

*IBM SPSS Modeler 16 - Manuale
dell'utente*

IBM

Nota

Prima di utilizzare queste informazioni e il relativo prodotto, leggere le informazioni in "Note" a pagina 179.

Informazioni sul prodotto

La presente edizione si applica alla versione 16, release 0, modifica 0 di IBM(r) SPSS(r) Modeler e a tutte le release e le modifiche successive fino a quando altrimenti indicato nelle nuove edizioni.

Indice

Capitolo 1. Informazioni su IBM SPSS

Modeler	1
Prodotti IBM SPSS Modeler	1
IBM SPSS Modeler	1
IBM SPSS Modeler Server	1
IBM SPSS Modeler Administration Console	2
IBM SPSS Modeler Batch	2
IBM SPSS Modeler Solution Publisher	2
Adattatori IBM SPSS Modeler Server per IBM SPSS Collaboration and Deployment Services	2
Edizioni di IBM SPSS Modeler	2
Documentazione di IBM SPSS Modeler	3
Documentazione di SPSS Modeler Professional	3
Documentazione di SPSS Modeler Premium	4
Esempi di applicazioni	5
Cartella Demos	5

Capitolo 2. Nuove funzionalità di IBM SPSS Modeler 16. 7

Capitolo 3. Panoramica su IBM SPSS

Modeler	11
Guida Introduttiva	11
Avvio di IBM SPSS Modeler	11
Avvio dalla riga di comando	11
Connessione a IBM SPSS Modeler Server	12
Modifica della directory Temp	14
Avvio di più sessioni di IBM SPSS Modeler.	14
Nozioni di base sull'interfaccia di IBM SPSS Modeler	15
Area del flusso di IBM SPSS Modeler.	15
Palette nodi	16
Manager di IBM SPSS Modeler	17
Progetti IBM SPSS Modeler	19
IBM SPSS Modeler Barra degli strumenti	19
Personalizzazione della barra degli strumenti	20
Personalizzazione della finestra di IBM SPSS Modeler	21
Modifica della dimensione dell'icona di un flusso	22
Utilizzo del mouse in IBM SPSS Modeler	22
Utilizzo dei tasti di scelta rapida	22
Stampa	24
Automazione di IBM SPSS Modeler	24

Capitolo 4. Nozioni sul data mining 25

Panoramica del data mining	25
Valutazione dei dati	26
Strategia per il data mining	27
Modello di elaborazione CRISP-DM	28
Tipi di modelli	29
Esempi di Oracle Data Mining	34

Capitolo 5. Creazione di flussi. 35

Panoramica sulla creazione dei flussi	35
---	----

Creazione dei flussi di dati	35
Utilizzo dei nodi	35
Utilizzo dei flussi	41
Descrizioni flusso	52
Esecuzione dei flussi	54
Utilizzo dei modelli	54
Aggiunta di commenti e annotazioni a nodi e flussi	55
Salvataggio dei flussi di dati.	60
Caricamento di file	62
Mappatura dei flussi di dati	62
Suggerimenti e tasti di scelta rapida	65

Capitolo 6. Gestione dei valori mancanti 67

Panoramica sui valori mancanti	67
Gestione dei valori mancanti	68
Gestione dei record con valori mancanti.	68
Gestione dei campi con valori mancanti	68
Assegnazione o riempimento dei valori mancanti.	69
Funzioni CLEM per i valori mancanti	69

Capitolo 7. Creazione di espressioni CLEM 71

Informazioni su CLEM	71
Esempi CLEM	71
Valori e tipi di dati	73
Espressioni e condizioni	74
Parametri flusso, sessione e Supernodo	74
Utilizzo di stringhe.	75
Gestione di vuoti e valori mancanti	76
Utilizzo di numeri	76
Utilizzo di date e ore	76
Riepilogo di più campi	77
Utilizzo di dati a risposta multipla	78
Generatore di espressioni.	79
Accesso al Generatore di espressioni	79
Creazione di espressioni	79
Selezione di funzioni	79
Selezione di campi, parametri e variabili globali	80
Visualizzazione o selezione di valori	80
Controllo di espressioni CLEM	81
Trova e sostituisci	81

Capitolo 8. Riferimenti al linguaggio CLEM 85

Panoramica di riferimento su CLEM	85
Tipi di dati CLEM	85
Interi	85
Numeri reali	86
Caratteri	86
Stringhe	86
Elenchi	86
Campi	87

