Strg+1	Im Expression Builder wird der Fokus mit dieser Tastenkombination zum Steuerelement für die Bearbeitung des Ausdrucks verschoben.
Strg+2	Im Expression Builder wird der Fokus mit dieser Tastenkombination zur Funktionsliste verschoben.
Strg+3	Mit dieser Tastenkombination wird der Fokus im Expression Builder zur Feldliste verschoben.

Tastenkombinationen für Tabellen

Tabellentastenkombinationen werden für Ausgabetafeln sowie für Tabellensteuerelemente in Dialogfeldern für Knoten, wie Typ, Filter und Verbinden, verwendet. In der Regel verwenden Sie die Tabulatortaste, um zwischen Tabellenzellen zu navigieren, und Strg+Tab, um die Tabellensteuerung zu verlassen. *Hinweis:* Es kann vorkommen, dass ein Sprachausgabeprogramm nicht sofort mit dem Lesen des Inhalts einer Zelle beginnt. Wenn Sie die Pfeiltasten ein- oder zweimal drücken, wird die Software zurückgesetzt und die Sprachausgabe beginnt.

Tabelle 35. Tastenkombinationen für Tabellen

Tastenkombination	Funktion
Strg+W	Mit dieser Tastenkombination wird die Kurzbeschreibung der ausgewählten Zeile in Tabellen gelesen. Zum Beispiel: "Die ausgewählten Werte aus Zeile 2 sind Geschlecht, Flag, m/w usw."
Strg+Alt+W	Mit dieser Tastenkombination wird die lange Beschreibung der ausgewählten Zeile in Tabellen gelesen. Zum Beispiel: "Die ausgewählten Werte aus Zeile 2 sind Feld = Geschlecht, Typ = Flag, Geschlecht = m/w usw."
Strg+D	Mit dieser Tastenkombination wird die Kurzbeschreibung des ausgewählten Bereichs in Tabellen gelesen. Zum Beispiel: "Auswahl ist eine Zeile mal sechs Spalten."
Strg+Alt+D	Mit dieser Tastenkombination wird die lange Beschreibung des ausgewählten Bereichs in Tabellen gelesen. Zum Beispiel: "Auswahl ist eine Zeile mal sechs Spalten. Die ausgewählten Spalten sind Feld, Typ, Fehlend. Die ausgewählte Zeile ist 1."
Strg+T	Diese Tastenkombination liefert in Tabellen eine Kurzbeschreibung der ausgewählten Spalten. Zum Beispiel: "Felder, Typ, Fehlend."
Strg+Alt+T	Diese Tastenkombination liefert in Tabellen eine lange Beschreibung der ausgewählten Spalten. Zum Beispiel: "Ausgewählte Spalten sind Felder, Typ, Fehlend."
Strg+R	Mit dieser Tastenkombination wird in Tabellen die Anzahl der in der Tabelle vorhandenen Datensätze angezeigt.
Strg+Alt+R	Mit dieser Tastenkombination wird in Tabellen die Anzahl der in der Tabelle vorhandenen Datensätze sowie Spaltennamen angezeigt.
Strg+I	Mit dieser Tastenkombination werden in Tabellen die Zelleninformation bzw. der Inhalt für die Zelle, auf der der Fokus ist, gelesen.

Tabelle 35. Tastenkombinationen für Tabellen (Forts.)

Tastenkombination	Funktion
Strg+Alt+I	Mit dieser Tastenkombination wird in Tabellen die lange Beschreibung der Zelleninformation (Spaltenname und Inhalt der Zelle) für die Zelle, auf der der Fokus ist, gelesen.
Strg+G	Diese Tastenkombination liefert für Tabellen kurze Allgemeine Auswahlinformationen.
Strg+Alt+G	Diese Tastenkombination liefert für Tabellen lange Allgemeine Auswahlinformationen.
Strg+Q	Ermöglicht in Tabellen ein schnelles Wechseln zwischen den Tabellenzellen. Mit Strg+Q werden lange Beschreibungen, wie "Geschlecht=Weiblich", gelesen, wenn Sie mit den Pfeiltasten durch die Tabelle blättern. Wenn Sie Strg+Q erneut auswählen, werden die kurzen Beschreibungen (Zelleninhalte) angezeigt.

Tastenkombinationen für Kommentare

Beim Arbeiten mit Kommentaren am Bildschirm können Sie die folgenden Tastenkombinationen verwenden.

Tabelle 36. Tastenkombinationen für Kommentare

Tastenkombination	Funktion
Alt+C	Aktiviert/inaktiviert die Funktion "Kommentare ein-/ausblenden".
Alt+M	Fügt einen neuen Kommentar ein, falls Kommentare angezeigt werden, und zeigt Kommentare an, wenn Sie gerade ausgeblendet sind.
Tabulator	Auf dem Streamerstellungsbereich wird zwischen allen Quellenknoten und Kommentaren im aktuellen Stream gewechselt.
Eingabe	Wenn ein Kommentar den Fokus aufweist, gibt dies den Beginn der Bearbeitung an.
Alt+Eingabetaste oder Strg+Tabulatortaste	Beendet die Bearbeitung und speichert vorgenommene Änderungen.
ESC	Bricht die Bearbeitung ab. Vorgenommene Änderungen gehen verloren.
Alt+Umschalt+Aufwärtspfeil	Verringert die Höhe des Textbereichs um eine Rasterzelle (oder ein Pixel), wenn die Funktion zum Ausrichten am Raster aktiviert (bzw. inaktiviert) ist.
Alt+Umschalt+Abwärtspfeil	Vergrößert die Höhe des Textbereichs um eine Rasterzelle (oder ein Pixel), wenn die Funktion zum Ausrichten am Raster aktiviert (bzw. inaktiviert) ist.
Alt+Umschalt+Linkspfeil	Verringert die Breite des Textbereichs um eine Rasterzelle (oder ein Pixel), wenn die Funktion zum Ausrichten am Raster aktiviert (bzw. inaktiviert) ist.
Alt+Umschalt+Rechtspfeil	Vergrößert die Breite des Textbereichs um eine Rasterzelle (oder ein Pixel), wenn die Funktion zum Ausrichten am Raster aktiviert (bzw. inaktiviert) ist.

Tastenkombinationen für Cluster-Viewer und Modellviewer

Für die Navigation in den Fenstern des Cluster-Viewers und des Modellviewers stehen Tastenkombinationen zur Verfügung.

Tabelle 37. Allgemeine Tastenkombinationen - Cluster-Viewer und Modellviewer

Tastenkombination	Funktion
Tabulator	Verschiebt den Fokus zum nächsten Bildschirmsteuerelement.
UMSCHALT + TAB	Verschiebt den Fokus zum vorherigen Bildschirmsteuerelement.
Abwärtspfeil	Wenn eine Dropdown-Liste den Fokus enthält, wird die Liste geöffnet bzw. erhält das nächste Element in der Liste den Fokus. Wenn ein Menü den Fokus enthält, wird der Fokus zum nächsten Element im Menü verschoben. Wenn sich der Fokus auf einem Piktogramm befindet, wird der Fokus zum nächsten Element im Set (oder zum ersten, wenn der Fokus auf dem letzten Piktogramm liegt) verschoben.
Aufwärtspfeil	Wenn eine Dropdown-Liste den Fokus enthält, wird der Fokus zum vorherigen Element in der Liste verschoben. Wenn ein Menü den Fokus enthält, wird der Fokus zum vorherigen Element im Menü verschoben. Wenn sich der Fokus auf einem Piktogramm befindet, wird der Fokus zum vorherigen Element im Set (oder zum letzten, wenn der Fokus auf dem ersten Piktogramm liegt) verschoben.
Eingabe	Schließt eine geöffnete Dropdown-Liste oder nimmt eine Auswahl in einem geöffneten Menü vor.
F6	Schaltet den Fokus zwischen dem linken und rechten Bereich des Fensters um.
Linkspfeil und Rechtspfeil	Wenn eine Registerkarte den Fokus enthält, wird der Fokus zur vorherigen bzw. nächsten Registerkarte verschoben. Wenn ein Menü den Fokus enthält, wird der Fokus zum vorherigen bzw. nächsten Menü verschoben.
Alt+Buchstabe	Wählt die Schaltfläche bzw. das Menü aus, in dessen Namen dieser Buchstabe unterstrichen ist.
ESC	Schließt ein geöffnetes Menü oder Dropdown-Listenfeld.

