IBM SPSS Modeler Entity Analytics 17.1 Benutzerhandbuch

Hinweis

Vor Verwendung dieser Informationen und des darin beschriebenen Produkts sollten die Informationen unter „Bemerkungen" auf Seite 59 gelesen werden.

Produktinformation

Diese Ausgabe bezieht sich auf Version 17, Release 1, Modifikation 0 von IBM(r) SPSS(r) Modeler und alle nachfolgenden Releases und Modifikationen, bis dieser Hinweis in einer Neuausgabe geändert wird.

Diese Veröffentlichung ist eine Übersetzung des Handbuchs
IBM SPSS Modeler Entity Analytics 17.1, User's Guide,
herausgegeben von International Business Machines Corporation, USA

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
August 2015

Inhaltsverzeichnis

Vorwort

IBM® SPSS Modeler ist die auf Unternehmensebene einsetzbare Data-Mining-Workbench von IBM. Mit SPSS Modeler können Unternehmen und Organisationen die Beziehungen zu ihren Kunden bzw. zu den Bürgern durch ein tief greifendes Verständnis der Daten verbessern. Organisationen verwenden die mithilfe von SPSS Modeler gewonnenen Erkenntnisse zur Bindung profitabler Kunden, zur Ermittlung von Cross-Selling-Möglichkeiten, zur Gewinnung neuer Kunden, zur Ermittlung von Betrugsfällen, zur Reduzierung von Risiken und zur Verbesserung der Verfügbarkeit öffentlicher Dienstleistungen.

Die visuelle Benutzerschnittstelle von SPSS Modeler erleichtert die Anwendung des spezifischen Fachwissens der Benutzer, was zu leistungsstärkeren Vorhersagemodellen führt und die Zeit bis zur Lösungserstellung verkürzt. SPSS Modeler bietet zahlreiche Modellierungsverfahren, beispielsweise Algorithmen für Vorhersage, Klassifizierung, Segmentierung und Assoziationserkennung. Nach der Modellerstellung ermöglicht IBM SPSS Modeler Solution Publisher die unternehmensweite Bereitstellung des Modells für Entscheidungsträger oder in einer Datenbank.

Informationen zu IBM Business Analytics

Die Software IBM Business Analytics liefert umfassende, einheitliche und korrekte Informationen, mit denen Entscheidungsträger die Unternehmensleistung verbessern können. Ein umfassendes Portfolio aus Anwendungen für Business Intelligence, Vorhersageanalyse, Finanz- und Strategiemanagement sowie Analysen bietet Ihnen sofort klare und umsetzbare Einblicke in die aktuelle Leistung und gibt Ihnen die Möglichkeit, zukünftige Ergebnisse vorherzusagen. Durch umfassende Branchenlösungen, bewährte Vorgehensweisen und professionellen Service können Unternehmen jeder Größe die Produktivität maximieren, Entscheidungen automatisieren und bessere Ergebnisse erzielen.

Als Teil dieses Portfolios unterstützt IBM SPSS Predictive Analytics-Software Unternehmen dabei, zukünftige Ereignisse vorherzusagen und proaktiv Maßnahmen zu ergreifen, um bessere Geschäftsergebnisse zu erzielen. Kunden aus Wirtschaft, öffentlichem Dienst und staatlichen Lehr- und Forschungseinrichtungen weltweit nutzen IBM SPSS-Technologie als Wettbewerbsvorteil für die Kundengewinnung, Kundenbindung und Erhöhung der Kundenumsätze bei gleichzeitiger Eindämmung der Betrugsmöglichkeiten und Minderung von Risiken. Durch die Einbindung von IBM SPSS-Software in ihre täglichen Operationen wandeln sich Organisationen zu "Predictive Enterprises", die Entscheidungen auf Geschäftsziele ausrichten und automatisieren und einen messbaren Wettbewerbsvorteil erzielen können. Wenn Sie weitere Informationen wünschen oder Kontakt zu einem Mitarbeiter aufnehmen möchten, besuchen Sie die Seite http://www.ibm.com/spss.

Technical Support

Kunden mit Wartungsvertrag können den Technical Support in Anspruch nehmen. Kunden können sich an den Technical Support wenden, wenn sie Hilfe bei der Arbeit mit IBM Produkten oder bei der Installation in einer der unterstützten Hardwareumgebungen benötigen. Zur Kontaktaufnahme mit dem Technical Support besuchen Sie die IBM Website unter http://www.ibm.com/support. Sie müssen bei der Kontaktaufnahme Ihren Namen, Ihre Organisation und Ihre Supportvereinbarung angeben.

Kapitel 1. Entitätsanalyse

Informationen zur Entitätsanalyse

IBM SPSS Modeler Entity Analytics fügt den IBM SPSS Modeler-Vorhersageanalysen eine zusätzliche 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.

Angenommen, Sie besitzen die folgenden Kundendatensätze aus zwei verschiedenen Quellen und sind sich nicht sicher, ob sie sich auf dieselbe Person oder auf zwei verschiedene Personen beziehen.

Quelle 1

```
Datensatznr.: 70001
Name: Hans Schmidt
Adresse: Hauptstraße 123
Steuernummer: 555001111
Führerschein: 0001133107
Kreditkarte: 10229127
```

Quelle 2

```
Datensatznr.: 9103
Name: JOHANN Schmidt
Geburtsdatum: 17.6.1934
Telefon: 555-1212
Kreditkarte: 10229128
E-Mail: jls@mail.com
IP-Adresse: 9.50.18.77
```

Es gibt keine exakten Übereinstimmungen in den Daten zwischen den beiden Datensätzen. Wenn wir jedoch noch eine dritte Quelle hinzunehmen, finden wir gemeinsame Attribute.

Quelle 3

```
Datensatznr.: 6251
Name: Hans Schmidt
Telefon: 555-1212
Führerschein: 0001133107
Kreditkarte: 10229132
```

Die Führerscheinnummer verknüpft die Datensätze in Quelle 1 und Quelle 3 und die Telefonnummer verknüpft die Quellen 2 und 3. Wir können also hinreichend sicher sein, dass sich alle drei Quellen auf dieselbe Person beziehen.

Aber was tun wir, wenn der Sachverhalt weniger klar ist? Möglicherweise haben wir nur sehr wenige Daten, die wir unserer Entscheidung zugrunde legen können. Betrachten Sie die beiden folgenden Datensätze.

Quelle 4

```
Datensatznr.: S45286
Name: Hans T. Schmidt Jr.
Adresse: Hauptstraße 456
Telefon: 703-555-2000
Geburtsdatum: 12.3.1984


Datensatznr.: S45287
Name: Hans T. Schmidt
Adresse: Hauptstraße 456
Telefon: 703-555-2000
Führerschein: 009900991
```

Offensichtlich ist dies nicht der Herr Schmidt aus den vorangegangenen Datensätzen: die Unterschiede sind groß genug, dass wir dies ausschließen können. Wir haben jedoch trotzdem ein Problem. Zwei verschiedene Datensätze aus derselben Datenquelle scheinen sich auf dieselbe Person zu beziehen. Handelt es sich um doppelte Datensätze? Wir können nicht sicher sein, solange wir keinen anderen zugehörigen Datensatz finden, der uns weitere Informationen liefert, möglicherweise einen Datensatz aus einer anderen Quelle.

Quelle 5

```
Datensatznr.: 769582-2
Name: Hans T. Schmidt Sr.
Adresse: Hauptstraße 456
Telefon: 703-555-2000
Führerschein: 009900991
Geburtsdatum: 25.6.1959
```

Dies löst das Problem. Bei den beiden Datensätzen in Quelle 4 handelt es sich nicht um Duplikate, sondern tatsächlich um einen Vater und einen Sohn mit demselben Namen und unter derselben Adresse, die dieselbe Telefonnummer verwenden. Bei einem manuellen System könnte eine wochenlange Suche erforderlich sein, um den einen Datensatz zu finden, der die Identitäten auflöst. Mit einem automatisierten Entitätsanalysesystem wird der Zeitaufwand für die Auflösung erheblich reduziert

Entitätsanalyse und Vorhersageanalyse

Wenn alle Ihre Daten aus einer einzigen, vollständigen und unzweideutigen Datenquelle bestehen würden, wäre es für IBM SPSS Modeler relativ einfach, etwaige Identitätskonflikte aufzulösen. Sie bräuchten lediglich die Vorhersageanalyse, um Ihre Daten in IBM SPSS Modeler einzulesen, die Verarbeitung durchzuführen und zuverlässige Ergebnisse zu erhalten.

In der realen Welt sieht die Sache jedoch normalerweise deutlich anders aus. Die Daten sind typischerweise alles andere als vollständig, häufig mehrdeutig und oftmals über viele verschiedene Datenquellen verstreut, in denen viele verschiedene Attribute mit wenigen sich überlappenden Feldern dokumentiert sind. Der Nutzen der Entitätsanalyse liegt zum Teil darin, dass Daten aus all den verschiedenen Quellen in einem einzigen, zentralen Bereich, dem sogenannten **Repository**, gesammelt werden. Das Entitätsanalysesystem untersucht die Daten dann bis ins kleinste Detail, um Konflikte aufzulösen. Dabei erhalten Datensätze, die von derselben Person oder Organisation stammen, eine eindeutige ID.

In der folgenden Tabelle werden die Unterschiede zwischen den beiden Analysetypen veranschaulicht.

Tabelle 1. Unterschiede zwischen Vorhersage- und Entitätsanalyse.

Eigenschaft	Vorhersageanalyse	Entitätsanalyse
Typen von Trainingsdaten	Beruht auf relativ kleinen Sets und numerischen Bereichen	Kann große Sets (Felder ohne Typ) wie Namen und Adressen nutzen

Tabelle 1. Unterschiede zwischen Vorhersage- und Entitätsanalyse (Forts.).

Eigenschaft	**Vorhersageanalyse**	**Entitätsanalyse**
Größe von Trainingsdaten	Ignoriert in der Regel große Sets (Felder ohne Typ)	Alle Daten werden verwendet
Verallgemeinerung	Algorithmus verallgemeinert über die gesamten Trainingsdaten hinweg, um ein prägnantes Modell zu erstellen.	Die Daten werden in Strukturen als persistent definiert, die sich für Entitätsabgleich und Beziehungserkennung eignen.
Betrugserkennung	Datensätze werden als potenziell betrügerisch gekennzeichnet, wenn sie typische Eigenschaften einer betrügerischen Anwendung aufweisen.	Datensätze werden als potenziell betrügerisch gekennzeichnet, wenn sie in Bezug zu bekannten betrügerischen Datensätzen stehen oder von denselben Personen, jedoch unter verschiedenen Identitäten, stammen.

Kapitel 2. Entitätsanalyse mit IBM SPSS Modeler

Verwendung der Entitätsanalyse mit IBM SPSS Modeler

Sie vermuten, dass bei Ihren Daten Identitätsprobleme vorliegen könnten. Personen können z. B. mehrmals vorkommen oder unterschiedliche Personen können scheinbar zusammengeführt worden sein oder fehlen. Wie kann IBM SPSS Modeler Entity Analytics Sie bei der Lösung dieser Probleme unterstützen? Im Folgenden wird eine Vorgehensweise vorgeschlagen, die Sie jedoch eventuell abwandeln müssen, um sie an Ihre speziellen Anforderungen anzupassen.

- Datenquelle in IBM SPSS Modeler einlesen
- Repository erstellen, in dem die Daten gespeichert werden können
- IBM SPSS Modeler mit dem Repository verbinden
- Datenfelder den Repository-Merkmalen zuordnen
- Daten in das Repository exportieren und Identitäten auflösen
- Aufgelöste Identitäten analysieren
- Neue Fälle anhand des Repositorys auflösen
- Etwaige notwendige Warnungen generieren (als Stapelvorgang oder in Echtzeit)

Sie benötigen nun Kenntnisse darüber, wie IBM SPSS Modeler funktioniert. IBM SPSS Modeler ist ein sehr benutzerfreundliches Tool, das auf der grafischen Darstellung eines Datenstreams beruht, der eine Reihe von Knoten durchläuft. Jeder Knoten steht für einen bestimmten Schritt des Arbeitsablaufs.

IBM SPSS Modeler selbst enthält eine große Palette an Knoten, die alle standardmäßigen Data-Mining-Funktionen abdecken. IBM SPSS Modeler Entity Analytics fügt Knoten speziell für die Verwendung in Entitätsanalysen hinzu. Hierbei handelt es sich um den EA-Exportknoten, den EA-Quellenknoten und den Prozessknoten "Streaming von EA".

Der Prozess wird durch die folgende Abbildung dargestellt.

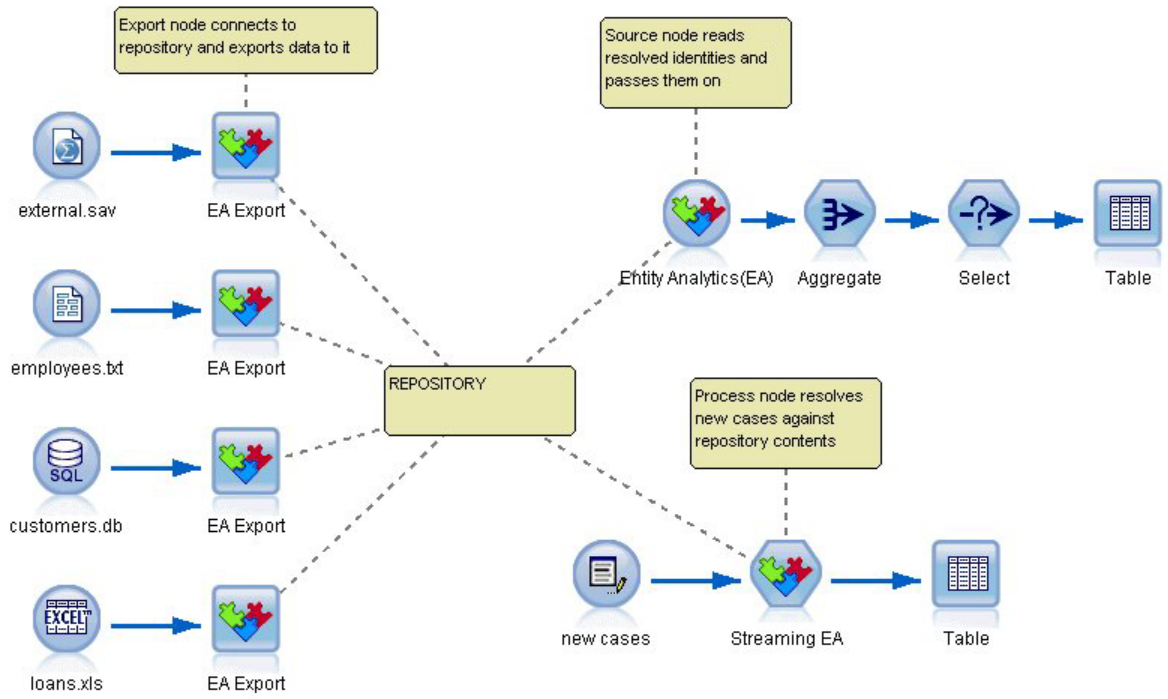


Abbildung 1. Der Entitätsanalyseprozess

Stufe 1: Einlesen der Datenquelle in SPSS Modeler

Ihre erste Aufgabe besteht darin, Ihre Daten mithilfe mindestens eines Quellenknotens in SPSS Modeler einzulesen. Quellenknoten werden in SPSS Modeler durch ein rundes Symbol gekennzeichnet.

Die Daten können in jedem beliebigen Format vorliegen, das von SPSS Modeler unterstützt wird, beispielsweise Textdateien, Datenbanktabellen, Kalkulationstabellen, XML-Dateien usw. Für jedes verschiedene Format ist jedoch jeweils ein entsprechender SPSS Modeler-Quellenknoten erforderlich. In der Abbildung handelt es sich um einen Datenbankquellenknoten.

Jede Datenquellendatei benötigt ein Feld, das die einzelnen Datensätze eindeutig kennzeichnet. Wenn eine Datenquelle kein derartiges Feld aufweist, können Sie problemlos eines in SPSS Modeler hinzufügen. Weitere Informationen finden Sie im Thema „Hinzufügen einer eindeutigen Datensatz-ID" auf Seite 14.

Weitere Informationen finden Sie im Thema „Verbinden mit einer Datenquelle" auf Seite 14.

Anmerkung: Es werden ausschließlich lateinische Zeichendaten unterstützt. Wenn die Daten sowohl aus Datensätzen mit lateinischen Zeichensätzen (z. B. Westeuropäisch) als auch aus Datensätzen mit nicht lateinischen Zeichensätzen bestehen, werden nur die Einträge für Daten aufgelöst, die lateinische Zeichen verwenden.

Stufe 2: Erstellen des Repositorys

Der Schwerpunkt all Ihrer Bemühungen im Bereich der Entitätsanalyse liegt auf dem Repository, dem zentralen Speicherbereich, in dem Sie alle Datensätze sammeln können.

Zur Erstellung eines Repositorys verbinden Sie zunächst die Datenquelle mit einem EA-Exportknoten, der durch das quadratische Symbol repräsentiert wird.

Aus dem Exportknoten können Sie ein neues Repository erstellen (oder ein bestehendes auswählen), das zur Aufnahme der exportierten Daten bereit ist.

Der Vorgang der Repository-Erstellung wird weiter unten ausführlich beschrieben. Weitere Informationen finden Sie im Thema „Einrichten eines Entitätsrepositorys (EA-Exportknotens)" auf Seite 13.

Anmerkung: Bei der Ausführung im Modus für einen fernen Server müssen Sie das Repository auf dem Modeler Server-Computer erstellen (d. h., es muss bei der Erstellung des Repositorys eine Verbindung von Modeler Client zu Modeler Server bestehen, sodass das EA-Repository auf dem Server-Computer erstellt wird).

Nachdem Sie ein Repository eingerichtet haben, können Sie seine Inhalte auf verschiedene Weisen verwalten. Weitere Informationen finden Sie im Thema „Konfigurieren eines Entitätsrepositorys" auf Seite 19.

Stufe 3: Verbinden von SPSS Modeler mit dem Repository

Nach der Erstellung des Repositorys müssen Sie es nun mit dem SPSS Modeler-Stream verbinden.

Weitere Informationen finden Sie im Thema „Optionen für das Entitätsrepository" auf Seite 16.

Stufe 4: Zuordnen der Eingabefelder zu Repository-Merkmalen

Datenquellen können mehrere Arten von Entitätsinformationen enthalten. Einige Informationstypen kommen bei den meisten Entitätsdatenquellen vor, während andere möglicherweise nur bei einer bestimmten Datenquelle vorkommen. Bei einem Entitätsrepository werden diese verschiedenen Informationstypen als **Merkmale** bezeichnet. Das Repository stellt eine Reihe von Merkmalen als Standard bereit, und Sie können auch eigene Merkmale erstellen.

Ein Repository-Merkmal ist ein einzelner Informationstyp, der zusammen mit einer Entitätsdatenquelle verwendet werden kann. Einige Merkmale (z. B. Vorname, Nachname, Geburtsdatum usw.) können mit vielen verschiedenen Datenquellen verwendet werden, während andere Merkmale nur bei einer bestimmten Datenquelle vorkommen. Ein Merkmal entspricht üblicherweise einem Feld in einem Datensatz oder einer Spalte in einer Datenbanktabelle.

Wenn Sie ein Repository erstellt und eine Verbindung damit hergestellt haben, kennzeichnen Sie ein Feld Ihrer Eingabedaten als Feld mit dem **eindeutigen Schlüssel**, das in der nachfolgenden Analyse verwendet wird. Außerdem ordnen Sie die Eingabedatenfeldern den entsprechenden Merkmalen im Repository zu. Anhand der Zuordnung zu vordefinierten Merkmalen erkennt das Entitätsrepository, welche Felder verglichen werden sollen und vor allem wie sie verglichen werden sollen. Der EA-Exportknoten enthält eine Zuordnungstabelle, in der Sie die Zuordnungen erstellen können.

Weitere Informationen finden Sie im Thema „Erstellen einer Zuordnung zwischen Eingabefeldern und Merkmalen (EA-Exportknoten)" auf Seite 17.

Stufe 5: Exportieren von Daten in das Repository und Auflösen von Identitäten

Für die einzelnen Datenquellen ist jeweils ein eigener EA-Exportknoten erforderlich. Wenn Ihre Daten also über mehrere verschiedene Quellen verstreut sind, enthält Ihr Stream möglicherweise mehrere Datenquellen, von denen jede mit einem separaten EA-Exportknoten verbunden ist. Weitere Informationen finden Sie im Thema „Verwendung der Entitätsanalyse mit IBM SPSS Modeler" auf Seite 5.

Wenn Sie über mehrere Datenquellen verfügen, können Sie auswählen, ob die Datensätze aus einer, einigen oder allen Datenquellen eingelesen werden sollen. Das Entitätsanalysesystem analysiert die von Ihnen ausgewählten Datensätze und fügt zu jedem ein ID-Feld mit der Bezeichnung "$EA_ID" hinzu. Wenn nun zwei oder mehr Datensätze, die sich auf zuvor mehrdeutige Identitäten bezogen, aufgelöst werden

können, sind die zu diesen Datensätzen hinzugefügten IDs im gesamten Repository eindeutig. Das System fügt auch ein Feld hinzu, das die Datenquelle angibt, aus der der Datensatz stammt.

Sie verbinden die einzelnen Datenquellenknoten jeweils mit ihrem eigenen EA-Exportknoten, ordnen die Eingabefelder den Repository-Funktionen zu und führen dann den Stream aus, um die Daten aus SPSS Modeler in das Repository zu exportieren und etwaige Identitätskonflikte in einem einzigen Vorgang aufzulösen. Stellen Sie sich zur Veranschaulichung der Funktionsweise vor, dass Sie die folgenden Datensätze in vier verschiedenen Datenquellen besitzen.

Externe Daten

Tabelle 2. Externe Daten

Name	Telefon	Kreditrisiko
Mike	555-1234	560
Joe	555-4567	780

Mitarbeiter

Tabelle 3. Mitarbeiter

Name	Adresse	Telefon
Michael	1234 5th Street	555-1234
Fred	543 1st Avenue	555-9876

Kunden

Tabelle 4. Kunden

Name	Adresse	Sparkonto
Susan	1234 5th Street	$1234
Joe	777 Oak Street	$5

Kredite

Tabelle 5. Kredite

Name	Adresse	Telefon	Kredit
Sue	1234 5th Street	555-1234	$10.000
Joseph	777 Oak Street	555-4567	$50.000

Wie wir gesehen haben, exportieren Sie die einzelnen Datenquellen nacheinander in das Repository. Dabei aktualisiert das Repository die Auflösung der einzelnen Datensätze. Im Repository wird jedem Datensatz ein ID-Feld (mit dem Namen *$EA-ID*) und ein Feld zur Quellenangabe (mit dem Namen *$EA-SRC*) vorangestellt, das die Datenquelle anzeigt, aus der der Datensatz stammt. In unserem Beispiel sieht der Repository-Inhalt nach dem Export aller vier Datenquellen wie folgt aus:

Tabelle 6. Beispiel für Repository-Inhalte nach Exportschritt.