Nur Cluster-Viewer

Im Cluster-Viewer kann ein Cluster-nach-Funktionen-Raster angezeigt werden.

So wählen Sie die Clusteransicht anstelle der Ansicht "Modellübersicht":

1. Drücken Sie wiederholt die Tabulatortaste, bis die Schaltfläche **Ansicht** ausgewählt ist.
2. Drücken Sie zweimal die Abwärtspfeiltaste, um **Cluster** auszuwählen.
Von hier aus können Sie eine einzelne Zelle im Raster auswählen:
3. Drücken Sie wiederholt die Tabulatortaste, bis Sie das letzte Symbol in der Visualisierungssymbolleiste erreichen.



Abbildung 17. Symbol "Visualisierungsbaum anzeigen"

4. Drücken Sie noch einmal die Tabulatortaste, anschließend die Leertaste und dann eine Pfeiltaste.

Nun stehen die folgenden Tastenkombinationen zur Verfügung:

Tabelle 38. Cluster-Viewer - Tastenkombinationen

Tastenkombination	Funktion
Pfeiltaste	Verschiebt den Fokus zwischen einzelnen Zellen im Raster. Die Anzei-ge Zellenverteilung im rechten Bereich ändert sich mit dem Verschie-ben des Fokus.
Strg+, (Komma)	Wählt die ganze Spalte im Raster aus, in der eine Zelle den Fokus ent-hält, bzw. hebt die Auswahl der ganzen Spalte auf. Sie fügen der Auswahl eine Spalte hinzu, indem Sie mithilfe der Pfeiltasten zu einer Zelle in der entsprechenden Spalte wechseln und erneut Strg+, drücken.
Tabulator	Verschiebt den Fokus aus dem Raster und zum nächsten Bildschirmsteuerelement.
UMSCHALT + TAB	Verschiebt den Fokus aus dem Raster und zum vorherigen Bildschirmsteuerelement.
F2	Wechselt in den Bearbeitungsmodus (nur Beschriftungen und Beschreibungszellen).
Eingabe	Speichert Bearbeitungsänderungen und beendet den Bearbeitungsmodus (nur Beschriftungen und Beschreibungszellen).
ESC	Beendet den Bearbeitungsmodus, ohne Änderungen zu speichern (nur Beschriftungen und Beschreibungszellen).

Beispiel für Tastenkombinationen: Erstellen von Streams

Um den Streamerstellungsprozess für Benutzer zu verdeutlichen, die auf die Tastatur oder ein Sprachausgabeprogramm angewiesen sind, folgt nun ein Beispiel für eine Streamerstellung ohne Verwendung der Maus. In diesem Beispiel erstellen Sie einen Stream, der einen Knoten vom Typ "Datei (var.)", einen Knoten vom Typ "Ableitung" und einen Knoten vom Typ "Histogramm" enthält. Gehen Sie dazu wie folgt vor:

1. **Starten Sie IBM SPSS Modeler.** Beim Start von IBM SPSS Modeler befindet sich der Fokus auf der Registerkarte "Favoriten" der Knotenpalette.
2. **Strg+Abwärtspfeil.** Verschiebt den Fokus von der Registerkarte selbst zum Inhalt der Registerkarte.
3. **Rechtspfeil.** Verschiebt den Fokus zum Knoten "Variable Datei".
4. **Leertaste.** Wählt den Knoten "Variable Datei" aus.
5. **Strg+Eingabetaste.** Nimmt den neuen Variablendateiknoten in den Streamerstellungsbereich auf. Mit dieser Tastenkombination bleibt die Auswahl am Knoten "Variable Datei", sodass der nächste hinzugefügte Knoten damit verbunden wird.
6. **Tabulatortaste.** Verschiebt den Fokus zurück in die Knotenpalette.
7. **Viermal Rechtspfeil.** So gelangen Sie zum Ableitungsknoten.
8. **Leertaste.** Mit dieser Taste wird der Ableitungsknoten ausgewählt.
9. **Alt+Eingabetaste.** Fügt den Ableitungsknoten im Erstellungsbereich ein und verschiebt die Auswahl zum Ableitungsknoten. Dieser Knoten kann nun mit dem nächsten hinzugefügten Knoten verbunden werden.
10. **Tabulatortaste.** Verschiebt den Fokus zurück in die Knotenpalette.
11. **Fünfmal Rechtspfeil.** Verschiebt den Fokus zum Knoten "Histogramm" in der Palette.
12. **Leertaste.** Wählt den Knoten "Histogramm" aus.
13. **Eingabetaste.** Fügt den Knoten dem Stream hinzu und verschiebt den Fokus in den Streamerstellungsbereich.

Fahren Sie mit dem nächsten Beispiel fort oder speichern Sie den Stream, falls Sie das nächste Beispiel zu einem späteren Zeitpunkt ausprobieren möchten.

Beispiel für Tastenkombinationen: Bearbeiten von Knoten

In diesem Beispiel verwenden Sie den im vorherigen Beispiel erstellten Stream. Der Stream besteht aus einem Knoten vom Typ "Datei (var.)", einem Ableitungsknoten und einem Knoten vom Typ "Histogramm". Die Anweisungen beginnen mit dem Fokus auf dem dritten Knoten im Stream, dem Knoten vom Typ "Histogramm".

1. **Strg+zweimal Linkspfeil.** Verschiebt den Fokus zurück zum Knoten "Datei (.var)".
2. **Eingabetaste.** Öffnet das Dialogfeld "Variable Datei". Wechseln Sie mithilfe der Tabulatortaste in das Feld "Datei" und geben Sie einen Dateipfad und einen Dateinamen ein, um die entsprechende Datei auszuwählen. Wechseln Sie dann mithilfe der Tastenkombination Strg+Tab in den unteren Bereich des Dialogfelds, drücken Sie die Tabulatortaste, bis die Schaltfläche "OK" ausgewählt ist, und drücken Sie die Eingabetaste, um das Dialogfeld zu schließen.
3. **Strg+Rechtspfeil.** Verschiebt den Fokus zum zweiten Knoten, einem Ableitungsknoten.
4. **Eingabetaste.** Öffnet das Dialogfeld "Ableitungsknoten". Blättern Sie mit der Tabulatortaste durch dieses Dialogfeld, um die gewünschten Felder auszuwählen und die Ableitungsbedingungen festzulegen. Wechseln Sie dann mithilfe der Tastenkombination Strg+Tab zur Schaltfläche "OK" und drücken Sie die Eingabetaste, um das Dialogfeld zu schließen.
5. **Strg+Rechtspfeil.** Verschiebt den Fokus zum dritten Knoten, einem Knoten vom Typ "Histogramm".
6. **Eingabetaste.** Öffnet das Dialogfeld "Histogrammknoten". Blättern Sie mit der Tabulatortaste durch dieses Dialogfeld, um die gewünschten Felder auszuwählen und Diagrammoptionen festzulegen. Für Dropdown-Listen drücken Sie die Abwärtspfeiltaste, um die Liste zu öffnen und ein Element zu markieren, und drücken Sie dann die Eingabetaste, um das Listenelement auszuwählen. Wechseln Sie dann mithilfe der Tabulatortaste zur Schaltfläche "OK" und drücken Sie die Eingabetaste, um das Dialogfeld zu schließen.