$EA-ID	$EA-SRC	Name	Telefon	Adresse	Kreditrisiko	Sparkonto	Kredit
1	Mitarbeiter	Michael	555-1234	1234 5th St			
1	Extern	Mike	555-1234		560		
2	Kunden	Joe		777 Oak St		$5	

Tabelle 6. Beispiel für Repository-Inhalte nach Exportschritt (Forts.).

$EA-ID	$EA-SRC	Name	Telefon	Adresse	Kreditrisiko	Sparkonto	Kredit
2	Extern	Joe	555-4567		780		
2	Kredite	Joseph	555-4567	777 Oak St			$50.000
3	Mitarbeiter	Fred	555-9876	543 1st Ave			
4	Kunden	Susan		1234 5th St		$1234	
4	Kredite	Sue	555-1234	1234 5th St			$10.000

Das Entitätsanalysesystem hat anhand der gemeinsamen Telefonnummer ermittelt, dass *Mike* im Dataset *Extern* dieselbe Person ist wie *Michael* im Dataset *Mitarbeiter*, und ihm die ID 1 zugewiesen.

Der Fall von *Joe* im Dataset *Extern* ist ein wenig komplizierter. Handelt es sich um dieselbe Person wie bei *Joe* in *Kunden*? Dies lässt sich aus diesen beiden Datenquellen allein unmöglich sagen. Wir verfügen jedoch über eine dritte Quelle, *Kredite*, die einen *Joseph* enthält. Nun haben wir eine Übereinstimmung: Die Telefonnummer von Joseph stimmt mit der von Joe im Dataset *Extern* überein. Auf dieser Grundlage ermittelt das System, dass es sich bei allen um dieselbe Person handelt, und weist ihnen die ID 2 zu.

Es gibt nicht mehrere Datensätze für *Fred*; er erhält also ID 3. *Susan* aus *Kunden* wird als dieselbe Person identifiziert wie *Sue* aus *Kredite*, da sie dieselbe Adresse haben; ihr wird also ID 4 zugewiesen.

Hinweis: Dies ist ein Beispiel für einen optimistischen Abgleich zum Zwecke der Illustration. Sie könnten ein pessimistischeres Regelset auswählen, sodass ein einfacher Name und eine Telefonnummer bzw. Adresse allein noch nicht ausreicht, um eine exakte Übereinstimmung festzustellen und beiden Datensätzen dieselbe ID zuzuweisen.

Stufe 6: Analysieren der aufgelösten Identitäten

Nach der Auflösung der Identitätskonflikte im Repository können Sie nun weitere Analysen und Verarbeitungsschritte an den Ergebnissen durchführen. Wenn Sie beispielsweise beim Vorliegen doppelter Datensätze für dieselbe Identität mögliche betrügerische Aktivitäten vermuten, kann es sinnvoll sein, einen Bericht zu erstellen, in dem die Duplikate aufgelistet werden.

Dazu erstellen Sie zuerst einen EA-Quellenknoten und verknüpfen ihn mit dem Repository.

Die allgemeine Ausgabe des Knotens besteht aus den folgenden Feldern:

- Vom System hinzugefügtes ID-Feld (*$EA-ID* im Beispiel für Schritt 5)
- Vom System hinzugefügtes Feld zur Quellenangabe (*$EA-SRC* im Beispiel für Schritt 5)
- In Schritt 4 hinzugefügtes Feld mit eindeutigem Schlüssel

Wenn Sie sich Beziehungen ansehen, wird darüber hinaus die folgende Ausgabe erzeugt. Weitere Informationen finden Sie im Thema „Auswählen einer Datenquelle" auf Seite 27.

- Der Abgrenzungsgrad zwischen Entitäten (*$EA-DEGREE*)
- Das übergeordnete Feld (*$EA-PARENT*)
- Das untergeordnete Feld (*$EA-CHILD*)
- Die Regel, die die Beziehung angibt (*$EA-RULE*)

Um die Ausgabe in SPSS Modeler anzuzeigen, können Sie einen SPSS Modeler-Ausgabeknoten, beispielsweise einen Tabellenknoten oder einen Berichtsknoten, anfügen und diesen Teil des Streams ausführen. Um die Ausgabe zusammenzufassen, die sehr groß werden könnte, können Sie Knoten für Datensatzoperationen mit aufnehmen, beispielsweise Aggregatknoten und Auswahlknoten.

Der EA-Quellenknoten wird weiter unten ausführlich beschrieben. Weitere Informationen finden Sie im Thema „Analysieren der aufgelösten Identitäten (EA-Quellenknoten)” auf Seite 26.

Stufe 7: Auflösen neuer Fälle mithilfe des Repositorys

Sie haben nun also die Identitäten aller Datensätze in allen Datenquellen aufgelöst. Doch was geschieht, wenn Sie eine Menge an neuen Datensätzen vergleichen möchten, um zu ermitteln, in welchem Zusammenhang sie mit Ihren bisherigen Erkenntnissen stehen, und so das Scoring zu verbessern? Hier kommt der Knoten "Streaming von EA" ins Spiel.

Zunächst fügen Sie einen neuen SPSS Modeler-Datenquellenknoten hinzu, um Ihre neuen Daten in den Stream einzulesen. Anschließend verbinden Sie diesen Quellenknoten mit einem Knoten des Typs "Streaming von EA". Wenn Sie die Ausgabe anzeigen möchten, fügen Sie einen Tabellenknoten zum Stream hinzu.

Wenn Sie diesen Teil des Streams ausführen, liest der Knoten "Streaming von EA" neue Datensätze ein und vergleicht sie mit dem Inhalt des Repositorys. Wenn übereinstimmende Datensätze im Repository gefunden werden, gibt der Knoten alle übereinstimmenden Datensätze zusammen mit dem neuen Datensatz aus und fügt die Felder für die ID und die Quellenangabe hinzu. Wenn keine Übereinstimmung gefunden wird, gibt der Prozessknoten nur den neuen Datensatz mit den hinzugefügten Feldern für die ID und die Quellenangabe aus.

Stellen Sie sich zur Veranschaulichung vor, dass das Repository derzeit aus den Inhalten besteht, die durch den EA-Quellenknoten ausgegeben wurden. Siehe Tabelle 6 auf Seite 8.

Nun erhalten wir folgende neuen Datensätze. Beziehen Sie sich auf Personen, die wir bereits kennen?

Tabelle 7. Mit Scoring zu bewertende neue Datensätze.

Name	Adresse	Telefon	Kredit
Suzan	1234 5th Street	555-1234	$100.000
Mark	888 9th Ave	555-9999	$60.000

Beim Vergleich der neuen Daten mit den bestehenden Repository-Inhalten findet der Knoten "Streaming von EA" eine Übereinstimmung zwischen dem ersten neuen Datensatz und der Person, die in den bestehenden Datensätzen die ID 4 trägt. Für den zweiten neuen Datensatz wird jedoch keine Übereinstimmung gefunden, weshalb ihr die neue, eindeutige ID 5 zugewiesen wird.

Der Knoten "Streaming von EA" fügt die Felder für die ID und die Quellenangabe hinzu und gibt die neuen Datensätze zusammen mit allen übereinstimmenden Datensätzen aus. Die Ausgabe sieht also wie folgt aus:

Tabelle 8. Ausgabe des Knoten "Streaming von EA".

$EA-ID	$EA-SRC	Name	Telefon	Adresse	Kreditrisiko	Sparkonto	Kredit
4	Kund	Susan		1234 5th St		$1234	
4	Kredit	Sue	555-1234	1234 5th St			$10.000
4	Neuer Kredit	Suzan	555-1234	1234 5th Street			$100.000
5	Neuer Kredit	Mark	555-9999	888 9th Ave			$60.000

Diese Ausgabe kann dann unter Verwendung der Entitätsanalyse-ID als Aggregationsschlüssel aggregiert und zur weiteren Verarbeitung an andere nachgeordnete Knoten weitergegeben werden.

Der Knoten "Streaming von EA" wird weiter unten ausführlich beschrieben.

Stufe 8: Generieren von Alerts

Auch hier können eventuell wieder potenziell verdächtige Aktivitäten aufgedeckt werden. In diesem Fall hat die Person mit ID 4 bereits einen Kredit über 10.000 Dollar aufgenommen und beantragt nun, unter einem leicht abweichenden Namen, einen weiteren Kredit in zehnfacher Höhe. Dies kann natürlich vollkommen in Ordnung sein und ohne betrügerische Absicht geschehen. Wenn eine solche Handlungsweise jedoch gemäß Ihren Geschäftsregeln als verdächtig gilt, kann es sinnvoll sein, einen genaueren Blick darauf zu werfen.

Sie könnten beispielsweise einen SPSS Modeler-Tabellenknoten oder -Berichtsknoten anfügen und ausführen, den Inhalt des zugehörigen Ausgabefensters ausdrucken und ihn von jemandem lesen und manuell Warnungen erstellen lassen. Alternativ könnten Sie die Ausgabe des Knotens "Streaming von EA" an ein Risikobewertungsmodell weiterleiten, das Sie zuvor in IBM SPSS Modeler erstellt haben, wodurch ein Score-Set ausgegeben wird, das Ihren Geschäftsregeln besser entspricht. Eine weitere Möglichkeit besteht darin, die Ausgabe zur weiteren Verarbeitung in eine Datenbank oder ein anderes Medium zu exportieren. Mit IBM SPSS Modeler steht Ihnen eine Vielzahl an Handlungsmöglichkeiten für Ihre speziellen Anforderungen zur Verfügung.

Kapitel 3. Entitätsanalyseaufgaben

Informationen zu den Aufgaben

In diesem Abschnitt werden die folgenden Aufgaben aus dem Bereich der Entitätsanalyse beschrieben.

- Einrichten eines Entitätsrepositorys
- Konfigurieren eines Entitätsrepositorys
- Analysieren der aufgelösten Identitäten
- Auflösen neuer Fälle anhand des Entitätsrepositorys
- Löschen des Inhalts eines Entitätsrepositorys
- Löschen eines Entitätsrepositorys
- Verwenden der Entitätsanalyse mit anderen IBM SPSS-Produkten
- Verwalten der Entitätsanalyse

Einrichten eines Entitätsrepositorys (EA-Exportknotens)

Die Einrichtung eines Entitätsrepositorys besteht aus folgenden Aufgaben:

1. Verbinden mit einer Datenquelle. Weitere Informationen finden Sie im Thema „Verbinden mit einer Datenquelle" auf Seite 14.
2. Erstellen des Repositorys. Weitere Informationen finden Sie im Thema „Erstellen des Repositorys" auf Seite 14.
3. Zuordnung der Eingabefelder in der Datenquelle zu Merkmalen im Repository. Weitere Informationen finden Sie im Thema „Erstellen einer Zuordnung zwischen Eingabefeldern und Merkmalen (EA-Exportknoten)" auf Seite 17.

Nachdem Sie die Zuordnungen eingerichtet haben, können Sie sie entweder für die aktuelle Datenquelle anzeigen oder für alle Datenquellen, die dem Repository bekannt sind. Weitere Informationen finden Sie im Thema „Anzeigen der Feldzuordnungen (EA-Exportknoten)" auf Seite 18.

Anmerkung: Ab Version 16 unterstützt SPSS Entity Analytics Repositorys in IBM DB2. Ein Repository ist spezifisch für eine Version von SPSS Modeler und kann nicht aus einer früheren Version importiert werden, Wenn Sie ein Repository haben und ein Upgrade auf SPSS Entity Analytics Version 16 durchführen, müssen Sie das Repository daher in der neuen DB2-Datenbank neu erstellen.

Entitätsrepository

Das Repository bietet einen zentralen Speicherbereich, der als Datencache für alle Entitätsinformationen dient. Da es sich um ein Live-Repository handelt, hat es nur einen einzigen Zustand. Es gibt also bei einem Entitätsrepository keine Versionserstellung und -verwaltung. Das Repository enthält den aktuellen Status sämtlicher Eingabedateien und kann sehr groß werden.

Sie können die Repository-Inhalte mithilfe einer benutzerfreundlichen grafischen Schnittstelle verwalten. Weitere Informationen finden Sie im Thema „Konfigurieren eines Entitätsrepositorys" auf Seite 19.

Wichtig: Ab Version 16 unterstützt IBM SPSS Modeler Entity Analytics Repositorys mit dem Produkt IBM DB2; vorherige Versionen von SPSS Entity Analytics unterstützten Repositorys mit IBM solidDB als Host. Wenn Sie über ein vorhandenes solidDB-Repository verfügen, müssen Sie dieses Repository in der neuen DB2-Datenbank erneut erstellen, wenn Sie ein Upgrade auf SPSS Entity Analytics Version 16 oder höher durchführen.

Hinweis: Die im Lieferumfang von IBM SPSS Modeler Premium enthaltene Version von IBM SPSS Modeler Entity Analytics unterstützt nur ein einzelnes Repository mit dem Produkt IBM DB2 als Host, das mit SPSS Entity Analytics gebündelt ist. Bei dieser Version müssen Sie ein bestehendes Repository löschen, bevor Sie ein neues erstellen. Es ist ein separat lizenziertes Upgrade für SPSS Entity Analytics (IBM SPSS Modeler Entity Analytics Unleashed) verfügbar, mit dem mehrere Repositorys auf demselben System vorhanden sein können. Jedes Repository kann mehr als 10 Millionen Zeilen enthalten und mehr als vier Prozessorkerne verwenden. Details hierzu erhalten Sie bei Ihrem lokalen IBM Support-Mitarbeiter.

Verbinden mit einer Datenquelle

Sie beginnen, indem Sie Ihre Quellendaten mit einem Quellenknoten in SPSS Modeler einlesen.

Herstellen der Verbindung zu einer Datenquelle

1. Doppelklicken Sie auf der Registerkarte "Quellen" auf der Knotenpalette unten im SPSS Modeler-Hauptfenster auf das Symbol, das dem Typ der Quellendaten entspricht. Dadurch wird ein Quellenknoten zum Bildschirmerstellungsbereich hinzugefügt.
2. Doppelklicken Sie im Bildschirmerstellungsbereich auf das Symbol, um das zugehörige Dialogfeld zu öffnen.
3. Geben Sie im Feld "Datei" den Standort und den Namen der Quellendatendatei ein.
4. Füllen Sie den Rest des Dialogfelds nach Bedarf aus (klicken Sie auf "Hilfe", wenn Sie weitere Informationen benötigen) und klicken Sie dann auf "OK".
5. Wenn die Quellendatendatei kein Feld enthält, mit dem die einzelnen Datensätze eindeutig identifiziert werden, fügen Sie mithilfe eines Ableitungsknotens ein solches Feld hinzu. Weitere Informationen finden Sie im Thema „Hinzufügen einer eindeutigen Datensatz-ID".

Anmerkung: Es werden ausschließlich lateinische Zeichendaten unterstützt. Wenn die Daten sowohl aus Datensätzen mit lateinischen Zeichensätzen (z. B. Westeuropäisch) als auch aus Datensätzen mit nicht lateinischen Zeichensätzen bestehen, werden nur die Einträge für Daten aufgelöst, die lateinische Zeichen verwenden.

Hinzufügen einer eindeutigen Datensatz-ID

Jede Datenquellendatei, die in das Entitätsrepository eingegeben wird, benötigt ein Feld, das die einzelnen Datensätze eindeutig identifiziert. Falls eine Datenquelle kein derartiges Feld enthält, können Sie eines mithilfe eines SPSS Modeler-Ableitungsknotens hinzufügen.

Hinzufügen einer eindeutigen Datensatz-ID zu einer Datenquellendatei

1. Klicken Sie im Bildschirmerstellungsbereich auf den Quellenknoten, den Sie in der vorangegangenen Aufgabe hinzugefügt haben.
2. Doppelklicken Sie auf der Registerkarte **Feldoperationen** der Knotenpalette auf das Symbol **Ableiten**, um einen Ableitungsknoten zum Quellenknoten hinzuzufügen.
3. Doppelklicken Sie im Bildschirmerstellungsbereich auf den Ableitungsknoten, um das zugehörige Dialogfeld zu öffnen.
4. Ersetzen Sie im Feld **Ableiten** den Standardnamen durch einen aussagekräftigen Namen (z. B. **ID**) für das hinzuzufügende ID-Feld.
5. Stellen Sie sicher, dass das Feld **Ableitungstyp** auf **Formel** gesetzt ist.
6. Setzen Sie **Feldtyp** auf `Stetig`.
7. Geben Sie im Textfeld **Formel** `@INDEX` ein und klicken Sie auf **OK**.

Erstellen des Repositorys

Sie müssen ein Repository erstellen, in dem alle Eingabedaten gespeichert werden.

Anmerkung: Bei der Ausführung im Modus für einen fernen Server müssen Sie das Repository auf dem Modeler Server-Computer erstellen (d. h., es muss bei der Erstellung des Repositorys eine Verbindung von Modeler Client zu Modeler Server bestehen, sodass das EA-Repository auf dem Server-Computer erstellt wird).

Erstellen eines Repositorys

1. Platzieren Sie auf der Registerkarte "Export" der SPSS Modeler-Knotenpalette einen EA-Exportknoten im Streamerstellungsbereich.

   Hinweis: Wenn Sie zum ersten Mal ein Repository erstellen, verwenden Sie einen EA-Exportknoten und verbinden Sie ihn mit dem SPSS Modeler-Quellenknoten, der die Daten enthält, die in das Repository eingegeben werden sollen (bzw. mit dem Ableitungsknoten, wenn Sie einen hinzugefügt haben, um ein Feld mit einer eindeutigen ID zu erhalten). Gehen Sie zur Verbindung der Knoten wie folgt vor.

   a. Klicken Sie mit der rechten Maustaste auf den SPSS Modeler-Quellenknoten.
   b. Wählen Sie die Option "Verbinden" aus.
   c. Klicken Sie auf den EA-Exportknoten.
2. Doppelklicken Sie dann auf den EA-Exportknoten, um das zugehörige Dialogfeld zu öffnen.
3. Klicken Sie auf die Liste **Entitätsrepository**.
4. Klicken Sie auf **<Durchsuchen...>**, um das Dialogfeld "Entitätsrepositorys" anzuzeigen.
5. Klicken Sie im Dialogfeld "Entitätsrepositorys" auf das Feld "Repository-Name".
6. Wählen Sie die Option **<Neues Repository erstellen...>** aus, um den Assistenten zum Erstellen von Repositorys anzuzeigen.

Assistent zum Erstellen von Repositorys

Schritt 1

Hier können Sie auswählen, ob ein lokales Repository mit dem Produkt IBM DB2 erstellt werden soll, das mit IBM SPSS Modeler Entity Analytics gebündelt ist, oder ob eine externe Datenbank für das Repository verwendet werden soll.

Lokales Repository erstellen. Geben Sie den Benutzernamen und das Kennwort eines Administrators für die IBM DB2-Datenbank an, die als Host für das zu erstellende Repository fungieren soll. Bestätigen Sie das Kennwort und klicken Sie auf **Weiter**.

Anmerkung: Im Benutzernamen sind weder Gedankenstriche noch Unterstriche zulässig.

Welche Berechtigungsnachweise für die IBM DB2-Datenbank verwendet werden müssen, hängt von Ihrem Betriebssystem ab. UNIX-Benutzer müssen den Benutzernamen g2user und das Kennwort G2password verwenden.

Für die Repository-Administrationsaufgaben innerhalb der EA-Knoten, wie z. B. das Erstellen oder Löschen eines Repositorys, sind zusätzliche Berechtigungen erforderlich. Unter UNIX muss der an IBM SPSS Modeler Server angemeldete Benutzer entweder der Rootbenutzer oder der Benutzer g2user sein und er muss ein Mitglied der Gruppe db2iadm1 sein. Unter Windows muss der Benutzer, der sich an IBM SPSS Modeler Server anmeldet, ein Mitglied der Gruppe DB2ADMNS sein, um die Repository-Administration vornehmen zu können.

Wenn Sie anschließend die Administratorberechtigungen ändern müssen, verwenden Sie dazu den Befehlszeileneditor für die Datenbank. Weitere Informationen finden Sie im Thema „Verwalten der Administratorberechtigungen für die Repository-Datenbank” auf Seite 34.

Anmerkung: Es ist nur eine Kombination aus Benutzername und Kennwort möglich. Alle Benutzer, die sich am Repository anmelden, verwenden denselben Benutzernamen und dasselbe Kennwort.

Externes Repository hinzufügen. Verwenden Sie diese Option, wenn Sie eine externe Datenbank als Host für das Repository verwenden möchten. Geben Sie den Speicherort der .ini-Datei der Datenbank in das Feld **.ini-Datei des Repositorys auswählen** ein und klicken Sie auf **Weiter**.

Schritt 2

Neuer Repository-Name. Geben Sie einen eindeutigen Namen für das neue Repository ein.

Konfiguration importieren aus. (Nur lokales Repository) Wenn die Konfiguration auf der eines bestehenden Repositorys beruhen soll, wählen Sie hier das Repository aus. Falls nicht, wählen Sie **Standard** aus. Weitere Informationen finden Sie im Thema „Konfigurieren eines Entitätsrepositorys" auf Seite 19.

Wenn Sie ein bestehendes Repository auswählen, geben Sie die Verbindungsdetails ein, sofern diese von den im vorherigen Bildschirm eingegebenen abweichen.

Klicken Sie auf **OK**, um das neue Repository zu erstellen und das Dialogfeld "Entitätsauflösungsinstanzen" anzuzeigen, über das Sie eine Verbindung zum Repository herstellen können.

Optionen für das Entitätsrepository

Das Dialogfeld "Entitätsrepositorys" enthält eine Reihe von Optionen, mit denen Sie ein Entitätsrepository erstellen, eine Verbindung damit herstellen und es konfigurieren und verwalten können.

Mit Repository verbinden. Mit diesen Optionen können Sie ein neues Entitätsrepository erstellen bzw. eine Verbindung mit einem bestehenden herstellen.

- **Repository-Name.** Zeigt das aktuelle Entitätsrepository an, sofern eines vorhanden ist. Wenn mehrere Repositorys vorhanden sind, können Sie in der Liste ein anderes Repository auswählen.

  Um ein neues Repository zu erstellen, wählen Sie die Option **<Neues Repository erstellen...>** aus. Dadurch wird ein Assistent gestartet, der Sie durch den Erstellungsvorgang führt.
- **Benutzername.** Geben Sie einen gültigen Benutzernamen für das ausgewählte Repository ein.
- **Kennwort.** Das Kennwort für diesen Benutzernamen.
- **Verbinden.** Klicken Sie auf diese Option, um eine Verbindung mit dem aktuellen Repository herzustellen.

Repository verwalten. In der Tabelle werden die Datenquellen aufgelistet, die in das aktuelle Repository (dasjenige, mit dem Sie verbunden sind) geladen wurden. Dabei wird auch jeweils die Anzahl der Datensätze in den einzelnen Datenquellen angezeigt.

- **Aktualisieren.** Aktualisiert die Tabellendaten zu Datenquellen und Datengröße, beispielsweise, wenn Sie eine neue Datenquelle hinzugefügt oder die Größe einer bestehenden Datenquelle geändert haben.
- **Alle löschen.** Entfernt alle Quellendaten aus dem Repository, behält jedoch alle Konfigurationsdetails bei. Sie können diese Option verwenden, wenn Sie die Konfigurationsinformationen noch brauchen können, jedoch alle Datensätze aus dem Repository entfernen möchten. Weitere Informationen finden Sie im Thema „Löschen des Inhalts eines Entitätsrepositorys" auf Seite 36.
- **Nicht verwendete löschen.** Entfernt die hervorgehobenen Quellendaten aus dem Repository, behält jedoch alle Konfigurationsdetails bei. Weitere Informationen finden Sie im Thema „Löschen nicht verwendeter Datenquellen aus einem Repository" auf Seite 36.
- **Quelle umbenennen.** Öffnet ein Dialogfeld, in dem Sie den Namen der hervorgehobenen Datenquelle ändern können.