Zu diesem Zeitpunkt können Sie weitere Knoten hinzufügen oder den aktuellen Stream ausführen. Beachten Sie beim Erstellen von Streams folgende Hinweise:

- Verwenden Sie beim manuellen Verbinden von Knoten die Taste F2, um den Startpunkt einer Verbindung zu erstellen. Navigieren Sie dann mithilfe der Tabulatortaste zum Endpunkt und drücken Sie Umschalt+Leertaste, um die Verbindung fertigzustellen.
- Verwenden Sie die Taste F3, um alle Verbindungen für einen ausgewählten Knoten im Erstellungsbereich zu zerstören.
- Nachdem Sie einen Stream erstellt haben, drücken Sie Strg+E, um den aktuellen Stream auszuführen.

Eine vollständige Liste mit Tastenkombinationen ist vorhanden. Weitere Informationen finden Sie im Thema „Tastenkombinationen für die Navigation im Hauptfenster“ auf Seite 175.

Verwenden eines Sprachausgabeprogramms

Auf dem Markt sind einige Sprachausgabeprogramme erhältlich. IBM SPSS Modeler ist für die Unterstützung von JAWS für Windows mithilfe der Java Access Bridge konfiguriert, die zusammen mit IBM SPSS Modeler installiert wird. Wenn JAWS installiert ist, starten Sie JAWS einfach vor dem Start von IBM SPSS Modeler, um das Produkt zu verwenden.

Aufgrund der Natur der einmaligen grafischen Darstellung des Data-Mining-Prozesses von IBM SPSS Modeler werden die Diagramme und Grafiken optimal visuell verwendet. Es ist jedoch möglich, dass Sie Entscheidungen basierend auf der Ausgabe und der Modelle treffen, die unter Verwendung eines Sprachausgabeprogramms textuell angezeigt werden.

Hinweis: Auf 64-Bit-Clientmaschinen funktionieren bestimmte Technologien für behindertengerechte Bedienung nicht. Das liegt daran, dass die Java Access Bridge nicht für den 64-Bit-Betrieb konfiguriert ist.

Verwenden der Wörterbuchdatei von IBM SPSS Modeler

Eine Wörterbuchdatei von IBM SPSS Modeler (*Awt.JDF*) steht für die Verwendung mit JAWS zur Verfügung. So verwenden Sie diese Datei:

1. Navigieren Sie zum Unterverzeichnis */accessibility* Ihrer IBM SPSS Modeler-Installation und kopieren Sie die Wörterbuchdatei (*Awt.JDF*).
2. Kopieren Sie die Datei in das Verzeichnis, in dem sich Ihre JAWS-Scripts befinden.

Unter Umständen befindet sich bereits eine Datei namens *Awt.JDF* auf Ihrem System, falls andere JAVA-Anwendungen ausgeführt werden. In diesem Fall können Sie diese Wörterbuchdatei eventuell nicht verwenden, ohne die Wörterbuchdatei manuell zu bearbeiten.

Verwenden eines Sprachausgabeprogramms mit HTML-Ausgabe

Wenn Sie unter Verwendung eines Sprachausgabeprogramms eine Ausgabe anzeigen, die in IBM SPSS Modeler als HTML angezeigt wird, können eventuell einige Schwierigkeiten auftreten. Es sind mehrere Ausgabetypen betroffen, darunter:

- Ausgabe, die auf der Registerkarte "Erweitert" für Knoten vom Typ "Regression", "Logistische Regression" und "Faktor/PCA" angezeigt wird
- Ausgabe des Berichtsknotens

In jedem dieser Fenster oder Dialogfelder gibt es ein Tool in der Symbolleiste, das zum Aufrufen der Ausgabe in Ihrem Standardbrowser verwendet wird, der Unterstützung für ein Standardsprachausgabeprogramm bietet. Sie können das Sprachausgabeprogramm dann verwenden, um die Ausgabeinformationen zu übermitteln.

Eingabehilfen im Fenster "Interaktiver Baum"

Die Standardanzeige eines Entscheidungsbaummodells im Fenster "Interaktiver Baum" kann zu Problemen mit Sprachausgabeprogrammen führen. Um eine Version mit Eingabehilfen aufzurufen, wählen Sie in den Menüs "Interaktiver Baum":

Ansicht > Fenster mit Eingabehilfen

Die Anzeige ist ähnlich wie die Standardbaumübersicht, kann jedoch von JAWS ordnungsgemäß gelesen werden. Mit den Standardpfeiltasten können Sie sich nach oben, nach unten, nach rechts oder nach links bewegen. Bei der Navigation im Fenster mit Eingabehilfen ändert sich der Fokus im Fenster "Interaktiver Baum" entsprechend. Verwenden Sie die Leertaste, um die Auswahl zu ändern, oder Strg+Leertaste, um die aktuelle Auswahl zu erweitern.

Tipps zur Verwendung

Es gibt mehrere Tipps, wie die IBM SPSS Modeler-Umgebung für Sie benutzerfreundlicher gestaltet werden kann. Im Folgenden erhalten Sie allgemeine Tipps für die Arbeit in IBM SPSS Modeler.

- **Verlassen von erweiterten Textfeldern.** Verwenden Sie Strg+Tab, um die erweiterten Textfelder zu verlassen. *Hinweis:* Strg+Tab wird auch zum Beenden von Tabellensteuerelementen verwendet.
- **Verwenden der Tabulatortaste anstelle der Pfeiltasten.** Verwenden Sie bei der Auswahl von Optionen für ein Dialogfeld die Tabulatortasten, um zwischen den Optionsfeldern hin- und herzuwechseln. Die Pfeiltasten funktionieren in diesem Zusammenhang nicht.
- **Dropdown-Listen.** In einer Dropdown-Liste für Dialogfelder können Sie entweder die Abbruchtaste oder die Leertaste verwenden, um ein Element auszuwählen, und dann die Liste schließen. Sie können die Abbruchtaste auch verwenden, um die Dropdown-Listen zu schließen, die nicht geschlossen werden können, wenn Sie mit der Tabulatortaste zu einem anderen Steuerelement wechseln.
- **Ausführungsstatus.** Wenn Sie einen Stream für eine große Datenbank ausführen, kann JAWS mit dem Lesen des Streamstatus hinterherhinken. Drücken Sie die Steuertaste in regelmäßigen Abständen, um den Statusbericht zu aktualisieren.

- **Verwenden der Knotenpaletten.** Wenn Sie zum ersten Mal eine Registerkarte der Knotenpaletten aufrufen, liest JAWS manchmal "Gruppenfeld" anstatt den Namen des Knotens. In diesem Fall können Sie Strg+Rechtspfeil und dann Strg+Linkspfeil verwenden, um das Sprachausgabeprogramm zurückzusetzen und den Knotennamen zu hören.
- **Lesen von Menüs.** Beim erstmaligen Öffnen eines Menüs kann es gelegentlich vorkommen, dass JAWS den ersten Menüeintrag nicht lesen kann. Wenn Sie vermuten, dass dies der Fall gewesen sein könnte, verwenden Sie die Abwärtspfeiltaste und dann die Aufwärtspfeiltaste, um das erste Element im Menü zu hören.
- **Untermenüfenster.** JAWS liest die erste Ebene eines überlappenden Menüs nicht. Wenn Sie beim Sprechen eine Unterbrechung hören, während Sie durch das Menü blättern, drücken Sie die Rechtspfeiltaste, um die untergeordneten Menüelemente zu hören.

Wenn auf Ihrem System IBM SPSS Modeler Text Analytics installiert ist, können Ihnen außerdem die folgenden Tipps die Arbeit mit der interaktiven Workbench-Schnittstelle erleichtern.

- **Aufrufen von Dialogfeldern.** Möglicherweise müssen Sie die Tabulatortaste drücken, um beim Aufrufen eines Dialogfelds den Fokus auf das erste Steuerelement zu legen.
- **Verlassen von erweiterten Textfeldern.** Verwenden Sie Strg+Tab, um die erweiterten Textfelder zu verlassen und zum nächsten Steuerelement überzugehen. *Hinweis:* Strg+Tab wird auch zum Beenden von Tabellensteuerelementen verwendet.
- **Eingabe des ersten Buchstabens zum Auffinden von Elementen in der Baumliste.** Beim Suchen nach einem Element im Fensterbereich für Kategorien, im Fensterbereich für die extrahierten Ergebnisse oder im Bibliotheksbaum können Sie den ersten Buchstaben des Elements eingeben, wenn sich der Fokus im entsprechenden Fensterbereich befindet. Dadurch wird das nächste Vorkommen eines Elements ausgewählt, das mit dem eingegebenen Buchstaben beginnt.
- **Dropdown-Listen.** In einer Dropdown-Liste für Dialogfelder können Sie die Leertaste verwenden, um ein Element auszuwählen, und dann die Liste schließen.