  Anmerkung: Dadurch wird die Datenquelle im Repository umbenannt. Sie müssen diesen neuen Datenquellennamen in allen vorhandenen Export- oder Streamingknoten, von denen er referenziert wird, erneut auswählen.

Gesamtes Repository löschen. Löscht die aktuellen Repository-Inhalte sowie die Konfigurationsdetails vollständig. Weitere Informationen finden Sie im Thema „Löschen eines Entitätsrepositorys" auf Seite 37.

Repository konfigurieren. Zeigt ein Fenster an, in dem Sie das aktuelle Repository konfigurieren können. Weitere Informationen finden Sie im Thema „Konfigurieren eines Entitätsrepositorys" auf Seite 19.

Erstellen einer Zuordnung zwischen Eingabefeldern und Merkmalen (EA-Exportknoten)

Das Repository bietet standardmäßig eine Reihe vordefinierter Merkmale. Verschiedene Datenquellen können für Informationstypen, die sich auf dasselbe Merkmal beziehen, unterschiedliche Feldnamen verwenden (z. B. **Adresse1** oder **Adresszeile 1**). Um Dopplungen zu vermeiden, müssen die Felder der Eingabedatenquellen bestimmten Repository-Merkmalen zugeordnet werden. Sie brauchen nicht jedes Feld im Dataset zuzuordnen, nur diejenigen, die vermutlich demselben Merkmal in anderen Datasets entsprechen.

Wenn eine Datenquelle Felder verwendet, die anderen Datentypen entsprechen, die nicht im Repository vordefiniert sind, können Sie über das Fenster "Repository-Konfiguration" neue Merkmale erstellen. Weitere Informationen finden Sie im Thema „Konfigurieren eines Entitätsrepositorys" auf Seite 19.

Erstellen einer Zuordnung zwischen Eingabefeldern und Merkmalen

1. Verbinden Sie einen EA-Exportknoten mit einem Datenquellenknoten im Streamerstellungsbereich. Jeder verwendete Datenquellenknoten muss mit seinem eigenen EA-Exportknoten verbunden sein.
2. Öffnen Sie den EA-Exportknoten, um die Registerkarte "Eingaben" anzuzeigen, die Optionen für die Zuordnung der Eingabefelder enthält. Weitere Informationen finden Sie im Thema „Repository-Eingabeoptionen für die Zuordnung".
3. Wählen Sie im EA-Exportknoten die Registerkarte "Repository" aus, um entweder die Zuordnungszuweisungen für die aktuelle Datenquelle oder für alle Datenquellen anzuzeigen, wenn Sie mehrere verwenden.
4. Klicken Sie auf **Zuordnung exportieren**, um eine Gruppe von Zuordnungszuweisungen zu speichern (z. B. zur Verwendung mit einem Exportknoten für eine andere Datenquelle).

Wenn Sie mit der Zuordnung des ersten Datenquellenknotens fertig sind, wiederholen Sie den Vorgang für alle anderen Datenquellenknoten, die verwendet werden sollen.

Repository-Eingabeoptionen für die Zuordnung

Die Registerkarte "Eingaben" enthält die Optionen zur Zuordnung von Datenquellenfeldern zu Repository-Merkmalen, die für den Export in das Repository bereit sind. Richten Sie die Zuordnungszuweisungen auf dieser Registerkarte ein. Klicken Sie optional auf die Registerkarte "Repository", um die Zuordnung für andere Datenquellen anzuzeigen, und klicken Sie dann auf **Ausführen**, um die Daten in das Repository zu exportieren.

Wenn Sie bereits eine Gruppe von Zuordnungen in einer XML-Datei gespeichert haben, können Sie diese durch Klicken auf **Zuordnung importieren** verwenden.

Modus. Behalten Sie die Standardauswahl **Zu Repository hinzufügen** bei, wenn Sie die Datensätze der Quellendatei zu den bestehenden Inhalten des Repositorys hinzufügen möchten. Wenn Sie die Repository-Inhalte löschen, jedoch die Konfigurationsdateien beibehalten möchten, bevor Sie die Quellendatensätze hinzufügen, wählen Sie die Datei **Repository-Inhalt vor Export löschen** aus.

Entitätsrepository. Zeigt das aktuelle Entitätsrepository an, sofern eines vorhanden ist. Wenn mehrere Repositorys vorhanden sind, können Sie in der Liste ein anderes Repository auswählen. Wählen Sie zur Erstellung eines neuen Repositorys die Option **<Durchsuchen...>** aus, um ein Dialogfeld anzuzeigen, in dem Sie das Repository erstellen können. Weitere Informationen finden Sie im Thema „Optionen für das Entitätsrepository" auf Seite 16.

Zu Entitätstyp zuordnen. Eine Liste der im Repository definierten Entitätstypen (d. h. Merkmalsets). Wählen Sie einen aus der Liste aus oder zeigen Sie mithilfe der Option **<Neuen Entitätstyp hinzufü-**

gen...> das Fenster "Repository-Konfiguration" an, in dem Sie einen neuen Entitätstyp definieren können. Weitere Informationen finden Sie im Thema „Konfigurieren eines Entitätsrepositorys" auf Seite 19.

Quellentag. Eine Liste von Tags, die Datenquellen angeben, die dem Repository derzeit bekannt sind. Wählen Sie einen Tag aus der Liste aus oder erstellen Sie mit **<Neuen Quellentag hinzufügen...>** einen Tag für eine neue Datenquelle.

Eindeutiger Schlüssel. (Erforderlich) Das für die eindeutigen IDs für die Datensätze zu verwendende Eingabefeld.

Zuordnungstabelle. In dieser Tabelle können Sie jedes Eingabefeld einem entsprechenden Merkmal im Repository zuordnen. Wenn im ausgewählten Entitätstyp kein geeignetes Merkmal vorhanden ist, können Sie hier ein neues Merkmal erstellen.

- **Feld.** Die Menge an Eingabefeldern in der ausgewählten Datenquelle. Zu jedem Feld gehört ein Symbol, das das Messniveau (also den Datentyp) für das Feld angibt.
- **Zugeordnet zu Merkmal.** Um ein Feld einem Merkmal zuzuordnen, doppelklicken Sie auf diese Spalte (oder drücken Sie die Leertaste) in der Zeile des betreffenden Felds und wählen Sie ein Merkmal aus der Liste aus. Wenn kein geeignetes Merkmal verfügbar ist, können Sie mit der Option **<Neues Merkmal hinzufügen...>** das Fenster "Repository-Konfiguration" anzeigen, in dem Sie ein neues Merkmal für diesen Entitätstyp definieren können. Weitere Informationen finden Sie im Thema „Konfigurieren eines Entitätsrepositorys" auf Seite 19.
- **Verwendung.** Gibt den Kontext eines bestimmten Felds an, wenn mehrere Kontexte möglich sind, beispielsweise die private und die dienstliche Telefonnummer. Für die Merkmale "Adresse" und "Telefon" sind voreingestellte Verwendungstypen verfügbar. Des Weiteren können Sie für alle Merkmale Ihre eigenen Verwendungstypen erstellen. Zur Festlegung eines anderen Verwendungstyps als dem standardmäßig eingestellten Typ **(Auto)** klicken Sie in dieser Spalte auf die gewünschte Zeile und wählen Sie entweder einen der bestehenden Verwendungstypen aus (falls vorhanden) oder klicken Sie auf die Option **Verwendung hinzufügen...**, um einen neuen Verwendungstyp zu erstellen. Weitere Informationen finden Sie im Thema „Verwalten der Elementtypen" auf Seite 23.

Zuordnung importieren. Importiert eine zuvor exportierte Menge an Zuordnungen zwischen Feldern und Merkmalen aus einer externen XML-Datei. Diese kann nützlich sein, wenn Sie verschiedene Datenquellen mit denselben Zuordnungsanforderungen verwenden, da Sie dadurch nicht dieselben Zuordnungen für die verschiedenen Quellen erneut zu definieren brauchen.

Zuordnung exportieren. Exportiert die Menge der in der Zuordnungstabelle angegebenen Zuordnungen zwischen Feldern und Merkmalen in eine externe XML-Datei.

Anzeigen der Feldzuordnungen (EA-Exportknoten)

Klicken Sie auf der Registerkarte "Repository" auf die Schaltfläche **Aktualisieren**, um anzuzeigen, welchen Repository-Merkmalen Eingabefelder zugeordnet sind. Sie können dies entweder für die aktuelle Datenquelle (die Datenquelle, die durch den mit diesem Exportknoten verknüpften Quellenknoten gesteuert wird) oder für alle Datenquellen anzeigen.

Eingaben anzeigen für. Wählen Sie eine Option aus, um die Zuordnungen entweder für die aktuelle Datenquelle anzuzeigen oder für alle Datenquellen, die dem Repository bekannt sind.

Aktualisieren. Aktualisiert die Anzeige für die ausgewählte Eingabeoption.

Merkmale. Eine Liste aller Merkmale, die über Zuordnungen in den angezeigten Datenquellen verfügen. Merkmale ohne Zuordnung werden nicht angezeigt.

<Datenquelle>. In den einzelnen Spalten sind zugeordnete Felder in einer bestimmten Datenquelle für jedes Merkmal aufgeführt, für das eine Zuordnung definiert wurde.

Konfigurieren eines Entitätsrepositorys

Sie können die Repository-Inhalte über das Fenster "Repository-Konfiguration" verwalten. Dieses bietet eine benutzerfreundliche visuelle Benutzerschnittstelle für das gesamte Repository.

Wenn Sie vorhaben, mehrere Repositorys mit denselben oder ähnlichen Konfigurationen zu verwenden, können Sie eine Grundkonfiguration einrichten und in eine Datei exportieren, die Sie später in andere Repositorys importieren können. Weitere Informationen finden Sie im Thema „Wiederverwenden von Repository-Konfigurationen" auf Seite 26.

Anmerkung: Ab Version 16 unterstützt SPSS Entity Analytics Repositorys in IBM DB2. Ein Repository ist spezifisch für eine Version von SPSS Modeler und kann nicht aus einer früheren Version importiert werden, Wenn Sie ein Repository haben und ein Upgrade auf SPSS Entity Analytics Version 16 durchführen, müssen Sie das Repository daher in der neuen DB2-Datenbank neu erstellen.
Vorsicht:
Wenn Sie die Konfiguration eines Repositorys ändern und speichern, das bereits Daten enthält, werden Sie möglicherweise aufgefordert, die Repository-Inhalte zu löschen und die Daten erneut zu laden. Dadurch wird ein inkonsistenter Zustand des Repositorys vermieden.

Einrichten einer Repository-Konfiguration

1. Öffnen Sie einen beliebigen EA-Knoten.
2. Klicken Sie auf die Liste **Entitätsrepository**.
3. Klicken Sie auf **<Durchsuchen...>**, um das Dialogfeld "Entitätsauflösungsinstanzen" anzuzeigen.
4. Klicken Sie im Dialogfeld "Entitätsauflösungsinstanzen" auf die Liste **Repository-Name**.
5. Wählen Sie das Repository aus, für das Sie die Konfiguration einrichten möchten.
6. Wenn noch keine Verbindung besteht, geben Sie den Benutzernamen für den Administrator und das zugehörige Kennwort ein und klicken Sie auf **Verbinden**.
7. Wenn die Schaltfläche **Repository konfigurieren** aktiviert ist, klicken Sie darauf, um das Fenster "Repository-Konfiguration" anzuzeigen.
8. Erstellen Sie die Konfigurationsdetails, wie in den folgenden Abschnitten beschrieben.

Der Navigationsbereich auf der linken Seite des Fensters "Repository-Konfiguration" enthält eine Baumstruktur, über die Sie die verschiedenen Eigenschaften des Repositorys verwalten können.

Tabelle 9. Hauptelemente des Fensters "Repository-Konfiguration".

Abschnitt	**Beschreibung**	
Datenquellen	Zeigt die Zuordnungen aller Datenquellen zu den verschiedenen Repository-Merkmalen an.	Weitere Informationen finden Sie im Thema „Anzeigen der Datenquellenzuordnungen" auf Seite 20.
Merkmale	Dient zum Erstellen eines neuen Merkmals bzw. zum Duplizieren, Bearbeiten oder Löschen eines bestehenden Merkmals.	Weitere Informationen finden Sie im Thema „Verwalten der Repository-Merkmale" auf Seite 20.
Entitätstypen	Dient zum Erstellen eines neuen Merkmals bzw. zum Verwalten bestehender Merkmale (Duplizieren, Umbenennen, Anfügen oder Entfernen von Merkmalen, Löschen).	Weitere Informationen finden Sie im Thema „Verwalten der Elementtypen" auf Seite 23.
Auflösungsregeln	Dient zur Festlegung des Schwellenwerts für den Entitätsabgleich.	Weitere Informationen finden Sie im Thema „Festlegen des Schwellenwerts für den Entitätsabgleich" auf Seite 25.

Anzeigen der Datenquellenzuordnungen

Im Abschnitt "Datenquellen" des Fensters "Repository-Konfiguration" bietet der Eintrag "Alle Quellen" eine schreibgeschützte Anzeige der Zuordnungen aller Datenquellen zu den verschiedenen Repository-Merkmalen.

Klicken Sie auf **Aktualisieren**, um die Liste zu aktualisieren, wenn neue Datenquellen in das Repository aufgenommen wurden.

Hinweis: Hier kann keine Datenquelle zum Repository hinzugefügt werden. Die einzige Möglichkeit zum Hinzufügen von Datenquellen besteht darin, einen SPSS Modeler-Quellenknoten zu erstellen und ihn mit einem EA-Exportknoten zu verbinden. Weitere Informationen finden Sie im Thema „Verbinden mit einer Datenquelle" auf Seite 14.

Verwalten der Repository-Merkmale

Ein Repository-Merkmal ist ein einzelner Informationstyp, der zusammen mit einer Entitätsdatenquelle verwendet werden kann. Einige Merkmale (z. B. Vorname, Nachname, Geburtsdatum usw.) können mit vielen verschiedenen Datenquellen verwendet werden, während andere Merkmale nur bei einer bestimmten Datenquelle vorkommen. Ein Merkmal kann ein oder mehrere Elemente enthalten; jedes Element entspricht üblicherweise einem Feld in einem Datensatz oder einer Spalte in einer Datenbanktabelle.

Im Abschnitt "Merkmale" des Fensters "Repository-Konfiguration" bietet der Eintrag "Alle Merkmale" die Möglichkeit zur Verwaltung aller Repository-Merkmale. Sie haben folgende Möglichkeiten:

- Neues Merkmal erstellen
- Bestehendes Merkmal duplizieren (beispielsweise, um ein neues Merkmal auf der Grundlage eines bestehenden Merkmals zu erstellen)
- Bestehendes Merkmal bearbeiten
- Bestehendes Merkmal löschen

Anweisungen für diese Aufgaben erhalten Sie später in diesem Abschnitt.

Die Merkmalsliste zeigt alle Merkmale, die in diesem Repository definiert wurden. In den Spalten in der Liste werden die verschiedenen Eigenschaften angezeigt, die ein Merkmal aufweisen kann.

Merkmal. Der Name des Merkmals. Ein Schlosssymbol neben dem Namen eines Merkmals zeigt an, dass das Merkmal gesperrt ist. Gesperrte Merkmale können nicht gelöscht oder dupliziert werden. Die einzige Änderung an gesperrten Merkmalen, die gespeichert werden kann, ist die Änderung des Anonymisierungsattributs.

Häufigkeit. Gibt an, wie viele Entitäten denselben Wert für dieses Merkmal aufweisen können. Gültige Werte sind **Eins** (z. B. für eine Passnummer), **Wenige** (z. B. für eine Adresse) oder **Viele** (z. B. für ein Geburtsdatum).

Exklusivität. Gibt an, dass eine Entität diesen Merkmalstyp typischerweise nur einmal aufweisen sollte. So hätten hier z. B. die Merkmale "Geburtsdatum" oder "Steuernummer" den Wert **Ja**, während die Merkmale "Adresse" oder "Kreditkartennummer" den Wert **Nein** hätten, da eine Entität mehrere Adressen oder Kreditkartennummern haben kann.

Stabilität. Gibt den Stabilitätswert des betreffenden Merkmals an (d. h. ob es *unwahrscheinlich* ist, dass es sich während der Lebensdauer einer Entität ändert). Das Merkmal "Geburtsdatum" beispielsweise hätte den Wert **Ja**, da es sich niemals ändert, das Merkmal "Adresse" dagegen hätte den Wert **Nein**, da hierfür die Wahrscheinlichkeit einer Änderung recht hoch und das Merkmal somit weniger stabil ist. *Hinweis*: Das Merkmal "Geschlecht" ist in der Regel lebenslang stabil. Da es aufgrund von fehlerhaften Daten aber häufig falsch angegeben wird, erhält das Merkmal bei der Standardkonfiguration den Wert **Nein**.

Anonymisieren. Gibt an, ob das Merkmal anonymisiert wurde. Die Eingaben lauten entweder **Ja** oder **Nein**. Weitere Informationen finden Sie im Thema „Anonymisieren der Repository-Merkmale" auf Seite 23.

Erstellen eines neuen Merkmals

1. Führen Sie eine der folgenden Aktionen aus.
   - Klicken Sie auf die Schaltfläche "Neues Merkmal erstellen" (die oberste Schaltfläche rechts im Bildschirm).
   - Klicken Sie mit der rechten Maustaste im Navigationsbereich links im Bildschirm auf **Alle Merkmale** und wählen Sie die Option **Neues Merkmal** aus.
2. Arbeiten Sie das Dialogfeld "Merkmal hinzufügen/bearbeiten" ab. Weitere Informationen finden Sie im Thema „Hinzufügen bzw. Bearbeiten von Merkmalen" auf Seite 22.

Duplizieren eines bestehenden Merkmals

1. Wählen Sie in der Spalte **Merkmal** auf der rechten Bildschirmseite das zu duplizierende Merkmal aus.
2. Klicken Sie auf die Schaltfläche "Ausgewähltes Merkmal duplizieren" (die zweite Schaltfläche rechts im Bildschirm).
3. Arbeiten Sie das Dialogfeld "Merkmal hinzufügen/bearbeiten" ab. Weitere Informationen finden Sie im Thema „Hinzufügen bzw. Bearbeiten von Merkmalen" auf Seite 22.

Bearbeiten eines bestehenden Merkmals

VORSICHT: Wenn Sie ein Merkmal oder ein Element eines Merkmals bearbeiten, löschen oder anonymisieren und das Repository bereits Daten enthält, sollten Sie den Repository-Inhalt anschließend löschen und die Daten neu laden. Dadurch wird ein inkonsistenter Zustand des Repositorys vermieden.

1. Wählen Sie in der Spalte **Merkmal** auf der rechten Bildschirmseite das zu bearbeitende Merkmal aus. *Hinweis*: Sie können nur Merkmale auswählen, die Sie selbst erstellt haben, keine vom System bereitgestellten Merkmale.
2. Klicken Sie auf die Schaltfläche "Ausgewähltes Merkmal bearbeiten" (die dritte Schaltfläche rechts im Bildschirm).
3. Arbeiten Sie das Dialogfeld "Merkmal hinzufügen/bearbeiten" ab. Weitere Informationen finden Sie im Thema „Hinzufügen bzw. Bearbeiten von Merkmalen" auf Seite 22.

Löschen eines bestehenden Merkmals

VORSICHT: Wenn Sie ein Merkmal oder ein Element eines Merkmals bearbeiten, löschen oder anonymisieren und das Repository bereits Daten enthält, sollten Sie den Repository-Inhalt anschließend löschen und die Daten neu laden. Dadurch wird ein inkonsistenter Zustand des Repositorys vermieden.

1. Wählen Sie in der Spalte **Merkmal** auf der rechten Bildschirmseite das zu löschende Merkmal aus. *Hinweis*: Sie können nur Merkmale löschen, die Sie selbst erstellt haben, keine vom System bereitgestellten Merkmale.
2. Führen Sie eine der folgenden Aktionen aus.
   - Klicken Sie auf die Schaltfläche "Ausgewähltes Merkmal löschen" (die unterste Schaltfläche rechts im Bildschirm).
   - Klicken Sie mit der rechten Maustaste im Navigationsbereich links im Bildschirm auf **Alle Merkmale** und wählen Sie die Option **Löschen** aus.
3. Klicken Sie auf **Weiter**, um das Löschen des Merkmals zu bestätigen.

   Vorsicht:
   Das Löschen von Merkmalen kann nicht rückgängig gemacht werden.

Hinzufügen bzw. Bearbeiten von Merkmalen

VORSICHT: Wenn Sie ein Merkmal oder ein Element eines Merkmals bearbeiten, löschen oder anonymisieren und das Repository bereits Daten enthält, sollten Sie den Repository-Inhalt anschließend löschen und die Daten neu laden. Dadurch wird ein inkonsistenter Zustand des Repositorys vermieden.

Im Dialogfeld "Merkmal hinzufügen/bearbeiten" können Sie ein neues Repository-Merkmal erstellen oder ein bestehendes Merkmal duplizieren bzw. bearbeiten.

Hinweis: Wenn ein vorhandenes Merkmal gesperrt ist, können Sie seine Details in diesem Dialogfeld nicht bearbeiten.

Merkmaltyp. Eine Beschriftung, die den Typ der Informationen angibt, auf die sich das Merkmal bezieht. Diese Beschriftung bildet den ersten Teil der Merkmal-ID.

Beschreibung. Eine kurze Textbeschreibung des Merkmalstyps, nur zu Informationszwecken.

Häufigkeit. Gibt an, wie viele Entitäten denselben Wert für dieses Merkmal aufweisen können. Gültige Werte sind **Eins** (z. B. für eine Passnummer), **Wenige** (z. B. für eine Adresse) oder **Viele** (z. B. für ein Geburtsdatum).

Exklusivität. Gibt an, dass eine Entität diesen Merkmalstyp typischerweise nur einmal aufweisen sollte. So hätten hier z. B. die Merkmale "Geburtsdatum" oder "Steuernummer" den Wert **Ja**, während die Merkmale "Adresse" oder "Kreditkartennummer" den Wert **Nein** hätten, da eine Entität mehrere Adressen oder Kreditkartennummern haben kann.

Stabilität. Gibt den Stabilitätswert des betreffenden Merkmals an (d. h. ob es *unwahrscheinlich* ist, dass es sich während der Lebensdauer einer Entität ändert). Das Merkmal "Geburtsdatum" beispielsweise hätte den Wert **Ja**, da es sich niemals ändert, das Merkmal "Adresse" dagegen hätte den Wert **Nein**, da hierfür die Wahrscheinlichkeit einer Änderung recht hoch und das Merkmal somit weniger stabil ist. *Hinweis*: Das Merkmal "Geschlecht" ist in der Regel lebenslang stabil. Da es aufgrund von fehlerhaften Daten aber häufig falsch angegeben wird, erhält das Merkmal bei der Standardkonfiguration den Wert **Nein**.