In den folgenden Themen werden weitere zusätzliche Tipps erörtert.

Kompatibilität mit anderer Software

Beim Testen von IBM SPSS Modeler mit Sprachausgabeprogrammen, wie JAWS, hat unser Entwicklungsteam festgestellt, dass die Verwendung eines Systems Management Servers (SMS) in Ihrem Unternehmen eventuell mit der Funktion von JAWS kollidiert, Java-basierte Anwendungen, wie IBM SPSS Modeler, zu lesen. Die Inaktivierung von SMS behebt diese Situation. Weitere Informationen zu SMS finden Sie auf der Microsoft-Website.

JAWS und Java

Die verschiedenen Versionen von JAWS bieten unterschiedliche Unterstützung für Java-basierte Software-Anwendungen. IBM SPSS Modeler funktioniert zwar mit allen neueren Versionen von JAWS, bei einigen Versionen können jedoch kleinere Probleme auftreten, wenn Sie zusammen mit Java-basierten Systemen eingesetzt werden. Besuchen Sie die Website von JAWS für Windows unter <http://www.FreedomScientific.com>.

Verwenden von Diagrammen in IBM SPSS Modeler

Die visuelle Anzeige von Informationen, wie Histogramme, Evaluierungsdiagramme, Multiplots und Streudiagramme, sind mit einem Sprachausgabeprogramm schwer zu interpretieren. Beachten Sie jedoch, dass die Webdiagramme und Verteilungen mithilfe der Textübersicht im Ausgabefenster angezeigt werden können.

Kapitel 15. Unicode-Unterstützung

Unicode-Unterstützung in IBM SPSS Modeler

IBM SPSS Modeler 10.0 ist für IBM SPSS Modeler und IBM SPSS Modeler Server vollständig Unicode-kompatibel. Dadurch können Daten mit anderen Anwendungen ausgetauscht werden, die Unicode unterstützen, einschließlich mehrsprachiger Datenbanken, und das ohne jeden Informationsverlust, der durch die Konvertierung in ein oder aus einem Codierungsschema verursacht werden könnte, das für eine bestimmte Ländereinstellung spezifisch ist.

- IBM SPSS Modeler speichert Unicode-Daten intern und kann mehrsprachige Daten, die als Unicode in Datenbanken gespeichert sind, verlustfrei lesen und schreiben.
- IBM SPSS Modeler kann UTF-8-codierte Textdateien lesen und schreiben. Für den Import und Export von Textdateien wird standardmäßig die für die Ländereinstellung spezifische Codierung verwendet, doch UTF-8 wird als Alternative unterstützt. Diese Einstellung kann in den Import- bzw. Exportknoten der Datei festgelegt werden. Alternativ kann die Standardcodierung im Dialogfeld "Stromeigenschaften" geändert werden. Weitere Informationen finden Sie im Thema „Festlegen allgemeiner Optionen für Streams“ auf Seite 44.
- Statistics-, SAS- und Textdateien, die in der für die Ländereinstellung spezifischen Codierung gespeichert sind, werden beim Import in UTF-8 konvertiert und beim Export wieder in das ursprüngliche Format zurückkonvertiert. Wenn beim Schreiben in eine Datei Unicode-Zeichen vorhanden sind, die im Zeichensatz der Ländereinstellung nicht vorhanden sind, werden diese Zeichen ersetzt und eine Warnung wird angezeigt. Dies sollte nur passieren, wenn die Daten aus einer Datenquelle importiert wurden, die Unicode unterstützt (Datenbank oder UTF-8-Textdatei) und Zeichen aus einer anderen Ländereinstellung oder aus mehreren Ländereinstellungen oder Zeichensätzen enthält.
- Bilder aus IBM SPSS Modeler Solution Publisher sind UTF-8-codiert und vollständig zwischen Plattformen und Ländereinstellungen übertragbar.

Informationen zu Unicode

Das Ziel des Unicode-Standards besteht darin, ein konsistentes Verfahren zur Codierung von mehrsprachigem Text bereitzustellen, sodass dieser problemlos über Grenzen, Ländereinstellungen und Anwendungen hinweg verwendet werden kann. Der Unicode-Standard, der mittlerweile in Version 4.0.1 vorliegt, definiert einen Zeichensatz, der eine Obermenge aller derzeit häufig verwendeten Zeichensätze weltweit darstellt, und weist jedem Zeichen einen eindeutigen Namen und Codepunkt zu. Die Zeichen und ihre Codepunkte stimmen mit denen des in ISO-10646 definierten Universal Character Set ("universeller Zeichensatz", Abk. UCS) überein. Weitere Informationen finden Sie auf der Unicode-Homepage.

Bemerkungen

Diese Informationen wurden für weltweit angebotene Produkte und Dienstleistungen erarbeitet.

Möglicherweise bietet IBM die in dieser Dokumentation beschriebenen Produkte, Services oder Funktionen in anderen Ländern nicht an. Informationen über die gegenwärtig im jeweiligen Land verfügbaren Produkte und Services sind beim zuständigen IBM Ansprechpartner erhältlich. Hinweise auf IBM Lizenzprogramme oder andere IBM Produkte bedeuten nicht, dass nur Programme, Produkte oder Services von IBM verwendet werden können. Anstelle der IBM Produkte, Programme oder Services können auch andere, ihnen äquivalente Produkte, Programme oder Services verwendet werden, solange diese keine gewerblichen oder anderen Schutzrechte von IBM verletzen. Die Verantwortung für den Betrieb von Produkten, Programmen und Services anderer Anbieter liegt beim Kunden.

Für in diesem Handbuch beschriebene Erzeugnisse und Verfahren kann es IBM Patente oder Patentanmeldungen geben. Mit der Auslieferung dieses Handbuchs ist keine Lizenzierung dieser Patente verbunden. Lizenzanforderungen sind schriftlich an folgende Adresse zu richten (Anfragen an diese Adresse müssen auf Englisch formuliert werden):

IBM Director of Licensing
IBM Europe, Middle East & Africa
Tour Descartes
2, avenue Gambetta
92066 Paris La Defense
France

Trotz sorgfältiger Bearbeitung können technische Ungenauigkeiten oder Druckfehler in dieser Veröffentlichung nicht ausgeschlossen werden. Die hier enthaltenen Informationen werden in regelmäßigen Zeitabständen aktualisiert und als Neuausgabe veröffentlicht. IBM kann ohne weitere Mitteilung jederzeit Verbesserungen und/oder Änderungen an den in dieser Veröffentlichung beschriebenen Produkten und/oder Programmen vornehmen.

Verweise in diesen Informationen auf Websites anderer Anbieter werden lediglich als Service für den Kunden bereitgestellt und stellen keinerlei Billigung des Inhalts dieser Websites dar. Das über diese Websites verfügbare Material ist nicht Bestandteil des Materials für dieses IBM Produkt. Die Verwendung dieser Websites geschieht auf eigene Verantwortung.

Werden an IBM Informationen eingesandt, können diese beliebig verwendet werden, ohne dass eine Verpflichtung gegenüber dem Einsender entsteht.

Lizenznehmer des Programms, die Informationen zu diesem Produkt wünschen mit der Zielsetzung: (i) den Austausch von Informationen zwischen unabhängig voneinander erstellten Programmen und anderen Programmen (einschließlich des vorliegenden Programms) sowie (ii) die gemeinsame Nutzung der ausgetauschten Informationen zu ermöglichen, wenden sich an folgende Adresse:

IBM Software Group
ATTN: Licensing
200 W. Madison St.
Chicago, IL; 60606
USA

Die Bereitstellung dieser Informationen kann unter Umständen von bestimmten Bedingungen - in einigen Fällen auch von der Zahlung einer Gebühr - abhängig sein.

Die Lieferung des in diesem Dokument beschriebenen Lizenzprogramms sowie des zugehörigen Lizenzmaterials erfolgt auf der Basis der IBM Rahmenvereinbarung bzw. der Allgemeinen Geschäftsbedingungen von IBM, der IBM Internationalen Nutzungsbedingungen für Programmpakete oder einer äquivalenten Vereinbarung.

Alle in diesem Dokument enthaltenen Leistungsdaten stammen aus einer kontrollierten Umgebung. Die Ergebnisse, die in anderen Betriebsumgebungen erzielt werden, können daher erheblich von den hier erzielten Ergebnissen abweichen. Einige Daten stammen möglicherweise von Systemen, deren Entwicklung noch nicht abgeschlossen ist. Eine Gewährleistung, dass diese Daten auch in allgemein verfügbaren Systemen erzielt werden, kann nicht gegeben werden. Darüber hinaus wurden einige Daten unter Umständen durch Extrapolation berechnet. Die tatsächlichen Ergebnisse können davon abweichen. Benutzer dieses Dokuments sollten die entsprechenden Daten in ihrer spezifischen Umgebung prüfen.

Alle Informationen zu Produkten anderer Anbieter stammen von den Anbietern der aufgeführten Produkte, deren veröffentlichten Ankündigungen oder anderen allgemein verfügbaren Quellen. IBM hat diese Produkte nicht getestet und kann daher keine Aussagen zu Leistung, Kompatibilität oder anderen Merkmalen machen. Fragen zu den Leistungsmerkmalen von Produkten anderer Anbieter sind an den jeweiligen Anbieter zu richten.

Aussagen über Pläne und Absichten von IBM unterliegen Änderungen oder können zurückgenommen werden und repräsentieren nur die Ziele von IBM.

Diese Veröffentlichung enthält Beispiele für Daten und Berichte des alltäglichen Geschäftsablaufs. Sie sollen nur die Funktionen des Lizenzprogramms illustrieren und können Namen von Personen, Firmen, Marken oder Produkten enthalten. Alle diese Namen sind frei erfunden; Ähnlichkeiten mit tatsächlichen Namen und Adressen sind rein zufällig.

Marken

IBM, das IBM Logo und ibm.com sind Marken oder eingetragene Marken der IBM Corp in den USA und/oder anderen Ländern. Weitere Produkt- und Servicenamen können Marken von IBM oder anderen Unternehmen sein. Eine aktuelle Liste der IBM Marken finden Sie auf der Webseite „Copyright and trademark information“ unter www.ibm.com/legal/copytrade.shtml.

Intel, das Intel-Logo, Intel Inside, das Intel Inside-Logo, Intel Centrino, das Intel Centrino-Logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium und Pentium sind Marken oder eingetragene Marken der Intel Corporation oder der Tochtergesellschaften des Unternehmens in den USA und anderen Ländern.

Linux ist eine eingetragene Marke von Linus Torvalds in den USA, anderen Ländern oder beidem.

Microsoft, Windows, Windows NT und das Windows-Logo sind Marken der Microsoft Corporation in den USA und/oder anderen Ländern.

UNIX ist eine eingetragene Marke von The Open Group in den USA und anderen Ländern.

Java und alle auf Java basierenden Marken und Logos sind Marken oder eingetragene Marken der Oracle Corporation und/oder ihrer verbundenen Unternehmen.

Weitere Produkt- und Servicenamen können Marken von IBM oder anderen Unternehmen sein.

Index

Sonderzeichen

@BLANK (Funktion) 73, 96, 121
@DIFF (Funktion) 115
@FIELD (Funktion) 73, 122
@FIELDS_BETWEEN (Funktion) 73, 81, 122
@FIELDS_MATCHING (Funktion) 73, 81, 122
@INDEX (Funktion) 115
@LAST_NON_BLANK (Funktion) 115, 121
@MAX (Funktion) 115
@MEAN (Funktion) 115
@MIN (Funktion) 115
@MULTI_RESPONSE_SET (Funktion) 82, 122
@NULL (Funktion) 73, 96, 121
@OFFSET (Funktion) 115
 Erläuterungen zur Leistung 171
@PARTITION_FIELD (Funktion) 122
@PREDICTED (Funktion) 122
@SDEV (Funktion) 115
@SINCE (Funktion) 115
@SUM (Funktion) 115
@TARGET (Funktion) 122
@TESTING_PARTITION (Funktion) 122
@THIS (Funktion) 115
@TODAY (Funktion) 110
@TRAINING_PARTITION (Funktion) 122
@VALIDATION_PARTITION (Funktion) 122

Numerische Stichwörter

508-Kompatibilität 173

A

abs (Funktion) 101
Aggregatknoten
 Leistung 169
Aktualisieren von Modellen 142, 143
Aktualisierung
 Quellenknoten 44
allbutfirst (Funktion) 105
allbutlast (Funktion) 105
alphabefore (Funktion) 105
and (Operator) 100
Anmeldung bei IBM SPSS Modeler Server 12
Anmerkungen
 in Kommentare umwandeln 62
 Knoten 58, 63
 Ordner 153
 Projekt 152
 Streams 58, 63
Anwendungen 25

Anwendungsbeispiele 3
Anwendungsgebiete des Data-Mining 26
Anzeigeformate
 Dezimalstellen 46
 Symbol für Zifferngruppierung 46
 Währung 46
 Wissenschaftlich 46
 Zahlen 46
arccos (Funktion) 102
arcosh (Funktion) 102
arcsin (Funktion) 102
arcsinh (Funktion) 102
arctan (Funktion) 102
arctan2 (Funktion) 102
arctanh (Funktion) 102
Attribut 25
Auflisten aller Kommentare für einen Stream 62
Ausdrücke 89
Ausführung anhalten 19
Ausführungszeiten anzeigen 50
Ausgabe 17
Ausgabedateien
 speichern 65
Ausgabeknoten 37
Ausgabeobjekte
 in IBM SPSS Collaboration and Deployment Services Repository speichern 131
Ausschneiden 19
Automatisierung 75

B

Baumbasierte Analyse
 typische Anwendungen 25
Bedingungen 78
Befehlszeile
 IBM SPSS Modeler starten 11
Begrüßungsdiaologfeld 160
Beispiele
 Anwendungshandbuch 3
 Übersicht 5
Benachrichtigungen
 Optionen festlegen 158
Benennen von Knoten und Streams 63
Benutzer-ID
 IBM SPSS Modeler Server 12
Benutzerdefinierte Funktionen (UDFs) 84
Benutzerdefinierte Palettenerstellung 163
 Unterpalettenerstellung 164
Benutzeroptionen 158
Bereitstellung 126
Bereitstellung von Szenarios 139
Bereitstellungsoptionen
 Szenarios 139
Bereitstellungstyp 139

Berichte
 Ausgabe speichern 65
 Eigenschaften festlegen 154
 erzeugen 154
 zu Projekten hinzufügen 150
Beschriftungen
 anzeigen 44
 Variablen 146
 Wert 146
Beschriftungen, Objekt von IBM SPSS Collaboration and Deployment Services Repository 138
Bitweise Funktionen 103
Blättern
 Optionen festlegen 49

C

Cache
 aktivieren 157
 Cache einrichten 40
 leeren 41, 44
 speichern 41
Cachodateiknoten
 laden 65
cdf_chisq (Funktion) 102
cdf_f (Funktion) 102
cdf_normal (Funktion) 102
cdf_t (Funktion) 102
Champion-Challenger-Analyse 126, 139
Chi-Quadrat-Verteilung
 Wahrscheinlichkeitsfunktionen 102
CLEM 83
 Ausdrücke 78, 89
 Ausdrücke erstellen 83
 Ausdrücke überprüfen 85
 Beispiele 75
 Datentypen 89, 90, 91
 Einführung 24, 75
 Funktionen 84
 Sprache 89
CLEM-Ausdrücke
 Leistung 171
CLEM-Funktionen
 bitweise 103
 datetime 110
 fehlende Werte 73
 Globalwert 120
 Information 96
 Leerstellen und Nullen 121
 Liste der verfügbaren Funktionen 95
 logisch 100
 numerisch 101
 Sequenz 115
 Sonderfunktionen 122
 trigonometrisch 102
 Umwandlung 97
 Vergleich 98
 Wahrscheinlichkeit 102
 Zeichenfolge 105
 Zufall 104