Elementetabelle. Eine Liste der Elemente, aus denen dieses Merkmal besteht.

- **Element.** Der Name des Elements.
- **Beschreibung.** Eine kurze Beschreibung dessen, was das Element angibt.
- **Datentyp.** Typ der Daten, die für dieses Element verwendet werden können. Die folgenden Typen sind verfügbar: Zeichenfolge, Ganze Zahl, Reelle Zahl und Datum.

Schaltfläche "Neues Element hinzufügen". Fügt eine neue Zeile zur Elementtabelle hinzu, sodass Sie ein neues Element definieren können.

Schaltfläche "Element löschen". Löscht eine ausgewählte Zeile aus der Elementetabelle. Dieser Vorgang kann nicht rückgängig gemacht werden.

VORSICHT: Wenn Sie ein Merkmal oder ein Element eines Merkmals bearbeiten, löschen oder anonymisieren und das Repository bereits Daten enthält, sollten Sie den Repository-Inhalt anschließend löschen und die Daten neu laden. Dadurch wird ein inkonsistenter Zustand des Repositorys vermieden.

Anonymisieren. Zu Datenschutzzwecken können Sie auswählen, dass Ihre Daten anonymisiert werden, wenn sie einem Repository hinzugefügt werden. Wenn Sie diese Option für ein Merkmal aktivieren wollen, wählen Sie **Ja** aus. Weitere Informationen finden Sie im Thema „Anonymisieren der Repository-Merkmale” auf Seite 23.

Anonymisieren der Repository-Merkmale

Als Teil der Datensicherheit möchten Sie eventuell die Daten anonymisieren, wenn sie dem Repository hinzugefügt werden, um das Risiko zu verringern, dass persönliche Daten versehentlich offengelegt werden.

Wenn anonymisierte Daten in ein Repository exportiert werden, ist eine Anonymisierungsmethode erforderlich, die weiterhin eine Entitätsauflösung mit den anonymisierten Daten ermöglicht. Wenn z. B. zwei Datensätze für die Kreditkartendetails einer Person mit "anon_s21" und "anon_s9271" anonymisiert werden, verlieren sie ihre Beziehung. Wenn Sie jedoch eine interne Hintergrundverknüpfung zwischen den Datensätzen verwenden, kann das System weiterhin verstehen, dass ein Name die Kurzform des anderen Namens ist.

Die Hintergrundverknüpfungen und IDs, die eine Verknüpfung Ihrer anonymisierten Daten ermöglichen, werden generiert, wenn Sie ein Repository erstellen. Sie sind für das Repository eindeutig. Die verschlüsselten Daten werden intern gespeichert und dann gelesen, wenn ein Stream eine Verbindung zu einem Repository herstellt.

Wenn Sie Ihr Repository konfigurieren, können Sie angeben, ob die einzelnen Merkmale anonymisiert werden oder nicht. Wenn ein Merkmal anonymisiert wird, werden alle seine Element anonymisiert; und das Merkmal wird unabhängig von seinem Verwendungstyp immer anonymisiert. Weitere Informationen finden Sie im Thema „Hinzufügen bzw. Bearbeiten von Merkmalen” auf Seite 22.

Anmerkung: Stellen Sie sicher, dass Sie nicht alle Felder für SPSS Entity Analytics anonymisieren; andernfalls können Sie nicht erkennen, welche Daten zurückkommen. Es wird empfohlen, mindestens ein Feld (auch wenn es nur eine Zeilennummer ist) ohne Anonymität zu belassen, sodass Sie die spätere erneute Zusammenführung mit Ihren ursprünglichen Daten steuern können.

Eine Spalte in der Merkmalliste im Fenster "Repository-Konfiguration" zeigt an, welche Merkmale für eine Anonymisierung festgelegt wurden. Die Eingaben lauten entweder **Ja** oder **Nein**.

Anmerkung: Wenn ein vorhandenes Repository Daten enthält, bevor Merkmale anonymisiert werden, müssen Sie zunächst alle Daten löschen. Andernfalls erfolgt kein Abgleich zwischen anonymisierten und nicht anonymisierten Merkmalen.

Verwalten der Elementtypen

Ein **Entitätstyp** ist eine benannte Menge an Repository-Merkmalen, die logisch zusammengehören. Beispielsweise könnte ein Entitätstyp, der für die Verwendung mit einem Kundendataset vorgesehen ist, aus Merkmalen wie "Name", "Geburtsdatum", "Geschlecht", "Adresse", "Telefonnummer" usw. bestehen.

Das IBM SPSS Modeler Entity Analytics-Repository wird mit einer Standardauswahl an Entitätstypen bereitgestellt und Sie können Ihre eigenen hinzufügen.

Im Abschnitt "Entitätstypen" des Fensters "Repository-Konfiguration" werden die verschiedenen Entitätstypen aufgeführt, die erstellt wurden. Sie haben folgende Möglichkeiten:

- Neuen Entitätstyp erstellen
- Bestehenden Entitätstyp duplizieren (beispielsweise, um einen neuen Entitätstyp auf der Grundlage eines bestehenden Entitätstyps zu erstellen)
- Merkmale zu einem Entitätstyp hinzufügen
- Merkmale aus einem Entitätstyp entfernen
- Entitätstyp umbenennen
- Entitätstyp löschen

Entitätstyp. Der Name des ausgewählten Entitätstyps.

Merkmal. Eine Liste gültiger Merkmale, aus denen dieser Entitätstyp besteht.

Verwendungstyp. (Optional) Gibt verschiedene Kontexte an, in denen dieses Merkmal verwendet werden könnte. Doppelklicken Sie auf diese Spalte, um einen Verwendungstyp hinzuzufügen bzw. zu bearbeiten. Trennen Sie dabei die einzelnen Verwendungstypen durch Komma und Leerzeichen voneinander. Die Werte, die Sie hier angeben, definieren die Werte, die im EA-Exportknoten bzw. im Knoten "Streaming von EA" angezeigt werden, wenn ein Benutzer auf der Registerkarte "Eingaben" auf die Spalte "Verwendung" für ein Merkmal klickt. Weitere Informationen finden Sie im Thema „Repository-Eingabeoptionen für die Zuordnung" auf Seite 17.

Allgemeine Informationen zu Verwendungstypen:

- Verwendungstypen sind beliebige Beschriftungen.
- Sie können aus nahezu jeder Texteingabe einen Verwendungstyp erstellen. Sie werden jedoch daran gehindert, Leerzeichen und ungültige Zeichen einzugeben.
- Ihre eingegebenen Daten werden während der Eingabe automatisch in Großbuchstaben geändert.
- Sie können über beliebig viele Verwendungstypen verfügen.
- Verwendungstypen müssen nicht aussagekräftig sein, es ist jedoch bei der späteren Zuordnung hilfreich, wenn Sie eine Namenskonvention verwenden, die für Sie und andere Benutzer sinnvoll ist.
- Während der Zuordnung wird eine Warnung angezeigt, wenn Sie einige Elemente einem Verwendungstyp und andere Elemente einem anderen Verwendungstyp zugeordnet haben.

Normalerweise wird ein Fehler angezeigt, wenn Sie versuchen, zwei Felder derselben Kombination aus Merkmal und Element zuzuordnen. Verwendungstypen sind eine Möglichkeit, mindestens zwei Felder derselben Kombination aus Merkmal und Element zuzuordnen und einen Abgleich zwischen diesen durchzuführen.

Wenn Sie z. B. die zwei separaten Merkmale *PRIVATADRESSE* und *GESCHÄFTSADRESSE* definiert haben, würde kein Abgleich zwischen diesen stattfinden. Wenn eine Entität eine *PRIVATADRESSE* aufweist, die mit der *GESCHÄFTSADRESSE* einer anderen Entität identisch ist, findet kein Abgleich statt, da es sich um verschiedene Merkmale handelt. Wenn Sie jedoch ein einziges Merkmal mit verschiedenen Verwendungstypen wiederverwenden, wird bei der Auflösung angenommen, dass *ADRESSE.GESCHÄFT* mit *ADRESSE.PRIVAT* identisch ist.

Sie können Verwendungstypen für verschiedene Merkmale wiederverwenden oder Sie können verschiedene Verwendungstypen verwenden, z. B. *PRIV* und *GESCH* für Telefonnummern und *PRIVAT* und *GESCHÄFT* für Adressen. Dies spielt keine Rolle, da wir nicht telefonnummernübergreifend mit Adressen abgleichen. Eine konsistente Verwendung wird Ihnen jedoch später die Identifikation und Gruppierung von Feldern erleichtern.

Wenn mehrere Entitätstypen in ein einziges Repository eingegeben werden, spielt es keine Rolle, um welche Verwendungstypen es sich handelt, vorausgesetzt, Sie verwenden dasselbe Merkmal. Wenn Sie z. B. *GESCH* und *PRIV* als Verwendungstypen von *ADRESSE* für den Entitätstyp *UNTERNEHMEN* definieren, wird dies trotzdem mit *GESCHÄFT* und *PRIVAT* als Verwendungstypen von *ADRESSE* für *PERSON* abgeglichen.

Erstellen eines neuen Entitätstyps

1. Klicken Sie mit der rechten Maustaste im Navigationsbereich links im Bildschirm auf **Entitätstypen**.
2. Wählen Sie die Option **Neuer Entitätstyp** aus.
3. Geben Sie einen eindeutigen Namen für den Entitätstyp ein und klicken Sie auf "OK".
4. Fügen Sie Merkmale zu dem Entitätstyp hinzu (siehe nächster Abschnitt).

Hinzufügen von Merkmalen zu einem Entitätstyp

1. Wählen Sie den Entitätstyp im Navigationsbereich links im Bildschirm aus.

2. Klicken Sie auf die Schaltfläche "Merkmal anfügen" (die obere Schaltfläche rechts im Bildschirm).
3. Wählen Sie mindestens ein Merkmal aus der Liste der verfügbaren Merkmale (zur Auswahl mehrerer Merkmale klicken Sie bei gedrückter Steuertaste auf die Merkmale) und klicken Sie auf "OK".

Entfernen von Merkmalen aus einem Entitätstyp

1. Wählen Sie den Entitätstyp im Navigationsbereich links im Bildschirm aus.
2. Wählen Sie mindestens ein Merkmal aus der Tabelle der angefügten Merkmale auf der rechten Bildschirmseite aus. Klicken Sie bei gedrückter Steuertaste auf die Merkmale, um mehrere Merkmale auszuwählen.
3. Klicken Sie auf die Schaltfläche "Merkmal lösen" (die untere Schaltfläche rechts im Bildschirm).

Duplizieren eines bestehenden Entitätstyps

1. Klicken Sie mit der rechten Maustaste im Navigationsbereich links im Bildschirm auf den zu duplizierenden Entitätstyp.
2. Wählen Sie die Option **Entitätstyp duplizieren** aus.
3. Geben Sie einen eindeutigen Namen für den neuen Entitätstyp ein und klicken Sie auf "OK".
4. Fügen Sie nach Bedarf Merkmale zum Entitätstyp hinzu bzw. entfernen Sie sie daraus (siehe Anweisungen weiter oben).

Umbenennen eines Entitätstyps

VORSICHT: Wenn Sie ein Merkmal oder ein Element eines Merkmals bearbeiten, löschen oder anonymisieren und das Repository bereits Daten enthält, sollten Sie den Repository-Inhalt anschließend löschen und die Daten neu laden. Dadurch wird ein inkonsistenter Zustand des Repositorys vermieden.

1. Klicken Sie mit der rechten Maustaste im Navigationsbereich links im Bildschirm auf den umzubenennenden Entitätstyp.
2. Wählen Sie die Option **Umbenennen** aus.
3. Geben Sie den neuen Namen für den Entitätstyp ein und klicken Sie auf "OK".

Löschen eines Entitätstyps

VORSICHT: Wenn Sie ein Merkmal oder ein Element eines Merkmals bearbeiten, löschen oder anonymisieren und das Repository bereits Daten enthält, sollten Sie den Repository-Inhalt anschließend löschen und die Daten neu laden. Dadurch wird ein inkonsistenter Zustand des Repositorys vermieden.

1. Klicken Sie mit der rechten Maustaste im Navigationsbereich links im Bildschirm auf den zu löschenden Entitätstyp.
2. Wählen Sie **Löschen** aus.
3. Klicken Sie auf **OK**, um das Löschen des Entitätstyps zu bestätigen.

   Vorsicht:
   Das Löschen von Entitätstypen kann nicht rückgängig gemacht werden.

Festlegen des Schwellenwerts für den Entitätsabgleich

Im Abschnitt "Auflösungsregeln" des Fensters "Repository-Konfiguration" wählen Sie den Schwellenwert aus, bei dem der Entitätsabgleich erfolgen soll.

Beim Erstellen des Repositorys ist der Abgleich auf den Standardschwellenwert voreingestellt.

Wählen Sie die Option **Festgelegt für aggressive Auflösung** aus, wenn Sie in Ihren Datensätzen nicht genug Treffer finden, um eine Entitätsauflösung durchzuführen.

Wählen Sie **Festgelegt für Standardauflösung**, um von einer der anderen Einstellungen zum Standardschwellenwert zurückzukehren.

Wählen Sie die Option **Festgelegt für konservative Auflösung** aus, wenn Sie zu viele Treffer finden.

Um ein Repository für Entitäten und Beziehungen zu erstellen, wählen Sie **Beziehungen einschließen** aus. Beachten Sie, dass diese Option nur verfügbar ist, wenn Sie über das separat lizenzierte Upgrade (IBM SPSS Modeler Entity Analytics Unleashed) verfügen.

Wiederverwenden von Repository-Konfigurationen

Wenn Sie bereits eine Konfiguration eingerichtet haben und sie für ein anderes Repository verwenden möchten, können Sie die bestehende Konfiguration in eine XML-Datei exportieren und die Datei in das andere Repository (das Ziel-Repository) importieren. Dies ist nur innerhalb einer vorhandenen Einrichtung möglich. Sie können z. B. keine Repository-Konfiguration von einer IBM SPSS Modeler-Version auf eine andere oder von einem Datenbanktyp auf einen anderen migrieren.

Wiederverwenden einer bestehende Konfiguration

1. Zeigen Sie das Fenster "Repository-Konfiguration" für das Repository an, dessen Konfiguration Sie verwenden möchten. Weitere Informationen finden Sie im Thema „Konfigurieren eines Entitätsrepositorys” auf Seite 19.
2. Wählen Sie im Menü in diesem Fenster Folgendes aus:

   Konfiguration > **Konfiguration exportieren**
3. Wählen Sie im Dialogfeld "Speichern unter" den Namen und den Speicherort der XML-Exportdatei aus.
4. Zeigen Sie das Fenster "Repository-Konfiguration" für das Ziel-Repository an.
5. Wählen Sie im Menü in diesem Fenster Folgendes aus:

   Konfiguration > **Konfiguration importieren**
6. Wählen Sie im Dialogfeld "Öffnen" den Namen und den Speicherort der zuvor exportierten XML-Exportdatei aus und klicken Sie auf **Öffnen**.

Speichern der Konfigurationsänderungen

Speichern der Änderungen an der Konfiguration

Wählen Sie im Menü im Fenster "Repository-Konfiguration" Folgendes aus:

Datei > **Speichern**

Schließen des Konfigurationsfensters

Verlassen des Konfigurationsfensters

Wählen Sie im Menü im Fenster "Repository-Konfiguration" Folgendes aus:

Datei > **Beenden**

Klicken Sie bei nicht gespeicherten Änderungen an der Konfiguration auf **OK**, um die Änderungen zu speichern und das Dialogfeld zu verlassen, bzw. auf **Abbrechen**, um das Dialogfeld ohne Speichern zu verlassen.

Analysieren der aufgelösten Identitäten (EA-Quellenknoten)

Nachdem die Daten in das Repository exportiert wurden, können Sie den EA-Quellenknoten verwenden, um die aufgelösten Identitäten zur weiteren Analyse bzw. Verarbeitung an andere IBM SPSS Modeler-Knoten weiterzuleiten, beispielsweise, um einen Bericht zu erstellen, in dem die aufgelösten Identitäten aufgeführt sind.

Analysieren der aufgelösten Identitäten

1. Fügen Sie einen EA-Quellenknoten zu einem Stream hinzu.
2. Öffnen Sie den EA-Quellenknoten.
3. Wählen Sie auf der Registerkarte "Daten" das Entitätsrepository und mindestens eine seiner Eingabedatenquellen aus (klicken Sie auf **Aktualisieren**, um die Datensatzanzahl zu aktualisieren). Weitere Informationen finden Sie im Thema „Auswählen einer Datenquelle".
4. Fügen Sie weitere Knoten zum Stream hinzu, um die gewünschte Verarbeitung durchzuführen. Weitere Informationen finden Sie im Thema „Hinzufügen von Knoten zum Stream" auf Seite 28.

Auswählen einer Datenquelle

Auf der Registerkarte "Daten" wählen Sie mindestens eine Datenquelle im Repository aus, für die Sie weitere Verarbeitungsschritte durchführen. Zur Aktualisierung der Datensatzanzahl für die aufgelisteten Datenquellen klicken Sie auf **Aktualisieren**.

Entitätsrepository. Zeigt das aktuelle Entitätsrepository an, sofern eines vorhanden ist. Wenn mehrere Repositorys vorhanden sind, können Sie in der Liste ein anderes Repository auswählen. Wählen Sie zur Erstellung eines neuen Repositorys die Option **<Durchsuchen...>** aus, um ein Dialogfeld anzuzeigen, in dem Sie das Repository erstellen können. Weitere Informationen finden Sie im Thema „Optionen für das Entitätsrepository" auf Seite 16.

Datensätze aus Datenquellen einschließen. In dieser Tabelle werden die verschiedenen Datenquellen aufgelistet, die in das Repository eingegeben wurden, zusammen mit der Anzahl der Datensätze in den einzelnen Datenquellen. Aktivieren Sie das Kontrollkästchen **Einschließen** für diejenigen Datenquellen, die Sie für die weitere Analyse und Verarbeitung verwenden möchten. Wenn Sie alle Datenquellen auswählen oder abwählen wollen, klicken Sie dementsprechend entweder auf **Alle einschließen** oder auf **Alle ausschließen**.

Beziehungen. Wählen Sie den Typ der Beziehung aus, die in das Repository aufgenommen werden soll. Beachten Sie, dass diese Option nur verfügbar ist, wenn Sie über das separat lizenzierte Upgrade (IBM SPSS Modeler Entity Analytics Unleashed) verfügen und das Repository für die Aufnahme von Beziehungen konfiguriert wurde.

- **Keine Beziehungen.** Beziehungsdetails werden nicht verwendet.
- **Enge Beziehungen.** Wählt nur eng miteinander verwandete Entitäten aus. Die Enge einer Beziehung hängt von vielen Variablen ab, wie den Eigenschaften der zugeordneten Funktionen, den gemeinsam genutzten Funktionen und der Angabe, ob eine konservative oder eine aggressive Auflösung verwendet wird.
- **Alle Beziehungen.** Wählt alle zugehörigen Entitäten aus.

Max. Abgrenzungsgrad. Nur verfügbar, wenn **Enge Beziehungen** oder **Alle Beziehungen** ausgewählt ist. Wählen Sie den Wert für den Abgrenzungsgrad aus, der zum Angeben einer Beziehung verwendet werden soll. Wenn Ann und Bob sich beispielsweise kennen und John sowohl Ann als auch Bob kennt, gehören Ann und Bob über 2 Abgrenzungsgrade zueinander.

Ausgabeentitätstyp. Wenn das Repository Details enthält, wird über diese Option standardmäßig der erste im Repository aufgelistete Entitätstyp angezeigt. Enthält das Repository mehr als einen Entitätstyp, werden die auf der Registerkarte "Filter" angezeigten Funktionen in die Funktionen dieses Entitätstyps geändert. Sie können einen beliebigen der im Repository verwendeten Entitätstypen auswählen.

Umbenennen von Datendateien

Sie können die Registerkarte "Filter" verwenden, um eines der aufgelösten Identitätsfelder umzubenennen, die zur weiteren Verarbeitung nach unten im Stream weitergegeben werden. Es kann sinnvoll sein, ein aufgelöstes Identitätsfeld umzubenennen, beispielsweise um die Kompatibilität der Feldnamen beim Zusammenführen mit einem anderen Dataset weiter unten im Stream zu gewährleisten.

Die Felder mit ihren ursprünglichen Namen lauten wie folgt.

Tabelle 10. Aufgelöste Identitätsfelder

Feld	Beschreibung
$EA-ID	Entitäts-ID
$EA-SRC	Quellentag zur Angabe der Datenquelle, aus der die Datensätze stammen
$EA-KEY	Als eindeutiger Schlüssel in der Datenquellendatei gekennzeichnetes Feld

Hinweis: Sie können zwar auch die Registerkarte "Filter" verwenden, um Felder herauszufiltern, sollten dies hier jedoch nicht tun, da die aufgelösten Identitätsfelder das absolute Minimum sind, das für die EA-Verarbeitung benötigt wird.

Festlegen der Typinformationen für Datenfelder

Auf der Registerkarte "Typen" können Sie verschiedene Eigenschaften der aufgelösten Identitätsfelder anzeigen bzw. ändern, die zur weiteren Verarbeitung nach unten im Stream weitergegeben werden.

Sie können dieselben Eigenschaften ändern wie auf der Registerkarte "Typen" eines regulären SPSS Modeler-Typknoten, nämlich:

Tabelle 11. Typeigenschaften für Felder

Eigenschaft	Beschreibung
Messung	Das Messniveau (d. h. der Datentyp), das zur Beschreibung der Eigenschaften der Daten im Feld verwendet wird.
Werte	Bietet Optionen zum Lesen von Datenwerten aus dem Dataset.
Fehlend	Wird verwendet, um anzugeben, wie fehlende Werte für das Feld behandelt werden.
Überprüfen	Validierungsoptionen, die sicherstellen, dass Feldwerte den angegebenen Werten bzw. Bereichen entsprechen.
Rolle	Gibt an, wie das Feld verwendet wird, wenn die Daten an Modellierungsknoten oder ein Modellnugget weitergegeben werden.

Hinzufügen von Knoten zum Stream

Sie können verschiedene SPSS Modeler-Knoten zum Stream hinzufügen, um Analyse- oder Verarbeitungsvorgänge an der Ausgabe aus dem EA-Quellenknoten vorzunehmen. Beispielsweise könnten Sie eines oder mehrere der folgenden Elemente hinzufügen.

- Aggregat- oder Duplikatknoten zur Zusammenfassung der Ausgabe, die sehr groß sein kann
- Auswahlknoten zur Auswahl eines Subsets der Ausgabe
- Tabellenknoten zur Anzeige der Ausgabe aus dem EA-Quellenknoten
- Berichtknoten zum Drucken der Ausgabe in einem Bericht
- SPSS Modeler-Exportknoten zum Exportieren der Ausgabe in ein anderes Format, beispielsweise eine Kalkulationstabelle oder eine Datenbank

Weitere Informationen finden Sie in den Abschnitten zu Datensatzoperationsknoten, Ausgabeknoten und Exportknoten im Handbuch *IBM SPSS Modeler 15 Quellen-, Prozess- und Ausgabeknoten*.