- Client
 - Standardverzeichnis 158
- Codierung 44, 185
- Coordinator of Processes 13
- COP 13
- cos (Funktion) 102
- cosh (Funktion) 102
- count_equal (Funktion) 81, 98
- count_greater_than (Funktion) 81, 98
- count_less_than (Funktion) 81, 98
- count_non_nulls (Funktion) 98
- count_not_equal (Funktion) 81, 98
- count_nulls (Funktion) 73, 81, 98
- count_substring (Funktion) 105
- CRISP-DM 19, 149
 - Projektsicht 149
- CRISP-DM, Prozessmodell 27, 28

D

- Data Audit-Knoten
 - bei Datenexploration verwenden 25
 - beim Data-Mining verwenden 26
- Data-Mining 25
 - Anwendungsbeispiele 35
 - Strategie 27
- date_before (Funktion) 98
- Daten
 - Vorschau 43
- Datenbank
 - Funktionen 84
- Datensätze 25
 - fehlende Werte 72
- Datenstreams
 - erstellen 37
 - Erstellung 37
- Datentypen 77
 - in Parametern 52
- Datenzuordnung 68
- Datenzuordnungstool 66, 67
- datetime_date (Funktion) 97
- Datums-/Uhrzeitfunktionen
 - datetime_date 110
 - datetime_day 110
 - datetime_day_name 110
 - datetime_day_short_name 110
 - datetime_hour 110
 - datetime_in_seconds 110
 - datetime_minute 110
 - datetime_month 110
 - datetime_month_name 110
 - datetime_month_short_name 110
 - datetime_now datetime_second 110
 - datetime_time 110
 - datetime_timestamp 110
 - datetime_weekday 110
 - datetime_year 110
- Datums-/Uhrzeitwerte 80
- Datumsformate 46, 91, 92
- Datumsfunktionen 91, 92
 - @TODAY (Funktion) 110
 - date_before 98, 110
 - date_days_difference 110
 - date_in_days 110
 - date_in_months 110
 - date_in_weeks 110
 - date_in_years 110

- Datumsfunktionen (*Forts.*)
 - date_months_difference 110
 - date_weeks_difference 110
 - date_years_difference 110
- Datumswerte
 - bearbeiten 114
 - konvertieren 114
- Dezimalstellen
 - Anzeigeformate 46
- Dezimalzeichen
 - Zahlenanzeigeformate 44
- Diagramme
 - Ausgabe speichern 65
 - zu Projekten hinzufügen 150
- DIFF (Funktion) 115
- div (Funktion) 101
- Dokumentation 3
- Domänenname (Windows)
 - IBM SPSS Modeler Server 12
- Drucken 24
 - Streams 22, 40
- DTD 146
- Duplikatknoten
 - Leistung 169

E

- Eigenschaften
 - Berichtphasen 154
 - für Datenstreams 44
 - Projektdrner 153
- Einfügen 19
- Einführung 89
 - IBM SPSS Modeler 11, 157
- Eingabehilfen 173, 184
 - Beispiel 181, 182
 - Funktionen in IBM SPSS Modeler 173
 - Tipps in IBM SPSS Modeler 183
- Endknoten 37
- endstring (Funktion) 105
- Enterprise-Ansichtsknoten 139
- Entscheidungsbäume
 - Eingabehilfen 183
- Erforderliche Felder 68
- Ersetzen von Modellen 158
- Erstellungsbereich 15
- Evaluierungsknoten
 - Leistung 169
- exponential (Funktion) 101
- Export
 - PMML 146
 - Streambeschreibungen 57
- Expression Builder 177
 - Übersicht 83
 - verwenden 83
 - Zugriff 83

F

- f-Verteilung
 - Wahrscheinlichkeitsfunktionen 102
- Faktor 183
- Fall 25
- Farben
 - festlegen 160

- Fehlende Werte 72, 80
 - CLEM-Ausdrücke 73
 - in Datensätzen 72
- Fehlernachrichten 50
- Felder 25, 89, 91
 - in CLEM-Ausdrücken 84
 - Werte anzeigen 85
- Feldzuordnungen aufheben 66
- first_index (Funktion) 82, 98
- first_non_null (Funktion) 82, 98
- first_non_null_index (Funktion) 82, 98
- fracof (Funktion) 101
- Füllerknoten
 - fehlende Werte 73
- Funktionen 91, 92, 96, 115
 - @BLANK 73
 - @FIELD 83, 122
 - @GLOBAL_MAX 120
 - @GLOBAL_MEAN 120
 - @GLOBAL_MIN 120
 - @GLOBAL_SDEV 120
 - @GLOBAL_SUM 120
 - @PARTITION 122
 - @PREDICTED 83, 122
 - @TARGET 83, 122
 - Beispiele 75
 - benutzerdefinierte Funktionen (UDFs) 84
 - Datenbank 84
 - in CLEM-Ausdrücken 84
 - Umgang mit fehlenden Werten 73

G

- Ganze Zahlen 89
- Generierte Modelle (Palette) 17
- gleich (Operator) 98
- Globale Funktionen 120
- Globale Werte
 - in CLEM-Ausdrücken 84
- Grad
 - Maßeinheiten 46
- Grafiken
 - Ausgabe speichern 65
- Größe ändern 21
- größer als (Operator) 98

H

- hasendstring (Funktion) 105
- hasmidstring (Funktion) 105
- hasstartstring (Funktion) 105
- hassubstring (Funktion) 105
- Hauptfenster 15
- Hinweise
 - allgemeine Verwendung 68
- Hinzufügen
 - zu einem Projekt 150
- Hostname
 - IBM SPSS Modeler Server 12, 13
- Hotkeys 23
- HTML-Ausgabe
 - Sprachausgabeprogramm 183

I

- IBM InfoSphere Warehouse (ISW)
 - PMML-Export 146
- IBM SPSS Collaboration and Deployment Services 126
- IBM SPSS Collaboration and Deployment Services Enterprise-Ansicht 126, 139
- IBM SPSS Collaboration and Deployment Services Repository 125, 126
 - durchsuchen 127
 - Objekte abrufen 132
 - Objekte speichern 128
 - Objekte sperren und entsperren 135
 - Objekte und Versionen löschen 136
 - Objekteigenschaften 137
 - Ordner 135, 136
 - Projekte übertragen 152
 - Single Sign-on 127
 - Suche in 134
 - Verbindung 127
- IBM SPSS Modeler 1, 15
 - Dokumentation 3
 - Eingabehilfen 173
 - erste Schritte 11
 - Optionen 157
 - Tipps und Tastenkombinationen 68
 - über Befehlszeile ausführen 11
 - Übersicht 11, 157
- IBM SPSS Modeler Advantage 126, 139
- IBM SPSS Modeler Server 1
 - Benutzer-ID 12
 - Domänenname (Windows) 12
 - Hostname 12, 13
 - Kennwort 12
 - Portnummer 12, 13
- IBM SPSS Modeler Server-Verbindungen hinzufügen 13
- if, then, else (Funktionen) 100
- Import
 - PMML 146
- INDEX (Funktion) 115
- Informationsfunktionen 96
- integer_bitcount (Funktion) 103
- integer_leastbit (Funktion) 103
- integer_length (Funktion) 103
- Interaktiver Baum (Fenster)
 - Eingabehilfen 183
- intof (Funktion) 101
- is_date (Funktion) 96
- is_datetime (Funktion) 96
- is_integer (Funktion) 96
- is_number (Funktion) 96
- is_real (Funktion) 96
- is_string (Funktion) 96
- is_time (Funktion) 96
- is_timestamp (Funktion) 96
- isalphacode (Funktion) 105
- isendstring (Funktion) 105
- islowercode (Funktion) 105
- ismidstring (Funktion) 105
- isnumbercode (Funktion) 105
- isstartstring (Funktion) 105
- issubstring (Funktion) 105
- issubstring_count (Funktion) 105
- issubstring_lim (Funktion) 105
- isuppercode (Funktion) 105

J

- Java 184
- JAWS 173, 182, 184

K

- K-Means-Knoten
 - große Sets 44
 - Leistung 170
- Kennwort
 - IBM SPSS Modeler Server 12
- Klassen 19, 149, 150
- Klassierknoten
 - Leistung 169
- kleiner als (Operator) 98
- Knoten 11
 - aktivieren 39
 - Anmerkungen 58, 63
 - aus Palette entfernen 163
 - Ausführungszeiten 50
 - bearbeiten 40
 - benutzerdefinierte Palettenerstellung 163
 - benutzerdefinierte Unterpalettenerstellung 164
 - Datenvorschau 43
 - duplizieren 40
 - Einführung 37
 - hinzufügen 38, 40
 - in einem Stream inaktivieren 39
 - in einem Stream umgehen 39
 - in einem Stream verbinden 38
 - in IBM SPSS Collaboration and Deployment Services Repository speichern 131
 - inaktivieren 39, 40
 - Kommentare hinzufügen zu 58
 - laden 65
 - Leistung 169, 170
 - löschen 38
 - Optionen festlegen 40
 - Palette anzeigen 163
 - Palettenregisterkarte, Anpassung 165
 - Reihenfolge 167
 - speichern 64
 - sperren 43
 - Suchen 54
 - Verbindungen löschen 40
 - Vorschau von Daten 43
 - zu Projekten hinzufügen 150, 151
- Knoten aktivieren 39
- Knoten inaktivieren 39, 40
- Knotennamen 63
- Knotenpaletteauswahl 163
- Knowledge Discovery 25
- Kohonen-Knoten
 - große Sets 44
 - Leistung 170
- Komma 44
- Kommentare
 - alle in einem Stream auflisten 62
 - Tastenkombinationen 179
 - zu Knoten und Streams 58
- Konventionen 96
- Konvertierungsfunktionen 97
- Kopieren 19

L

- Laden
 - Knoten 65
 - Statusangaben 65
- Ländereinstellung
 - Optionen 157
- last_index (Funktion) 82, 98
- LAST_NON_BLANK (Funktion) 115
- last_non_null (Funktion) 82, 98
- last_non_null_index (Funktion) 82, 98
- Laufzeit, Eingabeaufforderungen während 52
- Leere Bereiche
 - aus Zeichenfolgen entfernen 79, 105
- Leerstellenbehandlung
 - CLEM-Funktionen 121
- Leerzeichen 72, 80
 - aus Zeichenfolgen entfernen 79, 105
- Leistung
 - CLEM-Ausdrücke 171
 - Modellierungsknoten 170
 - von Prozessknoten 169
- length (Funktion) 105
- Lineare Regression
 - als PMML exportieren 161
- Listen 89, 91
- locchar (Funktion) 105
- locchar_back (Funktion) 105
- log (Funktion) 101
- log10 (Funktion) 101
- Logische Funktionen 100
- Logistische Regression 183
 - als PMML exportieren 161
- lowertoupper (Funktion) 105

M

- Manager 17
- Maschinenlernen 25
- matches (Funktion) 105
- Maus
 - in IBM SPSS Modeler verwenden 22, 38
- max (Funktion) 98
- MAX (Funktion) 115
- max_index (Funktion) 82, 98
- max_n (Funktion) 81, 98
- MEAN (Funktion) 115
- mean_n (Funktion) 81, 101
- Mehrere IBM SPSS Modeler-Sitzungen 14
- Mehrfachantwortsets
 - in CLEM-Ausdrücken 82, 84
- member (Funktion) 98
- Merkmalauswahlknoten
 - fehlende Werte 72
- min (Funktion) 98
- MIN (Funktion) 115
- min_index (Funktion) 82, 98
- min_n (Funktion) 81, 98
- Minimierung 21
- Mittlere Maustaste
 - simulieren 38
 - Simulierung 22
- mod (Funktion) 101
- Modellaktualisierung 139

Modelle 57
 Aktualisierung 142, 143
 ersetzen 158
 Export 161
 in IBM SPSS Collaboration and Deployment Services Repository speichern 132
 zu Projekten hinzufügen 150
 Modellierung
 Verzweigung 57
 Modellierungsknoten 37
 Leistung 170
 Modellierungs-Palettenregisterkarte, Anpassung 165
 Modellnuggets 57

N

Nachrichten
 erzeugtes SQL anzeigen 48
 Navigieren
 Tastenkombinationen 174
 negate (Funktion) 101
 Netzknoten
 Leistung 170
 Neue Funktionen 7
 Neuronales Netz, Knoten
 große Sets 44
 Normalverteilung
 Wahrscheinlichkeitsfunktionen 102
 not (Operator) 100
 Nuggets 57
 definiert 17
 Nullen 80
 Numerische Funktionen 101

O

Objekte
 Eigenschaften 153
 Objekte aus dem IBM SPSS Collaboration and Deployment Services Repository abrufen 132
 Objekte in IBM SPSS Collaboration and Deployment Services Repository speichern 128
 Objekte von IBM SPSS Collaboration and Deployment Services Repository entsperren 135
 Objekte von IBM SPSS Collaboration and Deployment Services Repository sperren 135
 Objekteigenschaften, IBM SPSS Collaboration and Deployment Services Repository 137
 Öffnen
 Ausgabe 65
 Knoten 65
 Modelle 65
 Projekte 150
 Statusangaben 65
 Streams 65
 OFFSET (Funktion) 115
 oneof (Funktion) 104
 Operatoren
 in CLEM-Ausdrücken 84

Operatoren (*Forts.*)
 Zeichenfolgen verbinden 97
 Operatorenrangfolge 93
 Optionen 157
 anzeigen 160
 Benutzer 158
 für IBM SPSS Modeler 157
 PMML 161
 Streameigenschaften 44, 46, 47, 48, 49, 50
 Syntax 161
 or (Operator) 100
 Ordner, IBM SPSS Collaboration and Deployment Services Repository 135, 136

P

Palette der Modelle 132
 Paletten 15
 anpassen 162
 Palettenregisterkarte, Anpassung 165
 Parameter
 Eingabeaufforderungen während der Laufzeit 52
 in CLEM-Ausdrücken 84
 in Szenarios verwenden 141
 Modellerstellung 141
 Scoring 141
 Sitzung 52
 Stream 52
 Typ 52
 pi (Funktion) 102
 PMML
 Exportoptionen 161
 Modelle exportieren 146
 Modelle importieren 146
 PMML-Modelle
 lineare Regression 161
 logistische Regression 161
 Portnummer
 IBM SPSS Modeler Server 12, 13
 power (exponentielle Funktion) 101
 PowerPoint-Dateien 150
 Predictive Applications 139
 Projekte 19, 149
 Anmerkungen 152
 Berichte generieren 154
 CRISP-DM, Ansicht 149
 Eigenschaften festlegen 152
 Erstellung 150
 in IBM SPSS Collaboration and Deployment Services Repository 152
 in IBM SPSS Collaboration and Deployment Services Repository speichern 131
 Klassenansicht 150
 neu erstellen 150
 Objekte hinzufügen 151
 Objekteigenschaften 153
 Ordneigenschaften 153
 schließen 154
 Standardordner einrichten 150
 Protokolldateien
 erzeugtes SQL anzeigen 48
 Prozessknoten 37
 Leistung 169
 Punkt 44

Q

Qualitätsknoten
 fehlende Werte 72
 Quellenknoten 37
 Aktualisierung 44
 Datenzuordnung 67
 Quickinfos
 Anmerkungen zu Knoten 63