Vergleichen neuer Fälle mit dem Repository (Knoten "Streaming von EA")

Wenn Sie bereits eine Identitätsauflösung im Repository durchgeführt haben, können Sie den Knoten "Streaming von EA" verwenden, um neue Fälle, auf die Sie danach stoßen, mit dem Repository-Inhalt zu vergleichen. Dieser Knoten verarbeitet Datensätze aus einer neuen Datenquelle, vergleicht sie mit den bereits im Repository befindlichen Entitäten und leitet alle übereinstimmenden Datensätze für die weitere Verarbeitung weiter. Es kann festgelegt werden, dass die Übereinstimmungen exakt sein müssen, oder es kann auch eine lockerere Beziehung mit den bestehenden Entitäten zugelassen werden.

Ebenso wie der EA-Exportknoten verwendet auch der Knoten "Streaming von EA" einen einzelnen SPSS Modeler-Quellenknoten als Eingabe. Der Knoten "Streaming von EA" weist jedoch folgende Abweichungen auf: Während der Exportknoten Datensätze zu allen Entitäten ausgibt, die mit seinen Eingabedatensätzen in Beziehung stehen, gibt der Knoten "Streaming von EA" Datensätze nur zu den Entitäten aus, die in Beziehung mit Entitäten stehen, die bereits im Repository aufgelöst wurden. Weitere Informationen finden Sie im Thema „Ausgabe des Knoten "Streaming von EA"" auf Seite 32.

Vergleichen neuer Fälle mit dem Repository

1. Stellen Sie eine Verbindung zu der Datenquelle her, die die neuen Datensätze enthält, die mit den bestehenden Entitäten verglichen werden sollen. Weitere Informationen finden Sie im Thema „Verbinden mit einer Datenquelle" auf Seite 14.
2. Verbinden Sie auf der Registerkarte "Datensatzoperationen" einen Knoten vom Typ "Streaming von EA" mit dem Datenquellenknoten.
3. Doppelklicken Sie dann auf den EA-Exportknoten, um das zugehörige Dialogfeld zu öffnen.
4. Klicken Sie auf die Liste **Entitätsrepository**.
5. Klicken Sie auf **<Durchsuchen...>**, um das Dialogfeld "Entitätsrepositorys" anzuzeigen.
6. Klicken Sie im Dialogfeld "Entitätsrepositorys" auf das Feld "Repository-Name".
7. Klicken Sie auf den Namen des zu verwendenden Repositorys.
8. Geben Sie Benutzernamen und Kennwort für dieses Repository ein und klicken Sie auf **Verbinden**. Klicken Sie auf **OK**, wenn die Verbindung mit dem Repository hergestellt wurde.
9. Wählen Sie im Dialogfeld "Streaming von EA" den Entitätstyp aus, den Sie zuordnen wollen. Weitere Informationen finden Sie im Thema „Verwalten der Elementtypen" auf Seite 23.
10. Ordnen Sie die Eingabefelder in der Datenquelle Merkmalen im Repository zu. Weitere Informationen finden Sie im Thema „Erstellen einer Zuordnung zwischen Eingabefeldern und Merkmalen (Knoten "Streaming von EA")" auf Seite 30.
11. Optional können Sie die Datensätze im Repository in Echtzeit aktualisieren, während Sie für Ihre Daten ein Scoring durchführen. Weitere Informationen finden Sie im Thema „Erstellen einer Zuordnung zwischen Eingabefeldern und Merkmalen (Knoten "Streaming von EA")" auf Seite 30.
12. Klicken Sie auf die Registerkarte **Ausgaben**, um Details der verschiedenen Datenquellen anzuzeigen, die in das Repository eingegeben wurden, und die Auswahlkriterien für das Abrufen bestehender Entitäten festzulegen. Weitere Informationen finden Sie im Thema „Anzeigen der Feldzuordnungen und Datenquellen (Knoten "Streaming von EA")" auf Seite 31.
13. Klicken Sie auf die Registerkarte **Filter**, um Details zu den Eingabefeldern und Merkmalen anzuzeigen, die im Repository gespeichert sind. Alle Merkmale, die im Knoten nicht zugeordnet wurden, werden standardmäßig ausgefiltert; Sie können dies jedoch bei Bedarf ändern.
14. Klicken Sie auf **OK**, wenn der Knoten korrekt eingerichtet wurde.
15. Verbinden Sie einen Tabellenknoten mit dem Knoten "Streaming von EA" und führen Sie den Stream aus.

Im Ausgabefenster des Tabellenknotens werden alle abgerufenen Entitäten aufgeführt, die mit den neuen Datensätzen in der Datenquelle übereinstimmen. Die Ausgabefelder wurden mit dem Präfix **$EA-** versehen. Weitere Informationen finden Sie im Thema „Ausgabe des Knoten "Streaming von EA"" auf Seite 32.

Anmerkung: Möglicherweise tritt bei der Ausführung des Knotens "Streaming von EA" ein Fehler der Form **Im Serverdatenmodell wurde eine falsche Anzahl von Feldern entdeckt** auf. Dies kann passieren, wenn Sie die Repository-Konfiguration nach der Erstellung des Knotens "Streaming von EA" bearbeitet haben. Eine Bearbeitung der Konfiguration unter diesen Umständen kann zur Folge haben, dass sich Anzahl und Namen der Ausgabefelder des Knotens ändern. Sie können das Problem beheben, indem Sie den Knoten "Streaming von EA" öffnen und auf die Schaltfläche **Aktualisieren** klicken. Dadurch werden Anzahl und Namen der Ausgabefelder neu berechnet.

Erstellen einer Zuordnung zwischen Eingabefeldern und Merkmalen (Knoten "Streaming von EA")

Die Registerkarte **Eingaben** enthält die Optionen für das Zuordnen von Feldern in der Eingabe für diesen Knoten zu Repository-Merkmalen. Richten Sie die Zuordnungszuweisungen auf dieser Registerkarte ein oder wählen Sie die Registerkarte **Ansicht** aus, um Details aller Datenquellen im Repository anzuzeigen, und klicken Sie anschließend auf **OK**.

Wenn Sie bereits eine Gruppe von Zuordnungen in einer XML-Datei gespeichert haben, können Sie diese durch Klicken auf **Zuordnung importieren** verwenden.

Entitätsrepository. Zeigt das aktuelle Entitätsrepository an, sofern eines vorhanden ist. Wenn mehrere Repositorys vorhanden sind, können Sie in der Liste ein anderes Repository auswählen. Wählen Sie zur Erstellung eines neuen Repositorys die Option **<Durchsuchen...>** aus, um ein Dialogfeld anzuzeigen, in dem Sie das Repository erstellen können. Weitere Informationen finden Sie im Thema „Optionen für das Entitätsrepository" auf Seite 16.

Zu Entitätstyp zuordnen. Eine Liste der im Repository definierten Entitätstypen (d. h. Merkmalsets). Wählen Sie einen aus der Liste aus oder zeigen Sie mithilfe der Option **<Neuen Entitätstyp hinzufügen...>** das Fenster "Repository-Konfiguration" an, in dem Sie einen neuen Entitätstyp definieren können. Weitere Informationen finden Sie im Thema „Konfigurieren eines Entitätsrepositorys" auf Seite 19.

Persistente Suchen. Wenn Sie die Datensätze im Repository in Echtzeit aktualisieren wollen, während Sie für Ihre Daten ein Scoring durchführen, wählen Sie diese Option aus.

Quellentag. Nur bei Auswahl von **Persistente Suchen** verfügbar. Eine Liste von Tags, die Datenquellen angeben, die dem Repository derzeit bekannt sind. Wählen Sie einen Tag aus der Liste aus oder erstellen Sie mit **<Neuen Quellentag hinzufügen...>** einen Tag für eine neue Datenquelle.

Eindeutiger Schlüssel. Nur bei Auswahl von **Persistente Suchen** verfügbar. Das für die eindeutigen IDs für die Datensätze zu verwendende Eingabefeld.

Zuordnungstabelle. In dieser Tabelle können Sie jedes Eingabefeld einem entsprechenden Merkmal im Repository zuordnen. Wenn im ausgewählten Entitätstyp kein geeignetes Merkmal vorhanden ist, können Sie hier ein neues Merkmal erstellen.

- **Feld.** Die Menge an Eingabefeldern in der ausgewählten Datenquelle. Zu jedem Feld gehört ein Symbol, das das Messniveau (also den Datentyp) für das Feld angibt.
- **Zugeordnet zu Merkmal.** Um ein Feld einem Merkmal zuzuordnen, doppelklicken Sie auf diese Spalte (oder drücken Sie die Leertaste) in der Zeile des betreffenden Felds und wählen Sie ein Merkmal aus der Liste aus. Wenn kein geeignetes Merkmal verfügbar ist, können Sie mit der Option **<Neues Merkmal hinzufügen...>** das Fenster "Repository-Konfiguration" anzeigen, in dem Sie ein neues Merkmal für diesen Entitätstyp definieren können. Weitere Informationen finden Sie im Thema „Konfigurieren eines Entitätsrepositorys" auf Seite 19.

- **Verwendung.** Gibt den Kontext eines bestimmten Felds an, wenn mehrere Kontexte möglich sind, beispielsweise die private und die dienstliche Telefonnummer. Weitere Informationen finden Sie im Thema „Verwalten der Elementtypen” auf Seite 23.

Zuordnung importieren. Importiert eine zuvor exportierte Menge an Zuordnungen zwischen Feldern und Merkmalen aus einer externen XML-Datei. Diese kann nützlich sein, wenn Sie verschiedene Datenquellen mit denselben Zuordnungsanforderungen verwenden, da Sie dadurch nicht dieselben Zuordnungen für die verschiedenen Quellen erneut zu definieren brauchen.

Zuordnung exportieren. Exportiert die Menge der in der Zuordnungstabelle angegebenen Zuordnungen zwischen Feldern und Merkmalen in eine externe XML-Datei.

Anzeigen der Feldzuordnungen und Datenquellen (Knoten "Streaming von EA")

Auf der Registerkarte "Ausgabe" können Sie Details der verschiedenen Datenquellen anzeigen, die in das Repository eingegeben wurden. Dabei handelt es sich um Datenquellen, anhand derer die Eingaben für diesen Knoten verarbeitet werden, um nach übereinstimmenden Entitäten zu suchen und diese abzurufen. Klicken Sie zum Aktualisieren der Datensatzanzahl auf **Aktualisieren**.

Übereinstimmungen aus Datenquellen einschließen. In dieser Tabelle werden die verschiedenen Datenquellen aufgelistet, die im Repository verfügbar sind, zusammen mit der Anzahl der Datensätze in den einzelnen Datenquellen.

Übereinstimmungen. Mit den folgenden Optionen wird angegeben, wie eng die Informationen für Zuordnungen zwischen Feldern und Merkmalen, die Sie auf der Registerkarte "Eingaben" angeben, mit den in Frage kommenden Datensätzen (also den gesamten Repository-Inhalten) abgeglichen werden sollen. Je enger das Übereinstimmungskriterium, desto weniger Entitäten werden abgerufen.

Hinweis: Werden mehr als 20 Übereinstimmungen gefunden, werden nur die ersten 20 zurückgegeben.

- **Nur exakte Übereinstimmungen einschließen.** Dies ist das engste Übereinstimmungskriterium und führt zur Auswahl der geringsten Anzahl an Datensätzen. Verwenden Sie diese Option, wenn Sie nur die Entitäten abrufen möchten, die als exakte Übereinstimmung eingestuft werden.
- **Mögliche Übereinstimmungen einschließen.** Verwenden Sie diese Einstellung, wenn Sie sowohl übereinstimmende Entitäten als auch Entitäten mit den gleichen Kennungen einbeziehen möchten (Entitäten mit Merkmalen, für die eine Häufigkeit von `Eins` konfiguriert wurde, z. B. übereinstimmende Kreditkartennummern, Steuernummern usw.).
- **Alle Übereinstimmungen einschließen.** Verwenden Sie diese Option, wenn Sie die größtmögliche Anzahl an Entitäten im Repository anzeigen möchten, die gemeinsame Merkmale haben. Dies ist das unschärfste Übereinstimmungskriterium und führt zur Auswahl der größten Anzahl an Datensätzen. Bei dieser Option werden exakte Übereinstimmungen sowie Entitäten mit fast beliebigen gemeinsamen Merkmalen zurückgegeben (typischerweise Entitäten mit einer Häufigkeit von `Eins` oder `Wenige`). So würden z. B. Entitäten mit der gleichen Steuernummer und Entitäten mit ähnlichen Adressen einbezogen werden.

Beziehungen. Diese Option ist nur verfügbar, wenn das Repository für die Aufnahme von Beziehungen konfiguriert wurde. Um das Repository für die Aufnahme von Beziehungen zu konfigurieren, müssen Sie über das separat lizenzierte Upgrade (IBM SPSS Modeler Entity Analytics Unleashed) verfügen. Wählen Sie den Typ der Beziehung aus, die in die Ausgabe aufgenommen werden soll.

- **Keine Beziehungen.** Beziehungsdetails werden nicht verwendet.
- **Enge Beziehungen.** Wählt nur eng miteinander verwandete Entitäten aus. Die Enge einer Beziehung hängt von vielen Variablen ab, wie den Eigenschaften der zugeordneten Funktionen, den gemeinsam genutzten Funktionen und der Angabe, ob eine konservative oder eine aggressive Auflösung verwendet wird.
- **Alle Beziehungen.** Wählt alle zugehörigen Entitäten aus.

Max. Abgrenzungsgrad. Nur verfügbar, wenn **Enge Beziehungen** oder **Alle Beziehungen** ausgewählt ist. Wählen Sie den Wert für den Abgrenzungsgrad aus, der zum Angeben einer Beziehung verwendet werden soll. Wenn Ann und Bob sich beispielsweise kennen und John sowohl Ann als auch Bob kennt, gehören Ann und Bob über 2 Abgrenzungsgrade zueinander.

Ausgabeentitätstyp. Wenn das Repository Details enthält, wird über diese Option standardmäßig der erste im Repository aufgelistete Entitätstyp angezeigt. Enthält das Repository mehr als einen Entitätstyp, werden die auf der Registerkarte "Filter" angezeigten Funktionen in die Funktionen dieses Entitätstyps geändert. Sie können einen beliebigen der im Repository verwendeten Entitätstypen auswählen.

Ausgabe des Knoten "Streaming von EA"

Die Ausgabe des Knoten "Streaming von EA" besteht aus folgenden Feldern für jeden abgerufenen Datensatz.

Feld	Beschreibung
Feld1[, *Feld2*[, ... *FeldN*]]	Felder aus der Datenquelle, die die neuen Datensätze enthält.
$EA-ID	Entitäts-ID für diesen Datensatz im Repository.
$EA-SRC	Quellentag zur Angabe der Datenquelle, aus der dieser Datensatz stammt.
$EA-KEY	Wert des eindeutigen Schlüssels für diesen Datensatz in der Datenquellendatei.
$EA-SC	Feld, das die Genauigkeit der Übereinstimmung zwischen diesem Datensatz und einer beobachteten Entität im Repository angibt; Wert von 1,0 (schlechte Übereinstimmung) bis 10,0 (gute Übereinstimmung).
$EA-*Merkmal1*[, $EA-*Merkmal2*[, ... $EA-*MerkmalN*]]	Werte der zugeordneten Merkmale für diesen Datensatz im Repository.

Wenn Beziehungsfelder im Repository aktiviert sind und der Abgrenzungsgrad auf der Registerkarte "Ausgaben" mehr als Null beträgt, enthält die Ausgabe des Knotens "Streaming von EA" auch die folgenden Felder für jeden Datensatz, der abgerufen wird.

Feld	Beschreibung
$EA-DEGREE	Abgrenzungsgrad.
$EA-PARENT	ID des Datensatzes, von dem die Abgrenzung berechnet wird.
$EA-CHILD	ID des Datensatzes, für den die Abgrenzung berechnet wird.
$EA-RULE	

Verwenden von IBM SPSS Modeler Entity Analytics mit anderen IBM SPSS-Produkten

Es stehen Installationsprogramme zur Verfügung, die die Verwendung von IBM SPSS Modeler Entity Analytics in Verbindung mit folgenden Produkten ermöglichen:

- IBM SPSS Collaboration and Deployment Services
- IBM SPSS Modeler Batch für Windows
- IBM SPSS Modeler Solution Publisher

Sie müssen diese Installationsprogramme ausführen, bevor Sie die Funktionen von IBM SPSS Modeler Entity Analytics in Verbindung mit diesen Produkten verwenden können. Weitere Informationen finden Sie im Handbuch *IBM SPSS Modeler Premium - Installation*.

Nach der Installation müssen Sie mithilfe des Clients von IBM SPSS Collaboration and Deployment Services Deployment Manager eine Entity Analytics-Repository-Serverdefinition erstellen. Diese ist für die Verwendung eines IBM SPSS Modeler-Streams erforderlich, der einen Entity Analytics-Knoten in einem Job von IBM SPSS Collaboration and Deployment Services enthält (also für die Ausführung von Entity Analytics-Streams in IBM SPSS Collaboration and Deployment Services). Die Serverdefinition muss mit dem Repository-Namen im Stream übereinstimmen. Diese Definition gibt dem Stream an, wo sich das Repository befindet, und gibt ihm die erforderlichen Verbindungsinformationen an.

Verwaltungsaufgaben

Für Repositorys, die in Entity Analytics erstellt werden, wird unter Verwendung des Produkts IBM DB2 ein neuer Datenbankdienst erstellt. DB2 erfordert einige Verwaltungsaufgaben. Diese Aufgaben werden in der Regel vom Datenbankadministrator oder vom Systemadministrator ausgeführt. Es handelt sich dabei um folgende Aufgaben:

- Konfigurieren von Portzuweisungen
- Verwalten der Administratorberechtigungen für die Repository-Datenbank

Weitere Verwaltungsaufgaben, die möglicherweise durchgeführt werden müssen, gelten für alle Repositorys. Es handelt sich dabei um folgende Aufgaben:

- Verschieben des Repositorys in ein anderes Speicherverzeichnis
- Festlegen von Streameigenschaften für Datums-/Zeit- und Zeitmarkenfelder
- Anpassen der Einstellungen für die Zeitlimitüberschreitung
- Ausführen von IBM SPSS Modeler Entity Analytics mit SPSS Modeler-Client und SPSS Modeler Server auf demselben Windows-System
- Löschen des Inhalts eines Entitätsrepositorys
- Löschen eines Entitätsrepositorys
- Löschen eines Repositorys, wenn keine Verbindung zu ihm hergestellt werden kann

Konfigurieren von Portzuordnungen

Jedem DB2-Datenbankservice muss ein Port zugeordnet sein, der keinen anderen Services, die auf dem Computer ausgeführt werden, zugeordnet sein darf. Die Datenbankservices befinden sich auf demselben Computer, auf dem auch IBM SPSS Modeler Server ausgeführt wird (oder, wenn IBM SPSS Modeler ohne Verbindung mit IBM SPSS Modeler Server verwendet wird, auf dem Computer, auf dem IBM SPSS Modeler ausgeführt wird.

Standardmäßig weist Entity Analytics Ports im Bereich 1320 bis 1520 zu, beginnend mit Port 1320 für das erste erstellte Repository. Im Falle eines Konflikts können Sie die Zuweisung von Ports durch Bearbeiten der Datei *<Modeler Server-Installationspfad>/ext/bin/pasw.entityanalytics/ea.cfg* und Festlegung geeigneter Werte für die Einstellungen *min_port* und *max_port* konfigurieren. Im Folgenden finden Sie die Standardinhalte dieser Datei:

```
# Konfiguration des Portbereichs für die Entitätsanalayse

#

#     Dieser Portbereich steuert, welche Ports DB2-Datenbanken

#     (zum Speichern von Repositorys für Entitätsanalysen erstellt)
```

```
#     verwenden dürfen.  Konfigurieren Sie diesen Bereich, wenn der Standardportbereich

#     in Ihrem System einen Konflikt verursacht.

#

# Standardwert für min_port = 1320

# Standardwert für max_port = 1520

min_port, 1320

max_port, 1520
```

Verwalten der Administratorberechtigungen für die Repository-Datenbank

Der Benutzername und das Kennwort für die DB2-Datenbank, die als Host für ein Entitätsrepository fungiert, werden bei der Erstellung des Repositorys definiert. Wenn Ihnen die aktuellen Berechtigungsnachweise bekannt sind, können Sie diese Angaben mithilfe des DB2-SQL-Editors ändern.

Starten des DB2-SQL-Editors

1. Öffnen Sie an einem Client-Computer ein Eingabeaufforderungsfenster.
2. Geben Sie Folgendes ein:

   `cd` *`Modeler-Installationsverzeichnis`*`\ext\bin\pasw.entityanalytics\DB2\bin`

   Dabei ist *Modeler-Installationsverzeichnis* das Verzeichnis, in dem SPSS Modeler installiert wurde.
3. Geben Sie Folgendes ein:

   `solsql -c "C:\Dokumente und Einstellungen\Alle Benutzer\Anwendungsdaten\IBM\SPSS\Modeler\`*`Version`*`\EA\ repositories\`*`Repository-Name`*

   Dabei ist *Version* die Versionsnummer der SPSS Modeler-Installation und *Repository-Name* ist der Name des Repositorys.
4. Geben Sie an der Eingabeaufforderung den aktuellen Benutzernamen des Datenbankadministrators und das zugehörige Kennwort ein, um die Eingabeaufforderung `solsql>` anzuzeigen.

Ändern des Kennworts für den Datenbankadministrator

1. Geben Sie an der Eingabeaufforderung `solsql>` Folgendes ein:

   ```
   alter user Benutzername identified by Kennwort;
   commit work;
   ```

   Dabei ist *Benutzername* der aktuelle Benutzername des Datenbankadministrators und *Kennwort* ist das neue Kennwort.
2. Geben Sie "exit" ein, um den Editor zu schließen.
3. Starten Sie den SPSS Modeler-Client neu.

Informationen zu anderen Verwaltungsaufgaben, die mit der DB2-Datenbank ausgeführt werden können, finden Sie in der Dokumentation zu der entsprechenden Version von IBM DB2 unter http://publib.boulder.ibm.com/.

Verschieben des Repositorys in ein anderes Speicherverzeichnis

Standardmäßig gelten für die Repository-Dateien im Verzeichnis *EA* folgende Speicherorte:

- C:\Dokumente und Einstellungen\Alle Benutzer\Anwendungsdaten\IBM\SPSS\Modeler*Version*\EA (Windows-Systeme)
- *Modeler-Installationsverzeichnis*/ext/bin/pasw.entityanalytics/EA (UNIX-Systeme)

Da die zum Speichern des Repositorys verwendeten Dateien sehr groß werden können, müssen Sie sie möglicherweise auf einen anderen Datenträger bzw. in eine andere Partition verlagern, um mehr Speicherplatz verfügbar zu machen.

Verschieben des Repositorys in ein anderes Verzeichnis

1. Beenden Sie SPSS Modeler.
2. Verschieben Sie das Verzeichnis *EA* vom ursprünglichen Speicherort (siehe oben) an einen neuen Speicherort. Unter Windows könnten Sie es beispielsweise in ein Verzeichnis wie *F:\Daten\EA* verschieben.
3. Bearbeiten Sie die Datei *<Modeler-Server-Installationspfad>/ext/bin/pasw.entityanalytics/ea.cfg*, indem Sie folgende Option hinzufügen:

   `repository_data_directory, neuer_Speicherort`

   Dabei ist *neuer_Speicherort* das Verzeichnis, in das Sie das EA-Verzeichnis verschoben haben, z. B. *F:\data\EA*.

Festlegen von Streameigenschaften für Datums-/Zeit- und Zeitmarkenfelder

Wenn die Quellendaten Felder mit Datums-/Zeit- bzw. Zeitmarkendaten enthalten, müssen Sie sicherstellen, dass die entsprechenden Streameigenschaften auf das von IBM SPSS Modeler Entity Analytics erkannte Format gesetzt sind.

Festlegen des Formats für die Streameigenschaften

1. Klicken Sie im SPSS Modeler-Hauptmenü auf:

   Extras > **Streameigenschaften** > **Optionen**
2. Wählen Sie **Datum/Uhrzeit** aus.
3. Setzen Sie **Datumsformat** auf **JJJJ-MM-TT**.
4. Setzen Sie **Zeitformat** auf **HH:MM:SS**.
5. Klicken Sie auf **OK**.

Anpassen der Einstellungen für die Zeitlimitüberschreitung

Wenn in langsamen oder stark ausgelasteten Systemen Fehler auftreten, wenn Sie Repositorys erstellen oder auf Repositorys zugreifen, müssen Sie möglicherweise die Einstellungen für die Zeitlimitüberschreitung für das Starten und Stoppen der EA-Engine oder des EA-Datenbankservers erhöhen.

Anpassen der Zeitlimitüberschreitung für die EA-Engine

1. Beenden Sie SPSS Modeler.
2. Bearbeiten Sie die Datei *<Modeler-Server-Installationspfad>/ext/bin/pasw.entityanalytics/ea.cfg*, um den Wert der folgenden Option zu erhöhen:

`timeout, Wert`

Dabei gibt *Wert* den Wert der Zeitlimitüberschreitung für die EA-Engine an (der Standardwert ist 60).

Anpassen der Zeitlimitüberschreitung für den EA-Datenbankserver (nur DB2)

1. Beenden Sie SPSS Modeler.
2. Bearbeiten Sie die Datei *<Modeler-Server-Installationspfad>/ext/bin/pasw.entityanalytics/ea.cfg*, um den Wert der folgenden Option zu erhöhen:

`timeout, Wert`

Dabei gibt *Wert* den Wert der Zeitlimitüberschreitung für den DB2-Datenbankserver für die Entitätsanalyse an (der Standardwert ist 100).

Ausführen von IBM SPSS Modeler Entity Analytics mit SPSS Modeler-Client und SPSS Modeler Server auf demselben Windows-System

Wenn Sie IBM SPSS Modeler Entity Analytics sowohl auf SPSS Modeler-Client als auch auf SPSS Modeler Server auf demselben Windows-System installiert haben, verwenden Client und Server standardmäßig dasselbe Repository. Wenn getrennte Repositorys verwendet werden sollen, müssen Sie die Konfigurationsdatei *ea.cfg* auf **einem** der Systeme so bearbeiten, dass es einen anderen Portbereich und einen anderen Repository-Ordner verwendet.

Hinweis: Dieses Verfahren muss insbesondere dann ausgeführt werden, wenn Sie eine 32-Bit-Version des SPSS Modeler-Clients und eine 64-Bit-Version von SPSS Modeler Server (oder umgekehrt) verwenden.

1. Öffnen Sie die Datei *<Modeler-[Server-]Installationspfad>/ext/bin/pasw.entityanalytics/ea.cfg* für die Bearbeitung.
2. Ändern Sie die Einstellungen `min_port` und `max_port` so, dass andere Ports verwendet werden als beim anderen System. Weitere Informationen finden Sie im Thema „Konfigurieren von Portzuordnungen” auf Seite 33.
3. Ändern Sie die Einstellung `repository_data_directory` so, dass ein anderes Verzeichnis verwendet wird als beim anderen System.
4. Speichern und schließen Sie die Datei *ea.cfg*.

Löschen des Inhalts eines Entitätsrepositorys

Wenn Sie die Datensätze aus einem Entitätsrepository entfernen, die Konfigurationsinformationen jedoch beibehalten wollen, können Sie Daten aus dem Repository löschen.

Löschen aller Daten aus einem Repository

1. Öffnen Sie einen EA-Knoten.
2. Klicken Sie auf die Liste **Entitätsrepository**.
3. Klicken Sie auf **<Durchsuchen...>**, um das Dialogfeld "Entitätsauflösungsinstanzen" anzuzeigen.
4. Klicken Sie im Dialogfeld "Entitätsauflösungsinstanzen" auf die Liste **Repository-Name**.
5. Wählen Sie das Repository aus, dessen Inhalt Sie löschen möchten.
6. Wenn noch keine Verbindung besteht, geben Sie den Benutzernamen für den Administrator und das zugehörige Kennwort ein und klicken Sie auf **Verbinden**.
7. Wenn die Schaltfläche **Alle löschen** aktiviert ist, klicken Sie darauf.
8. Klicken Sie im Dialogfeld zum Löschen aller Datenquellen auf **Löschen**, um das Löschen des Repository-Inhalts zu bestätigen.

Löschen nicht verwendeter Datenquellen aus einem Repository

Wenn Sie in einem Entitätsrepository über eine Datenquelle verfügen, die Sie nicht mehr verwenden oder benötigen, können Sie die Quelle aus dem Repository löschen. Sie können mindestens eine zu löschende Datenquelle auswählen.

Löschen einer ausgewählten Datenquelle aus einem Repository

1. Öffnen Sie einen EA-Knoten.
2. Klicken Sie auf die Liste **Entitätsrepository**.
3. Klicken Sie auf **<Durchsuchen...>**, um das Dialogfeld "Entitätsauflösungsinstanzen" anzuzeigen.
4. Klicken Sie im Dialogfeld "Entitätsauflösungsinstanzen" auf die Liste **Repository-Name**.
5. Wählen Sie das Repository aus, aus dem Sie eine Datenquelle löschen wollen.
6. Wenn noch keine Verbindung besteht, geben Sie den Benutzernamen für den Administrator und das zugehörige Kennwort ein und klicken Sie auf **Verbinden**.

7. Wählen Sie in der Liste **Repository verwalten** die zu löschende Datenquelle aus. Wenn Sie weitere Datenquellen auswählen wollen, klicken Sie bei gedrückter Steuertaste auf die Datenquellen.
8. Wenn die Schaltfläche **Nicht verwendete löschen** aktiviert ist, klicken Sie darauf.
9. Klicken Sie im Dialogfeld zum Löschen nicht verwendeter Datenquellen auf **Löschen**, um das Löschen des Repository-Inhalts zu bestätigen.

Löschen eines Entitätsrepositorys

Wenn Sie ein Repository nicht mehr benötigen, können Sie es vollständig löschen.

Vorsicht: Dabei geschieht genau dies. **Dieser Vorgang kann nicht rückgängig gemacht werden.** Wenn Sie sich nicht sicher sind, verwenden Sie die Schaltfläche **Inhalt löschen**, um alle Quellendaten zu entfernen. Dabei bleibt die Repository-Konfiguration erhalten. Weitere Informationen finden Sie im Thema „Löschen des Inhalts eines Entitätsrepositorys" auf Seite 36.

Hinweis: Beim folgenden Verfahren wird davon ausgegangen, dass Sie über SPSS Modeler eine Verbindung mit dem Repository herstellen können und dass Ihnen der Benutzername und das Kennwort des Administrators für die Datenbank bekannt ist, die als Host für das Repository fungiert. Wenn dies nicht der Fall ist, verwenden Sie das Verfahren zum Löschen eines Repositorys, wenn keine Verbindung zu ihm hergestellt werden kann. Weitere Informationen finden Sie im Thema „Löschen eines Repositorys, wenn keine Verbindung zu ihm hergestellt werden kann".

Löschen eines Repositorys

1. Öffnen Sie einen EA-Knoten.
2. Klicken Sie auf die Liste **Entitätsrepository**.
3. Klicken Sie auf **<Durchsuchen...>**, um das Dialogfeld "Entitätsauflösungsinstanzen" anzuzeigen.
4. Klicken Sie im Dialogfeld "Entitätsauflösungsinstanzen" auf die Liste **Repository-Name**.
5. Wählen Sie das Repository aus, das Sie löschen möchten.
6. Wenn noch keine Verbindung besteht, geben Sie den Benutzernamen für den Administrator und das zugehörige Kennwort ein und klicken Sie auf **Verbinden**.
7. Wenn die Schaltfläche **Gesamtes Repository löschen** aktiviert ist, klicken Sie darauf.
8. Klicken Sie auf **Löschen**, um das Löschen des Repositorys zu bestätigen.
9. Klicken Sie auf **OK**, um die erfolgreiche Löschung zu bestätigen.

Löschen eines Repositorys, wenn keine Verbindung zu ihm hergestellt werden kann

Gehen Sie wie folgt vor, wenn Sie ein Entitätsrepository löschen möchten, aber entweder aufgrund von Konnektivitätsproblemen mit SPSS Modeler oder weil Sie den Benutzernamen bzw. das Kennwort vergessen haben, keine Verbindung zu ihm herstellen können.

Führen Sie die folgende Prozedur auf dem Computer durch, der als Host für die Repository-Datenbank fungiert.

Windows-Systeme

1. Öffnen Sie ein Fenster mit Eingabeaufforderung.
2. Geben Sie Folgendes ein:

   ```
   cd Modeler-Installationsverzeichnis
   cd ext\bin\pasw.entityanalytics
   delete_repository.bat Repository-Name
   ```

   Dabei ist *Modeler-Installationsverzeichnis* das Verzeichnis, in dem SPSS Modeler installiert ist, und *Repository-Name* ist der Name des Repositorys.

Anmerkung: Beim Repository-Namen muss die Groß-/Kleinschreibung beachtet werden.

3. Fahren Sie mit "Abschluss des Verfahrens" weiter unten in diesem Abschnitt fort.

UNIX-Systeme

1. Öffnen Sie eine Shell.
2. Geben Sie Folgendes ein:

   ```
   cd Modeler-Server-Installationsverzeichnis
   cd ext/bin/pasw.entityanalytics
   ./delete_repository.sh Repository-Name
   ```

   Dabei ist *Modeler-Server-Installationsverzeichnis* das Verzeichnis, in dem SPSS Modeler Server installiert ist, und *Repository-Name* ist der Name des Repositorys.

   Anmerkung: Beim Repository-Namen muss die Groß-/Kleinschreibung beachtet werden.

Abschluss des Verfahrens (alle Systeme)

1. Bestätigen Sie an der Eingabeaufforderung die Löschung des Repositorys durch Eingabe von `Y`.
2. Löschen Sie das Verzeichnis, das denselben Namen trägt wie das gelöschte Repository. Wenn Sie das Verzeichnis nicht löschen können, führen Sie einen Neustart des Computers durch und versuchen Sie es noch einmal.

Kapitel 4. Entitätsanalyse in Aktion

Informationen zu diesem Beispiel

In diesem Beispiel wird gezeigt, wie Sie durch die Hinzunahme der Entitätsanalyse die bereits beeindruckenden Ergebnisse, die Sie mit IBM SPSS Modeler erzielen können, weiter verbessern können.

In diesem Beispiel wird der Stream *loan_entity_analytics.str* verwendet, der auf die Datendatei *loan_applications.csv* verweist. Die Dateien stehen im Verzeichnis *Demos* jeder IBM SPSS Modeler-Installation zur Verfügung, bei der auch IBM SPSS Modeler Entity Analytics installiert ist. Das Verzeichnis *Demos* kann über die Programmgruppe "IBM SPSS Modeler" im Windows-Startmenü aufgerufen werden. Die Datei *loan_entity_analytics.str* befindet sich im Verzeichnis *Entity_Analytics*.

Hinweis: Damit Sie diesen Beispielstream ausführen können, müssen Sie ein Repository in Ihrem System erstellen. Erledigen Sie dies, bevor Sie mit diesem Beispiel fortfahren. Weitere Informationen finden Sie im Thema „Erstellen des Repositorys" auf Seite 14.

Beginnen wir mit einer vertrauten Situation: Die leitenden Angestellten einer Bank fragen sich, ob es bei bestimmten Kunden wahrscheinlich ist, dass es bei Krediten, die sie beantragt haben, zu einem Zahlungsausfall kommt. Die IT-Abteilung der Bank verwendet bereits seit langer Zeit SPSS Modeler, weshalb die Mitarbeiter der Abteilung bereits einen Stream erstellt und aus den bestehenden Daten zu ca. 700 früheren Krediten der Bank ein Vorhersagemodell erstellt haben. Diese Kredite wurden entweder zurückgezahlt, oder die Kunden kamen ihren Rückzahlungsverpflichtungen nicht nach.

Ursprüngliches Modell

Hier sehen Sie, wie die Bankmitarbeiter ihr Modell erstellten und welche Erkenntnisse sie daraus gewonnen haben.

Abbildung 2. Ursprünglicher Stream mit Modellierungsknoten

Neben den Details zu früheren Krediten enthält das Dataset *loan_applications.csv* Details zu 150 Kunden, über deren Kreditanträge noch nicht entschieden wurde. Dies ergibt insgesamt 850 Datensätze.

Nicht alle Felder aus dem Dataset sind für die Erstellung der Vorhersage von Nutzen. So können beispielsweise Namensfelder ignoriert werden. Der Typknoten filtert die zu ignorierenden Felder heraus, indem er deren Rolle auf **Keine** (None) setzt. Bei den für die Vorhersage zu verwendenden Feldern wird die Rolle auf **Eingabe** (Input) gesetzt, und bei dem Feld, dessen Wert das Modell vorherzusagen versucht, wird die Rolle auf **Ziel** (Target) gesetzt.

Abbildung 3. Im Typknoten festgelegte Feldrollen

Da das Modell seine Vorhersagen ausschließlich auf der Grundlage der Daten aus der Vergangenheit erstellen darf, enthält der Stream einen Auswahlknoten, der nur die Kredite mit aufnimmt, die *nicht* als "Pending" (Ausstehend) gekennzeichnet sind. Die 150 ausstehenden Kredite werden somit verworfen.

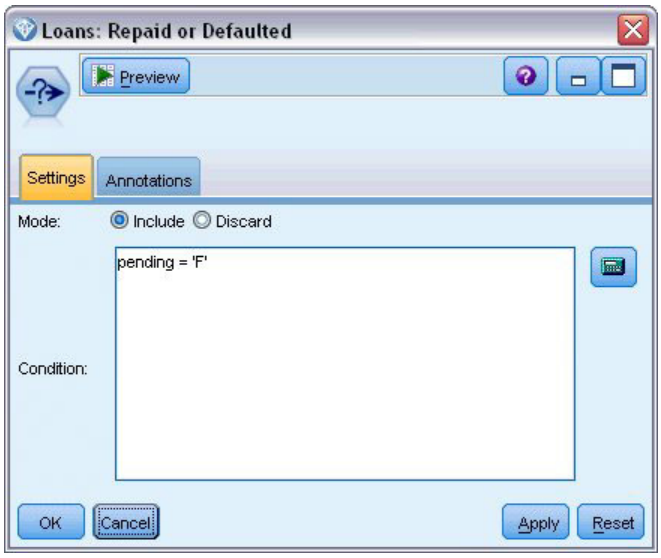


Abbildung 4. Verwerfen der ausstehenden Kreditanträge

Nachdem die ausstehenden Kredite verworfen wurden, werden nur die Details der verbleibenden 700 Kredite, die entweder zurückgezahlt oder nicht zurückgezahlt wurden, an den Modellierungsknoten weitergegeben. Die Bank könnte einen von mehreren SPSS Modeler-Algorithmen verwendet haben, um ein

gutes Modell zu erstellen. In diesem Fall wurde ein C&R-Baumknoten verwendet, mit dem ein Modell erstellt wird, das die wahrscheinlichen Kreditausfälle auf der Grundlage der früheren Leistung der Bankkunden vorhersagt.

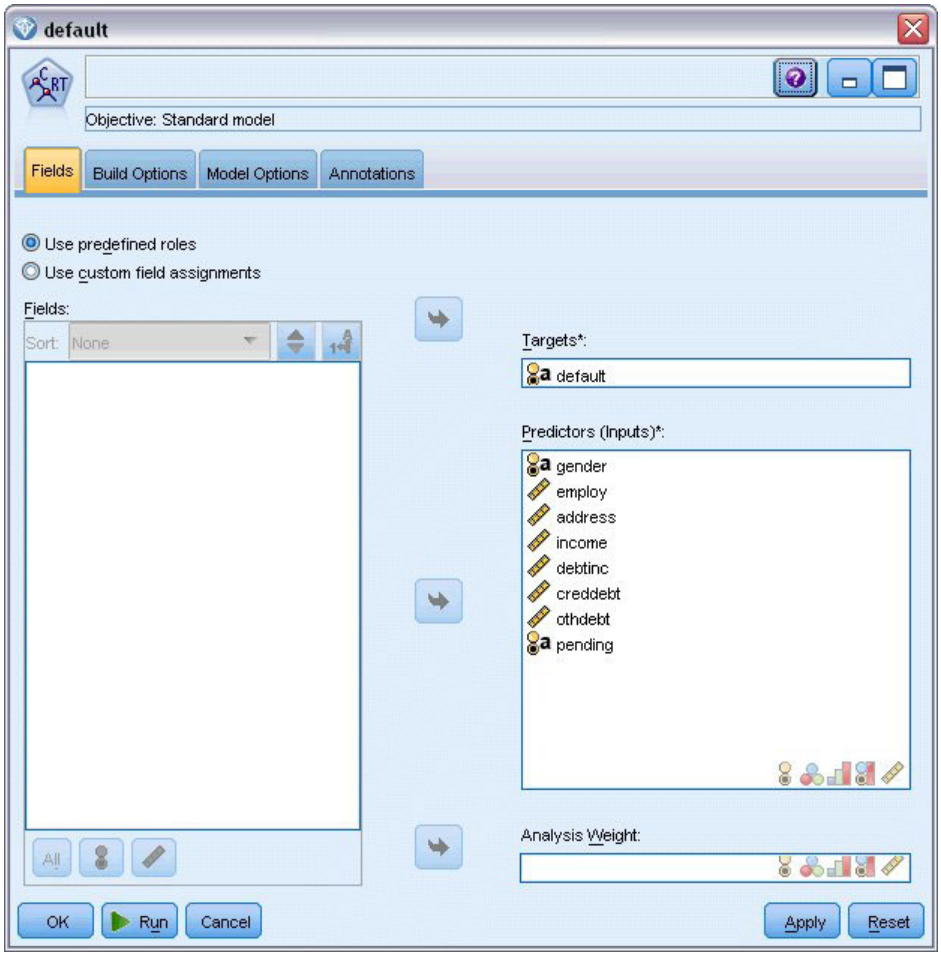


Abbildung 5. Zuweisen von Prädiktor- und Zielfeldern

Die zur Erstellung der Vorhersage verwendeten Felder werden als Prädiktorfelder gekennzeichnet und das Feld, dessen Wert mit dem Modell vorhergesagt werden soll - in diesem Fall **default** (Zahlungsausfall) - wird als Zielfeld festgelegt, wie zuvor vom Typknoten definiert.

Bei der Ausführung dieses Streams wird ein Modellnugget erstellt, das das aus den Prädiktorfeldern erstellte Modell enthält.

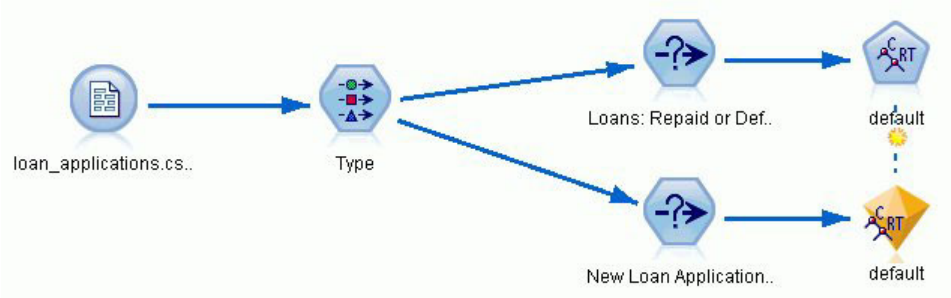


Abbildung 6. Stream mit hinzugefügtem Modellnugget

Nun kann der Analyst der Bank das Modell verwenden, um mit der Vorhersage zu beginnen, ob es bei Kunden mit ausstehenden Kreditanträgen mit hoher Wahrscheinlichkeit zu einem Zahlungsausfall kommt. Der Analyst verwendet das ursprüngliche Dataset und fügt einen Auswahlknoten ein, der dieses Mal nur die 150 als "Pending" (Ausstehend) gekennzeichneten Datensätze enthält, die zuvor verworfen worden waren. Der Analyst leitet diese Datensätze direkt in das Modell weiter und fügt dabei einen Verteilungsknoten für eine visuelle Darstellung der Vorhersagen des Modells hinzu.

Abbildung 7. Stream zur Auswahl neuer Kreditanträge mit hinzugefügten Verteilungsknoten

Der Verteilungsknoten zeigt die Verteilung der Werte des Felds *$R-default* im Modell. Dieses Feld wird dem Datenmodell bei der Ausführung vom C&R-Baumknoten hinzugefügt. Das Feld enthält die Vorhersage, ob die einzelnen neuen Bewerber ihren Kredit zurückzahlen oder nicht und wir verwenden dieses Feld später, um zu vergleichen, welche Wirkung die Hinzunahme der Entitätsanalyse hat.

Bei der Ausführung dieses Teils des Streams erfährt der Analyst aus der Ausgabe des Verteilungsknotens, dass 137 der 150 neuen Antragsteller ihre Kredite vermutlich zurückzahlen werden. Für die verbleibenden 13 wird ein Zahlungsausfall vorhergesagt, weshalb der Analyst der Bank vermutlich empfehlen wird, diese Anträge abzulehnen.

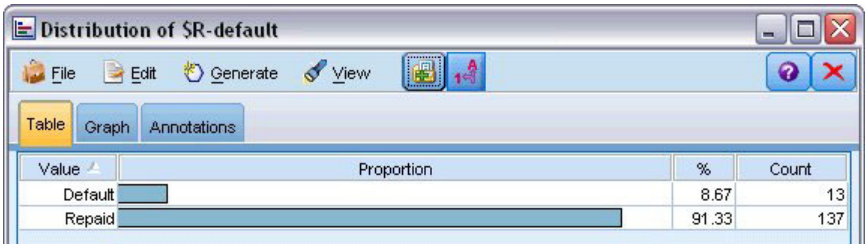


Abbildung 8. Ausgabe aus dem Verteilungsknoten ohne Entitätsanalyse

Hinzunahme der Entitätsanalyse

Betrachten wir nun, ob die Situation verbessert werden kann, wenn eine Entitätsanalyse mit in die Gleichung aufgenommen wird. Stellen Sie sich vor, Sie sind ein Experte auf dem Gebiet der Entitätsanalyse, der von der Bank hinzugezogen wurde, um mögliche betrügerische Einträge in den Kundendatensätzen in den Quellendaten zu untersuchen. Es kann möglicherweise doppelte Datensätze aufgrund von Fehlern beim Dateneintrag geben, es ist jedoch ebenfalls möglich, dass ein Antragsteller für einen Kredit versucht, seine wahre Identität zu verschleiern. In jedem Fall muss die Bank den tatsächlichen Sachverhalt aufdecken.