R

Radiant
 Maßeinheiten 46
 random (Funktion) 104
 random0 (Funktion) 104
 Rangfolge 93
 Reelle Zahlen 89, 90
 Regelerstellungsknoten
 laden 65
 Regelsets
 Auswertung 44
 Regression 183
 rem (Funktion) 101
 replace (Funktion) 105
 replicate (Funktion) 105
 round (Funktion) 101
 Rückgängig 19

S

SAS-Dateien
 Codierung 185
 Schlüsselwörter
 Anmerkungen zu Knoten 63
 Schriftarten 160, 161
 Scoring
 Verzweigung 57, 141, 142, 143
 Scriptfarben
 festlegen 161
 Scripts 24, 75
 SDEV (Funktion) 115
 sdev_n (Funktion) 81, 101
 Sequenzfunktionen 115
 Server
 Anmeldung 12
 Standardverzeichnis 158
 Suche nach COP für Server 13
 Verbindungen hinzufügen 13
 Set aus dichotomen Variablen
 in CLEM-Ausdrücken 82
 Sets 44
 Sets aus kategorialen Variablen
 in CLEM-Ausdrücken 82
 sign (Funktion) 101
 sin (Funktion) 102
 SINCE (Funktion) 115
 Single Sign-on 12
 Single Sign-on, IBM SPSS Collaboration and Deployment Services Repository 125, 127
 sinh (Funktion) 102
 Sitzungsparameter 52
 Skalierung der Streams für Ansicht 22
 skipchar (Funktion) 105
 skipchar_back (Funktion) 105
 Sonderfunktionen 122

- Sonderzeichen
 - aus Zeichenfolgen entfernen 79
- Sortierknoten
 - Leistung 169
- soundex (Funktion) 110
- soundex_difference (Funktion) 110
- Speicher
 - verwalten 157
- Speichern
 - Ausgabeobjekte 65
 - Knoten 64
 - mehrere Objekte 64
 - Statusangaben 64
 - Streams 64
- Sperren von Knoten 43
- Sprachausgabeprogramme 175, 177, 182, 183
 - Beispiel 181, 182
- Sprache
 - Optionen 157
- SQL-Erzeugung
 - Protokollierung 48
 - Vorschau 48
- sqrt (Funktion) 101
- Standard
 - Projektphase 150
- Startdialogfeld 160
- startstring (Funktion) 105
- Statistikdateien
 - Codierung 185
- Statusangaben
 - laden 65
 - speichern 64
- Stream 15
- Streambeschreibungen 55, 57
- Streameigenschaften
 - Analytic Server 49
- Streamerstellungsbereich
 - Einstellungen 49
- Streamnamen 63
- Streamparameter 52
- Streams 11
 - Anmerkungen 58, 63
 - ausführen 57
 - Ausführungszeiten anzeigen 50
 - Bereitstellungsoptionen 139
 - erstellen 37
 - Erstellung 37
 - in IBM SPSS Collaboration and Deployment Services Repository speichern 130
 - Knoten hinzufügen 38, 40
 - Knoten inaktivieren 39
 - Knoten umgehen 39
 - Knoten verbinden 38
 - Kommentare hinzufügen 58
 - laden 65
 - Optionen 44, 46, 47, 48, 49
 - Sicherungsdateien 64
 - Skalierung für Ansicht 22
 - speichern 64
 - umbenennen 55, 63
 - zu Projekten hinzufügen 150, 151
- Streams ausführen 57
- Streamsicherungsdateien
 - wiederherstellen 64
- Streamstandardcodierung 44

- stripchar (Funktion) 105
- strmember (Funktion) 105
- subscrs (Funktion) 105
- substring (Funktion) 105
- substring_between (Funktion) 105
- Suche nach COP für Verbindungen 13
- Suche nach Objekten Objekte hinzufügen
 - and Deployment Services Repository 134
- Suchen
 - Knoten in einem Stream 54
- SUM (Funktion) 115
- sum_n (Funktion) 81, 101
- Symbol für Zifferngruppierung
 - Zahlenanzeigeformate 44
- Symbole
 - Optionen festlegen 22, 49
- Symbolleiste 19
- System
 - Optionen 157
- Szenarios 139
 - Bereitstellungsoptionen 139
 - definiert 126

T

- t-Verteilung
 - Wahrscheinlichkeitsfunktionen 102
- Tabellen 177
 - Ausgabe speichern 65
 - zu Projekten hinzufügen 150
- Tage übertragen 46
- tan (Funktion) 102
- tanh (Funktion) 102
- Tastenkombinationen 174, 175, 177, 179
 - allgemeine Verwendung 68
 - Tastatur 23, 174, 175, 177, 179
- Temp-Verzeichnis 14
- testbit (Funktion) 103
- Textdatendateien
 - Codierung 185
- Textcodierung 44
- THIS (Funktion) 115
- time_before (Funktion) 98
- Tipps
 - allgemeine Verwendung 68
 - für Eingabehilfen 183
- to_date (Funktion) 97, 110
- to_dateline (Funktion) 110
- to_datetime (Funktion) 97
- to_integer (Funktion) 97
- to_number (Funktion) 97
- to_real (Funktion) 97
- to_string (Funktion) 97
- to_time (Funktion) 97, 110
- to_timestamp (Funktion) 97, 110
- Trigonometrische Funktionen 102
- trim (Funktion) 105
- trim_start (Funktion) 105
- trimend (Funktion) 105
- Typische Anwendungen 25
- Typknoten
 - fehlende Werte 73
 - Leistung 169

U

- Überprüfen von CLEM-Ausdrücken 85
- Umbenennen
 - Knoten 63
 - Streams 55
- Umgekehrter Schrägstrich in CLEM-Ausdrücken 90
- undef (Funktion) 121
- ungleich (Operator) 98
- unicode_char (Funktion) 105
- Unicode-Unterstützung 185
- unicode_value (Funktion) 105
- Unterpalette
 - auf Palettenregisterkarte anzeigen 164
 - aus Palettenregisterkarte entfernen 164
 - Erstellung 164
- uppertolower (Funktion) 105
- UTF-8-Codierung 44, 185

V

- value_at (Funktion) 82, 98
- Variablen 25
- Verbindungen
 - mit IBM SPSS Collaboration and Deployment Services Repository 127
 - Server-Cluster 13
 - zu IBM SPSS Modeler Server 12, 13
- Vergleichsfunktionen 98
- Verkettung von Zeichenfolgen 97
- Verrauschte Daten 26
- Versionsbeschriftungen, Objekt von IBM SPSS Collaboration and Deployment Services Repository 138
- Verteilungsfunktionen 102
- Verzeichnis
 - Standard 158
- Verzweigungen, Scoring und Modellierung 57, 141, 142, 143
- Visuelle Programmierung 15
- Vorlagen 67
- Vorlagenfelder 68
- Vorschau
 - Knotendaten 43

W

- Wahrscheinlichkeitsfunktionen 102
- Währungsanzeigeformat 46
- Warnungen 50
 - Optionen festlegen 158
- Werte 77
 - aus einem Data Audit anzeigen 85
 - zu CLEM-Ausdrücken hinzufügen 85
- Wichtige Felder 66, 68
- Wissenschaftliche Notation
 - Anzeigeformat 46
- Wörterbuchdatei 182

Z

- Zahlen 80, 89, 90
- Zahlenanzeigeformate 46
- Zeichen 89, 90
- Zeichenfolgefunktionen 105
- Zeichenfolgen 89, 90
 - abgleichen 79
 - Ersetzung 79
 - in CLEM-Ausdrücken ändern 79
- Zeit- und Datumsfunktionen 91, 92
- Zeitfelder
 - Konvertierung 114
- Zeitformate 46, 91, 92
- Zeitfunktionen 91, 92
 - time_before 98, 110
 - time_hours_difference 110
 - time_in_hours 110
 - time_in_mins 110
 - time_in_secs 110
 - time_mins_difference 110
 - time_secs_difference 110
- Zoomen 19
- Zuordnen von Feldern 66
- Zusammenführungsknoten
 - Leistung 169