Im Rahmen dieses Beispiels gehen wir davon aus, dass bereits ein Entitätsrepository erstellt wurde. Weitere Informationen finden Sie im Thema „Erstellen des Repositorys" auf Seite 14.

Übertragen der Quellendaten in das Repository

Zunächst müssen Sie einen EA-Exportknoten zum Datenquellenknoten hinzufügen, damit Sie die Quellendaten in das Entitätsrepository exportieren können.

Bevor Sie die Daten exportieren können, müssen Sie Felder in der Datenquelle Merkmalen im Entitätsrepository zuordnen. Dies ist notwendig, da verschiedene Datenquellen möglicherweise verschiedene Feldnamen für dieselbe Art von Informationen verwenden. Das Entitätsrepository stellt eine Standardmenge an Informationstypen (sogenannte Merkmale) bereit, um Dopplungen zu vermeiden.

Legen Sie im EA-Exportknoten die Details für das Repository fest: die Verbindungsdetails, den Quellentag (in diesem Fall zur Angabe der Datenquelle **TEST**), den Entitätstyp (das verwendete Merkmalset, d. h. das Set mit dem Namen **PERSON**) und das Feld mit dem eindeutigen Schlüssel (zur eindeutigen Angabe der einzelnen Datensätze). Verwenden Sie in diesem Fall das Feld **key** (Schlüssel) als eindeutigen Schlüssel.

Nun können Sie die Zuordnungen einrichten. In dem verwendeten Merkmalset gibt es Merkmale, die den Feldern *fname*, *mname*, *lname*, *generation*, *dob*, *gender*, *addr1*, *city*, *country*, *postcode*, *phone*, *email*, *ssn*, *drlic* und *passport* entsprechen.

Beginnen Sie, indem Sie die Zuordnung für *fname* festlegen. Doppelklicken Sie in der Spalte **Zugeordnet zu Merkmal** der Tabelle auf die Zeile *fname*, blättern Sie nach unten bis zum Eintrag **NAME.GIVEN_NAME** und klicken Sie darauf, um die Zuordnung zu erstellen.

Ordnen Sie nun die restlichen Felder, die entsprechende Merkmale aufweisen zu, sodass die vollständige Menge der Zuordnungen wie folgt aussieht.

Tabelle 12. Repository-Merkmalen zugeordnete Felder.

Feld	**Zugeordnet zu Merkmal**
fname	NAME.GIVEN_NAME
mname	NAME.MIDDLE_NAME
lname	NAME.SUR_NAME
generation	NAME.NAME_GEN
dob	DOB.DOB
gender	GENDER.GENDER
addr1	ADDRESS.ADDR1
city	ADDRESS.CITY
country	ADDRESS.COUNTRY
postcode	ADDRESS.POSTAL_CODE
phone	PHONE.PHONE_NUM
email	EMAIL_ADDR.ADDR
ssn	SSN.ID_NUM
drlic	DRLIC.ID_NUM
passport	PASSPORT.ID_NUM

Klicken Sie auf **Ausführen**, um die Daten in das Repository zu exportieren. Dies nimmt eine gewisse Zeit in Anspruch. Wenn das Dialogfeld "Ausführungs-Feedback" automatisch geschlossen wird, ist der Exportvorgang abgeschlossen.

Lesen der aufgelösten Identitäten

Beim Export der Daten in das Repository beginnt das Entitätsanalysesystem mit der Auflösung möglicher Identitätskonflikte, indem es eine eindeutige Entitäts-ID zuweist, die später als Feld *$EA-ID* angezeigt wird. (*Hinweis*: Dies ist nicht dasselbe Feld wie das Feld "Eindeutiger Schlüssel" im EA-Exportknoten, das zur eindeutigen Kennzeichnung von Datenquellen-Datensätzen dient.)

Der erste Schritt beim Lesen der aufgelösten Identitäten besteht darin, dem Stream einen EA-Quellenknoten hinzuzufügen. Dieser Quellenknoten sollte in dieser Phase noch völlig unverbunden sein.

Öffnen Sie den EA-Quellenknoten und legen Sie die Details für das Entity-Repository fest. Daraufhin wird eine Liste der Datenquellen angezeigt, die in das Repository exportiert wurden. Im vorliegenden Fall gibt es nur eine einzige.

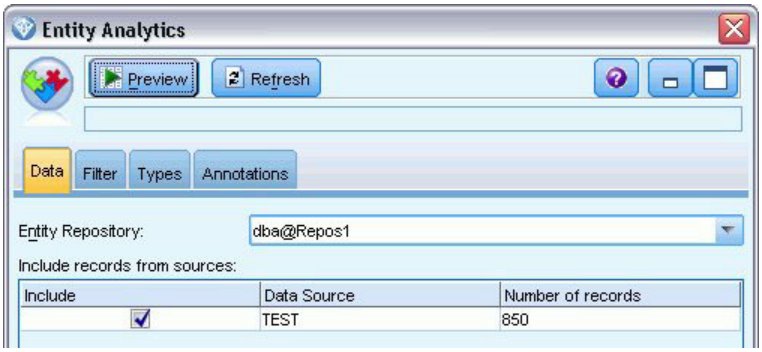


Abbildung 9. Auswählen einer Datenquelle im Repository

Aktivieren Sie das Kontrollkästchen für die Datenquelle **TEST** und klicken Sie auf **OK**.

Betrachten wir nun, wie das Entitätsanalysesystem die Daten bearbeitet hat. Verbinden Sie einen Tabellenknoten mit dem EA-Quellenknoten, öffnen Sie den Tabellenknoten und klicken Sie auf **Ausführen**, um das Ausgabefenster des Tabellenknotens anzuzeigen.

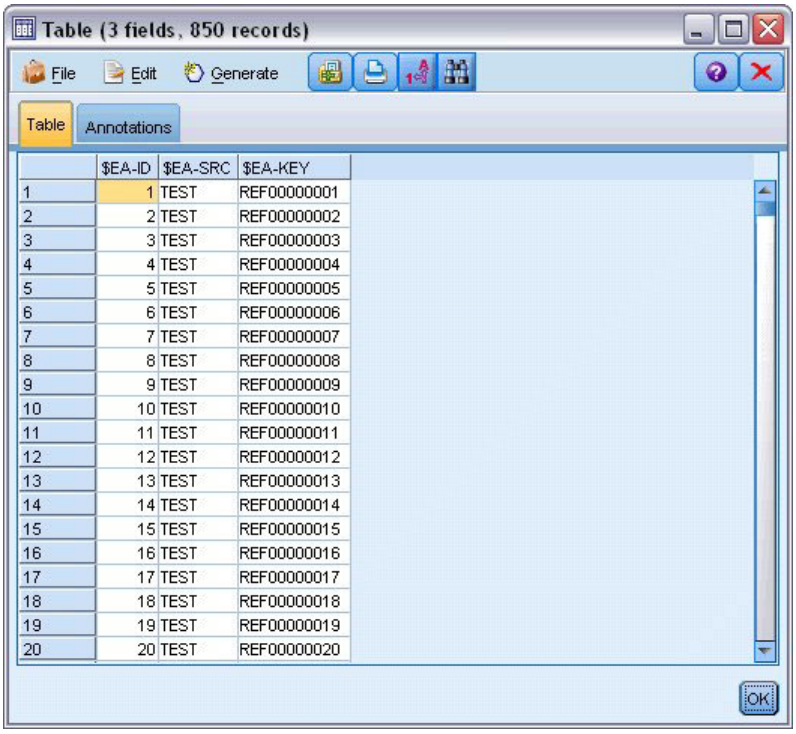


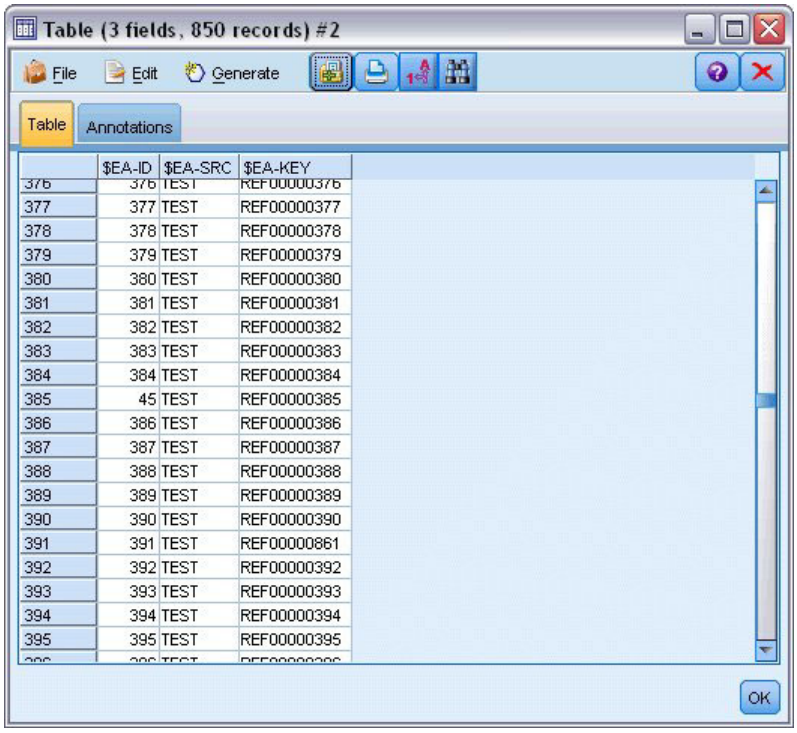
Abbildung 10. Ausgabe aus dem Tabellenknoten

Nur ein einziges Feld sieht bekannt aus, nämlich das mit der Beschriftung *$EA-KEY*. Dabei handelt es sich tatsächlich um das Feld *key* (Schlüssel) aus den Quellendaten und es ist enthalten, weil Sie es im EA-Exportknoten als Feld "Eindeutiger Schlüssel" ausgewählt hatten.

Das System hat jedoch noch zwei weitere Felder hinzugefügt. Das Feld *$EA-ID* ist nicht die eindeutige ID der Quellendatensätze, sondern die eindeutige ID der aufgelösten Identitäten. Der Unterschied wird in Kürze deutlich. Das Feld *$EA-SRC* gibt an, woher die Daten stammen. Hier steht **TEST**, da dies der Quellentag ist, den Sie im EA-Exportknoten zugewiesen haben.

Was ist mit den ganzen anderen Feldern in den Quellendaten geschehen? Keine Sorge, sie befinden sich noch immer im Repository. Es ist nur so, dass der EA-Quellenknoten aus Leistungsgründen nur das Mindestset an Feldern zur weiteren Verarbeitung nach unten in den Stream weitergibt.

Blättern Sie nun in der Ausgabe des Tabellenknotens nach unten bis zur Zeile 385.



	$EA-ID	$EA-SRC	$EA-KEY
376	376	TEST	REF00000376
377	377	TEST	REF00000377
378	378	TEST	REF00000378
379	379	TEST	REF00000379
380	380	TEST	REF00000380
381	381	TEST	REF00000381
382	382	TEST	REF00000382
383	383	TEST	REF00000383
384	384	TEST	REF00000384
385	45	TEST	REF00000385
386	386	TEST	REF00000386
387	387	TEST	REF00000387
388	388	TEST	REF00000388
389	389	TEST	REF00000389
390	390	TEST	REF00000390
391	391	TEST	REF00000861
392	392	TEST	REF00000392
393	393	TEST	REF00000393
394	394	TEST	REF00000394
395	395	TEST	REF00000395

Abbildung 11. Unterschied zwischen Tabellen-Ausgabezeilen und $EA-ID-Nummern

Beachten Sie, dass die *$EA-ID*-Nummer hier nicht in der richtigen Reihenfolge zu stehen scheint. Das Entitätsanalysesystem hat ermittelt, dass Datensatz REF00000385 auf die Person verweist, die als Entität 45 ermittelt wurde und die außerdem Datensatz REF00000045 aufweist. Wenn wir in der Ausgabe weiter nach unten blättern, gibt es weitere Nummern mit falscher Reihenfolge, beispielsweise in den Zeilen 485, 517, 520 usw. Wir sehen uns das besser ein wenig genauer an.

Heben wir zunächst die Tatsache hervor, dass das Feld *$EA-KEY* die Daten aus dem Feld *key* (Schlüssel) in den Quellendaten enthält, indem wir es in *key* umbenennen. Verbinden Sie einen Filterknoten mit dem EA-Quellenknoten und öffnen Sie den Filterknoten. Doppelklicken Sie auf die Zeichenfolge **$EA-KEY** in der zweiten Spalte **Feld** und geben Sie `key` ein.

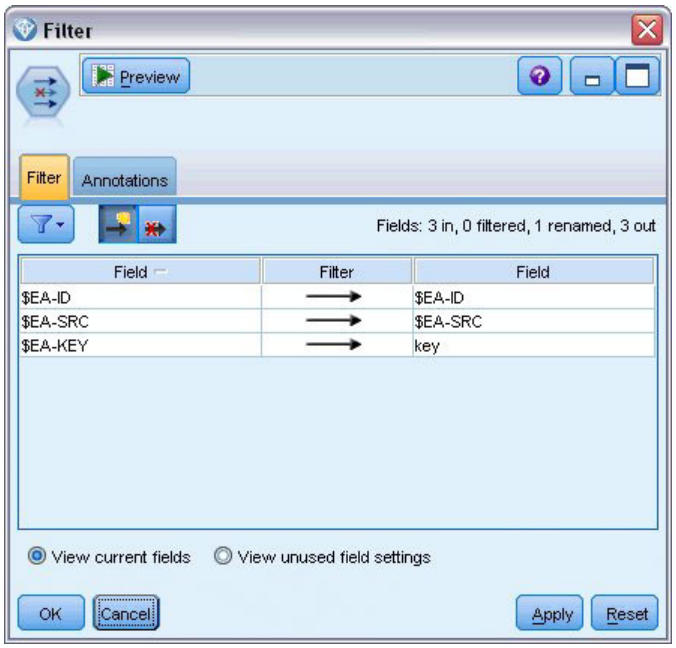


Abbildung 12. Umbenennen des Felds "$EA-KEY"

Klicken Sie auf **OK**, um den Filterknoten zu schließen.

Nun müssen wir die *$EA-ID*-Entitäts-IDs in aufsteigender Reihenfolge sortieren. Verbinden Sie einen Sortierknoten mit dem Filterknoten. Öffnen Sie den Sortierknoten, klicken Sie auf die oberste Schaltfläche neben der Tabelle **Sortieren nach**, wählen Sie **$EA-ID** und klicken Sie auf **OK**.

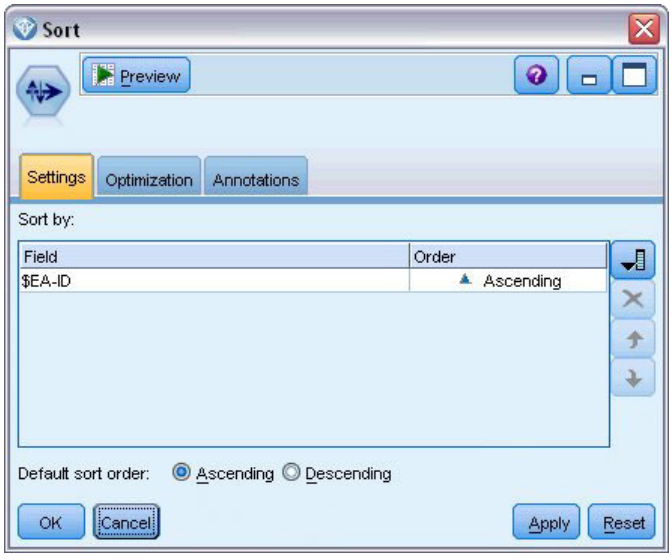


Abbildung 13. Sortieren der Entitäts-IDs in aufsteigender Reihenfolge

Behalten Sie die Sortierreihenfolge **Aufsteigend** bei und klicken Sie auf **OK**.

Nun müssen Sie ein zusätzliches Feld erstellen, das angibt, ob ein Datensatz eindeutig oder ein Duplikat ist. Verbinden Sie einen Ableitungsknoten mit dem Sortierknoten. Öffnen Sie den Ableitungsknoten und setzen Sie den Wert für **Ableitungsfeld** auf `IsDuplicate`. Wählen Sie in der Liste **Ableitungstyp** die Option **Flag** aus. Dadurch wird auch **Feldtyp** auf **Flag** gesetzt. Setzen Sie das Feld **Wahr-Wert** auf **Duplicate** und das Feld **Falsch-Wert** auf **Unique**.

Um doppelte Datensätze zu finden, verwenden Sie eine spezielle Sequenzfunktion, @OFFSET, die im Lieferumfang von SPSS Modeler enthalten ist.

Geben Sie Folgendes in das Feld **Wenn** ein:

```
'$EA-ID' = @OFFSET('$EA-ID',1) or '$EA-ID' = @OFFSET('$EA-ID',-1))
```

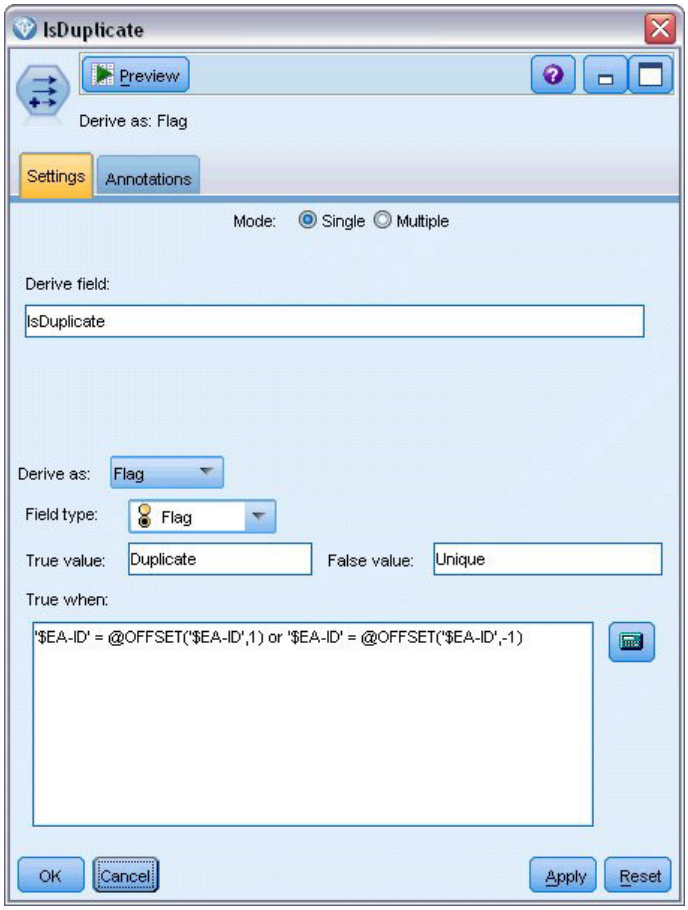
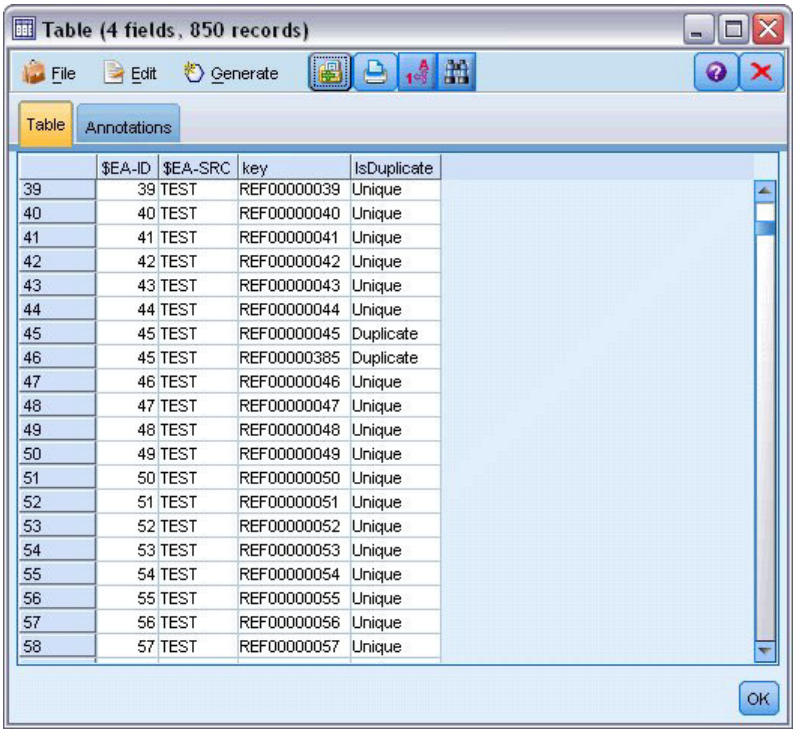


Abbildung 14. Festlegen der Bedingung im Ableitungsknoten

Wenn die Entitäts-IDs in aufsteigender Reihenfolge sortiert sind, können Sie mit der Funktion @OFFSET überprüfen, ob untereinander stehende Entitäts-IDs identisch sind, was bedeuten würde, dass es sich bei den Datensätzen um Duplikate handelt. In diesem Fall wird der zugehörige Wert für *IsDuplicate* auf *Duplicate* gesetzt; andernfalls wird er auf *Unique* gesetzt.

Klicken Sie auf **OK**, um den Knoten zu schließen.

Um die Wirkung des Ableitungsknotens anzuzeigen, verbinden Sie einen Tabellenknoten mit dem Ableitungsknoten, öffnen Sie den Tabellenknoten und klicken Sie auf **Ausführen**. Blättern Sie im Ausgabefenster des Tabellenknotens nach unten bis zur Zeile 45.



	$EA-ID	$EA-SRC	key	IsDuplicate
39	39	TEST	REF00000039	Unique
40	40	TEST	REF00000040	Unique
41	41	TEST	REF00000041	Unique
42	42	TEST	REF00000042	Unique
43	43	TEST	REF00000043	Unique
44	44	TEST	REF00000044	Unique
45	45	TEST	REF00000045	Duplicate
46	45	TEST	REF00000385	Duplicate
47	46	TEST	REF00000046	Unique
48	47	TEST	REF00000047	Unique
49	48	TEST	REF00000048	Unique
50	49	TEST	REF00000049	Unique
51	50	TEST	REF00000050	Unique
52	51	TEST	REF00000051	Unique
53	52	TEST	REF00000052	Unique
54	53	TEST	REF00000053	Unique
55	54	TEST	REF00000054	Unique
56	55	TEST	REF00000055	Unique
57	56	TEST	REF00000056	Unique
58	57	TEST	REF00000057	Unique

Abbildung 15. Ausgabe aus dem Ableitungsknoten

Erinnern Sie sich an die Anzeige der Ausgabe direkt aus dem EA-Quellenknoten? Das System hatte bereits ermittelt, dass Datensatz REF00000385 dieselbe Person referenziert wie Entität 45. Nun sind wir dabei einen Schritt weitergegangen und haben die Tatsache, dass es sich bei den Datensätzen REF00000045 und REF00000385 um Duplikate handelt, da sie beide Entität 45 referenzieren, durch ein Flag gekennzeichnet.

Blättern Sie im Ausgabefenster weiter nach unten, um die anderen Datensätze zu sehen, die als Duplikate gekennzeichnet sind.

Um einen Bericht zu erhalten, in dem die doppelten Datensätze aufgeführt sind, verbinden Sie einen Berichtsknoten (von der Registerkarte "Ausgabe" der Knotenpalette) mit dem Ableitungsknoten *IsDuplicate*. Öffnen Sie den Berichtsknoten, kopieren Sie den folgenden Text in das Eingabefeld der Registerkarte "Vorlage" und klicken Sie auf **Ausführen**.

```
<html>
<h1>Liste der doppelten Kundendatensätze

<h2>Dieser Bericht wurde generiert am: [@TODAY]

<h2>Doppelte Datensätze
<table>
 <tr>
   <td>Entitäts-ID</td>
   <td>Schlüssel</td>
 </tr>

#WHERE IsDuplicate = "Duplicate"

 <tr>
   <td>['$EA-ID']</td>
   <td>[key]</td>
 <tr>
```

```
#
</table>

</html>
```

Dadurch ergibt sich die folgende Ausgabe:

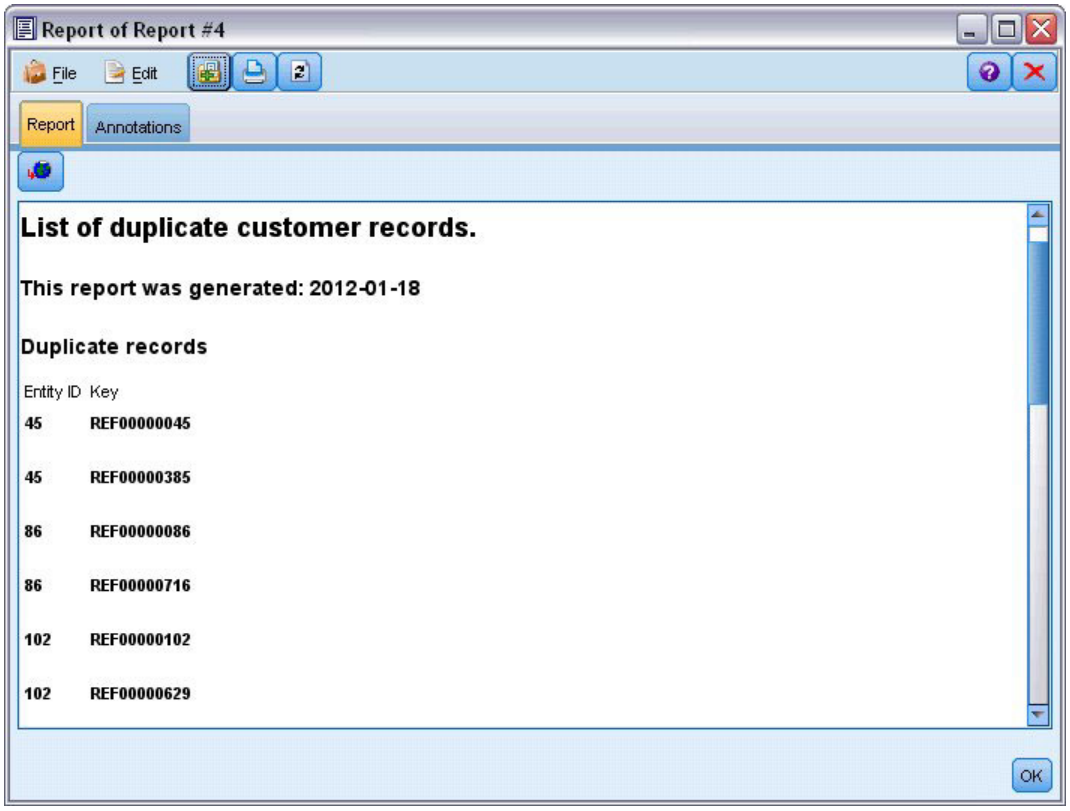


Abbildung 16. Ausgabe aus dem Berichtsknoten

In diesem Fall wird für den Bericht HTML-Format verwendet. Sie könnten jedoch auch XML- oder ASCII-Formatierung verwenden.

Vergleich der Entitätsanalyseausgabe mit dem ursprünglichen Modell

Der letzte Arbeitsschritt in diesem Beispiel besteht darin zu ermitteln, ob sich die Vorhersage durch die Hinzunahme einer Entitätsanalyse gegenüber der ursprünglichen Vorhersage der Bank verändert. Sie erinnern sich vielleicht, dass durch das ursprüngliche Modell 13 Kreditausfälle bei den 150 ausstehenden Anträgen vorhergesagt wurden. Sie verwenden nun einen Zusammenführungsknoten ("Mergen"), um die Ausgabe aus diesem Modell mit Informationen zu doppelten Datensätzen aus der Entitätsanalyse zusammenzuführen, um zu sehen, ob sich die Vorhersage dadurch ändert.

Zunächst müssen Sie sicherstellen, dass die durch die Entitätsanalyse neu hinzugekommenen Felder die richtigen Datentypen (bzw. *Messniveaus* laut SPSS Modeler Terminologie) aufweisen. Fügen Sie einen Typknoten zum Ableitungsknoten **IsDuplicate** hinzu, öffnen Sie den Typknoten und klicken Sie auf die Schaltfläche **Werte lesen**.

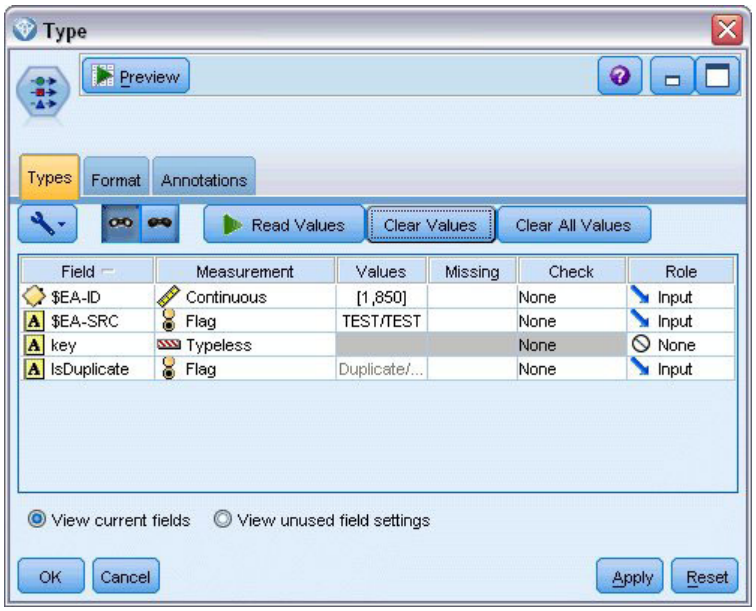


Abbildung 17. Einstellungen für den Typknoten

Sie können nun den Zusammenführungsknoten hinzufügen. Verbinden Sie ihn mit dem Typknoten und auch mit dem goldenen Nugget, das das ursprüngliche Modell enthält. Klicken Sie dazu mit der rechten Maustaste auf das goldene Nugget, wählen Sie **Verbinden** und klicken Sie dann auf den Zusammenführungsknoten, der nun zwei Eingabepfeile aufweisen sollte.

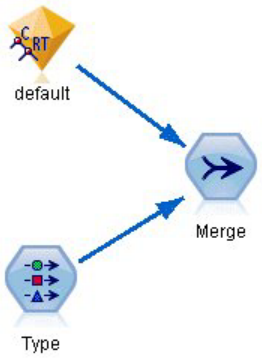


Abbildung 18. Eingaben für den Zusammenführungsknoten

Öffnen Sie den Zusammenführungsknoten, setzen Sie das **Zusammenführungsverfahren** auf **Schlüssel** und klicken Sie auf die Schaltfläche mit dem Rechtspfeil, um das Feld **key** aus dem Bereich **Mögliche Schlüsselfelder** in den Bereich **Verwendete Schlüsselfelder** zu verschieben, und klicken Sie dann auf **OK**.

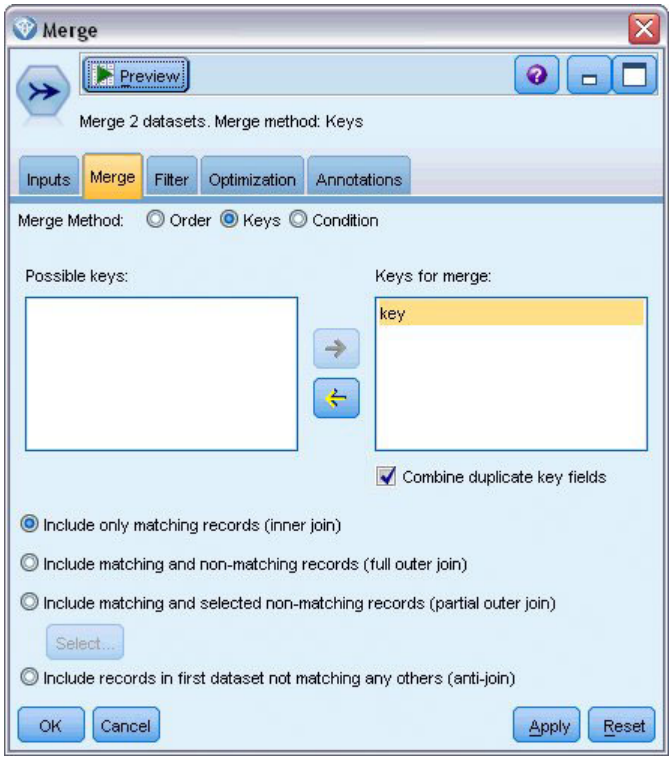


Abbildung 19. Angabe des Schlüsselfelds für die Zusammenführung

Sie sind nun fast so weit, dass Sie den Vergleich durchführen können. Wenn Sie nun jedoch einen Verteilungsknoten hinzufügen und ausführen würden, wären keinerlei Änderungen gegenüber der ursprünglichen Vorhersage zu bemerken. Auch wenn der Stream nun die Ausgaben aus dem ursprünglichen Modellnugget mit den neuen Feldern zusammenführt, die von der Entitätsanalyse erstellt wurden, wurde das Vorhersagefeld selbst (*$R-default*) im Datenmodell noch nicht mit den neuen Informationen aktualisiert.

Dazu verwenden Sie einen Füllerknoten, der Feldwerte ersetzen kann. Verbinden Sie einen Füllerknoten mit dem Zusammenführungsknoten und öffnen Sie den Füllerknoten.

Klicken Sie auf die obere Schaltfläche rechts neben **Felder ausfüllen**, blättern Sie ans Ende der Liste, wählen Sie **$R-default** aus und klicken Sie auf **OK**. Das ist das Feld, dessen Werte geändert werden sollen, wenn die im restlichen Dialogfeld angegebene Bedingung erfüllt ist.

Vergewissern Sie sich zur Angabe der Bedingung, dass **Ersetzen** auf**Anhand der Bedingung** gesetzt ist, und geben Sie dann im Feld **Bedingung** Folgendes ein:

```
default /="default" and IsDuplicate  = "Duplicate"
```

Geben Sie Folgendes in das Feld **Ersetzen durch** ein:

```
"default"
```

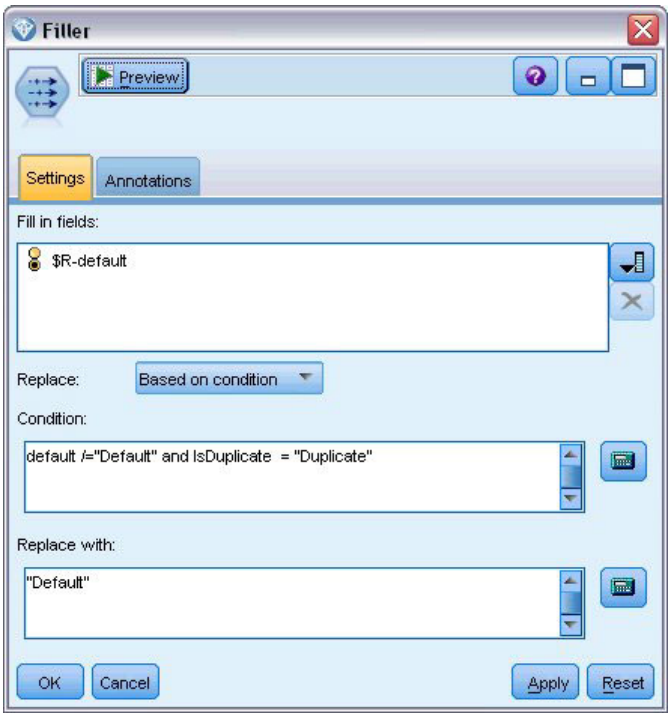


Abbildung 20. Angabe der Bedingung für das Ersetzen von Feldwerten

Diese Einstellungen müssen ein wenig erläutert werden. Die Bedingung gibt an, dass für jeden Datensatz, bei dem der Wert des Felds *default* (Zahlungsausfall) aus den ursprünglichen Daten nicht gleich **Default** ist und bei dem der Datensatz als Duplikat gekennzeichnet wurde, der Wert des Felds *$R-default* im Modell auf **Default** gesetzt wird.

Das Feld *$R-default* ist das Feld im Modell, das die Vorhersage enthält, ob ein Kunde den Kredit vermutlich nicht zurückzahlen wird. Auf diese Weise werden Kunden mit doppelten Datensätzen als Kunden mit möglichem Zahlungsausfall zum Modell hinzugefügt.

Klicken Sie auf **OK**, um den Füllerknoten zu schließen.

Nun können Sie schließlich anzeigen, welche Unterschiede sich durch die Entitätsanalyse ergeben haben. Verbinden Sie in der Diagrammpalette einen Verteilungsknoten mit dem Füllerknoten und öffnen Sie den Verteilungsknoten. Klicken Sie auf die Liste **Feld** und wählen Sie die Option **$R-default** aus.

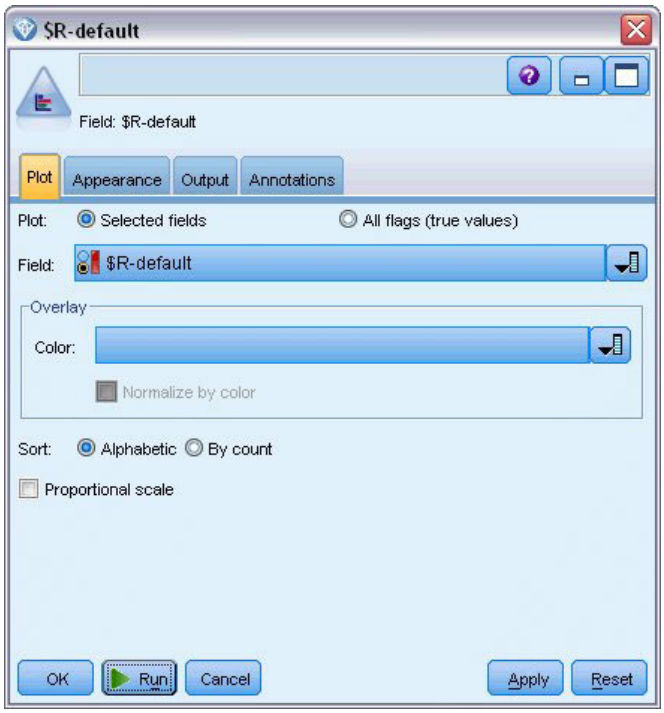


Abbildung 21. Einstellungen für den Verteilungsknoten

Klicken Sie auf **Ausführen**, um das Diagramm der neuen Vorhersage zu erstellen.

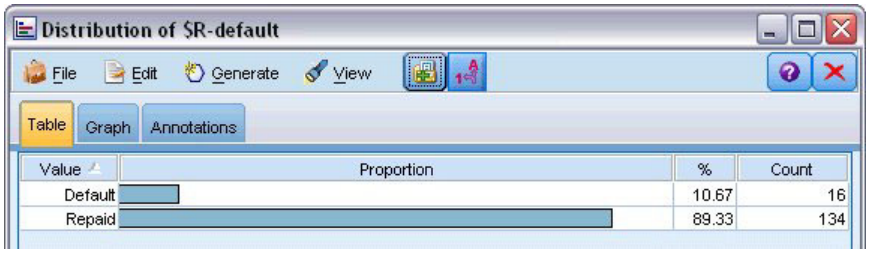


Abbildung 22. Ausgabe aus dem Verteilungsknoten nach der Entitätsanalyse

Es gibt nun 16 risikoreiche Anträge statt bisher 13. Diese zusätzlichen Anträge könnten sehr kostspielig werden, wenn es tatsächlich zu einem Kreditausfall kommen sollte, sodass Sie der Bank den Nutzen einer Ergänzung ihrer Risikobewertung durch Entitätsanalyse grafisch demonstrieren können.

Zusammenfassung

Dieses Beispiel hat gezeigt, wie Sie durch Entitätsanalyse Dopplungen von Datensätzen bei Daten zu Personen bzw. Organisationen ausmerzen und dadurch die Vorhersagequalität verbessern können.

Hinweis: Idealerweise sollten doppelte Datensätze vor allen sonstigen Verarbeitungsschritten entfernt werden. Sie könnten dies weiter ausbauen, indem Sie den Knoten "Automatisierte Datenaufbereitung" (ADP - Automated Data Preparation) verwenden, um Ihre Daten zu analysieren und Korrekturen zu identifizieren, problematische oder vermutlich überflüssige Felder auszuschließen, bei Bedarf neue Attribute abzuleiten und die Leistung durch intelligente Screeningverfahren zu verbessern.

Eine Kombination aus Entitätsanalyse und automatischer Datenvorbereitung kann dabei helfen, sicherzustellen, dass Sie mit möglichst sauberen Daten arbeiten.

Anhang. Scripteigenschaften für IBM SPSS Modeler Entity Analytics

Scripterstellung mit IBM SPSS Modeler Entity Analytics

Die Scripterstellung in IBM SPSS Modeler Entity Analytics ist ein leistungsstarkes Tool, mit dem Prozesse an der Benutzerschnittstelle automatisiert werden. Scripts können dieselben Arten von Aktionen durchführen, die Sie mit einer Maus oder einer Tastatur durchführen. So können Sie Aufgaben automatisieren, die bei einer manuellen Durchführung sehr viele Wiederholungen verlangen oder sehr viel Zeit beanspruchen. Weitere Informationen zur Verwendung von Scripts finden Sie im Handbuch *ScriptingAutomation.pdf*, das mit IBM SPSS Modeler bereitgestellt wird.

Allgemeine Eigenschaften

Eigenschaften, die alle IBM SPSS Modeler Entity Analytics-Knoten besitzen, sind in der folgenden Tabelle aufgelistet. Informationen zu speziellen Knoten finden Sie in den nachfolgenden Abschnitten.

Tabelle 13. Allgemeine Eigenschaften

Eigenschaftsname	Datentyp	Eigenschaftsbeschreibung
`entity_repository`	[*'Feld'*,*'Feld'*, ... ,*'Feld'*]	Die Repository-Verbindungszeichenfolge. Format: [*'Repository-Name'*, *'Benutzername'*, *'Kennwort'*] Beispiel: `entity_repository = [ 'repos1', 'dba', 'psw1' ]`
`entity_type`	*Zeichenfolge*	Der zu verwendende Entitätstyp (Merkmalset). Beispiel: `entity_type = 'PERSON'`

Eigenschaften von "entityanalytics_exportnode"

Der EA-Exportknoten ist ein Endknoten, der Entitätsdaten aus einer Datenquelle liest und die Daten zum Zweck der Entitätsauflösung in ein Repository exportiert.

Tabelle 14. Eigenschaften von "entityanalytics_exportnode"

Eigenschaften von `entityanalytics_exportnode`	Datentyp	Eigenschaftsbeschreibung
`mode`	`Add` `PurgeFirst`	Exportmodus. `Add` fügt dem vorhandenen Inhalt des Repositorys Datensätze aus Quellendateien hinzu. `PurgeFirst` entfernt vorhandene Inhalt vor dem Export.
`source_tag`	*Zeichenfolge*	Die Datenquellen-ID. Beispiel: `source_tag = 'CUST'`

Tabelle 14. Eigenschaften von "entityanalytics_exportnode" (Forts.)

Eigenschaften von `entityanalytics_exportnode`	**Datentyp**	**Eigenschaftsbeschreibung**
`unique_key_field`	*Zeichenfolge*	Eingabefeld für eindeutige IDs von Datensätzen. Beispiel: `unique_key_field = 'ID'`
`field_mapping`	[['*Feldname*' '*Merkmal.Element*' '*Verwendungstyp*']...]	Ordnet Eingabefelder dem entsprechenden Merkmal im Repository zu. Beispiel: `field_mapping = [['fname' 'NAME.GIVEN_NAME' ''] ['addr1' 'ADDRESS.ADDR1' 'PRIMARY']]` *Hinweis*: Wenn Sie *Verwendungstyp* auf die Entsprechung von "(Auto)" setzen wollen, verwenden Sie `''` wie im ersten Beispiel oben.

Eigenschaften von "entityanalytics_sourcenode"

Der EA-Quellenknoten liest die aufgelösten Entitäten aus dem Repository und gibt diese Daten zur weiteren Verarbeitung, beispielsweise zur Formatierung als Bericht, an den Stream weiter.

Tabelle 15. Eigenschaften von "entityanalytics_sourcenode"

Eigenschaften von `entityanalytics_sourcenode`	**Datentyp**	**Eigenschaftsbeschreibung**
`source_tags`	*Liste*	Liste der Tags für Datenquellen, die aus dem Repository extrahiert werden sollen. Beispiel: `source_tags=[ 'LOANS', 'CUSTOMERS' ]`

`Beziehungen`	`None`
	`Close`
	`All`
		Abgleichungskriterium zum Abrufen von Beziehungsdetails aus dem Repository. `None` gibt keine Beziehungen zurück.
		`Close` gibt enge Übereinstimmungen abhängig von Details wie dem Abgrenzungsgrad zurück.
		`All` gibt alle möglichen Beziehungen zurück.

`max_degree_separation`	*ganze Zahl*	Minimum 0, Maximum 3.
`output_entity_type`	*Zeichenfolge*	Liste der im Repository verwendeten Entitätstypen.

Eigenschaften von "entityanalytics_processnode"

Der Knoten "Streaming von EA" vergleicht neue Fälle mit den Entitätsdaten im Repository.

Tabelle 16. Eigenschaften von "entityanalytics_processnode"

Eigenschaften von `entityanalytics_processnode`	**Datentyp**	**Eigenschaftsbeschreibung**

`match`	`Exact`
	`ByIdentifier`
	`All`
		Abgleichungskriterium zum Abrufen von Entitäten aus dem Repository. `Exact` gibt nur exakte Übereinstimmungen zurück.
		`ByIdentifier` gibt exakte Übereinstimmungen und Entitäten zurück, die eine gemeinsame ID haben.
		`All` gibt alle möglichen Übereinstimmungen zurück.

`save_search_records`	*boolesch*	

`Beziehungen`	`None`
	`Close`
	`All`
		Abgleichungskriterium zum Abrufen von Beziehungsdetails aus dem Repository. `None` gibt keine Beziehungen zurück.
		`Close` gibt enge Übereinstimmungen abhängig von Details wie dem Abgrenzungsgrad zurück.
		`All` gibt alle möglichen Beziehungen zurück.

`max_degree_separation`	*ganze Zahl*	Minimum 0, Maximum 3.
`output_entity_type`	*Zeichenfolge*	Liste der im Repository verwendeten Entitätstypen.

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 Corporation 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