Date	87
Ora	88
Operatori CLEM.	88
Riferimenti alle funzioni	90
Convenzioni utilizzate nelle descrizioni delle funzioni	91
Funzioni di informazione	92
Funzioni di conversione	92
Funzioni di confronto	93
Funzioni logiche.	95
Funzioni numeriche	96
Funzioni trigonometriche.	97
Funzioni di probabilità	97
Operazioni bitwise sui numeri interi	98
Funzioni casuali	99
Funzioni stringa	99
Funzioni SoundEx.	104
Funzioni di data e ora	105
Funzioni sequenza.	109
Funzioni globali	113
Funzioni per la gestione di valori vuoti e null	114
Campi speciali	115

**Capitolo 9. Utilizzo di IBM SPSS
Modeler con un repository. 117**

Informazioni su IBM SPSS Collaboration and Deployment Services Repository	117
Archiviazione e deployment di oggetti di repository	118
Connessione al repository	119
Immissione di credenziali per il repository	119
Esplorazione dei contenuti di repository	119
Archiviazione di oggetti in repository	120
Impostazione delle proprietà degli oggetti.	120
Archiviazione di flussi	122
Archiviazione di progetti	122
Archiviazione di nodi	123
Archiviazione di oggetti di output	123
Archiviazione di modelli e palette di modelli	124
Recupero di oggetti da repository	124
Scelta di un oggetto da recuperare	125
Selezione della versione di un oggetto	125
Ricerca di oggetti nel repository	125
Modifica di oggetti nel repository	127
Creazione, ridenominazione ed eliminazione delle cartelle.	127
Blocco e sblocco degli oggetti del repository	127
Eliminazione di oggetti dal repository	127
Gestione delle proprietà degli oggetti di repository	128
Visualizzazione delle proprietà delle cartelle	128
Visualizzazione e modifica delle proprietà degli oggetti.	129
Gestione delle etichette di versione degli oggetti	130
Distribuzione di flussi	130
Opzioni di distribuzione dei flussi	131
Il ramo di calcolo del punteggio	133

**Capitolo 10. Esportazione in
applicazioni esterne 137**

Informazioni sull'esportazione in applicazioni esterne	137
Apertura di un flusso in IBM SPSS Modeler Advantage	137
Importazione ed esportazione di modelli come PMML	138
Tipi di modello che supportano il formato PMML	138

Capitolo 11. Progetti e report. 141

Introduzione ai progetti	141
Visualizzazione CRISP-DM	141
Visualizzazione Classi	142
Creazione di un progetto	142
Creazione di un nuovo progetto	142
Aggiunta di oggetti a un progetto	142
Trasferimento di progetti in IBM SPSS Collaboration and Deployment Services Repository	143
Impostazione delle proprietà del progetto	144
Annotazione di un progetto	144
Proprietà oggetto	145
Chiusura di un progetto.	145
Generazione di report	146
Salvataggio ed esportazione dei report generati	147

**Capitolo 12. Personalizzazione di IBM
SPSS Modeler 149**

Personalizzazione delle opzioni di IBM SPSS Modeler	149
Impostazione delle opzioni di IBM SPSS Modeler	149
Opzioni di sistema	149
Impostazione delle directory di default.	150
Impostazione delle opzioni utente	150
Personalizzazione della palette dei nodi	154
Personalizzazione del Manager palette	154
Modifica della visualizzazione di una scheda della palette.	156
Gestione dei nodi CEMI.	157

**Capitolo 13. Considerazioni sulle
performance dei flussi e dei nodi. 159**

Ordine dei nodi	159
Cache dei nodi	160
Performance: nodi di elaborazione	161
Performance: nodi Modelli	162
Performance: espressioni CLEM	163

**Capitolo 14. Accesso facilitato in IBM
SPSS Modeler 165**

Panoramica delle funzionalità per l'accesso facilitato in IBM SPSS Modeler	165
Tipi di funzionalità per l'accesso facilitato	165
Accesso facilitato per utenti con problemi di vista	165
Accesso facilitato per utenti non vedenti	166
Accesso facilitato mediante tastiera	166

Uso di un lettore di schermo	174
Suggerimenti per l'utilizzo	175
Interferenze con altri programmi	175
JAWS e Java	176
Utilizzo dei grafici in IBM SPSS Modeler	176

Capitolo 15. Supporto Unicode	177
Supporto Unicode in IBM SPSS Modeler	177

Note	179
Marchi	180

Indice analitico.	183
----------------------------------	------------

Capitolo 1. Informazioni su IBM SPSS Modeler

IBM® SPSS Modeler è un insieme di strumenti di data mining che consente di sviluppare rapidamente modelli predittivi con l'ausilio di competenze di business e di eseguirne la distribuzione nelle operazioni di business per migliorare i processi decisionali. Progettato secondo il modello CRISP-DM conforme agli standard di settore, IBM SPSS Modeler supporta l'intero processo di data mining, dai dati a risultati di business migliori.

IBM SPSS Modeler offre numerosi metodi di modellazione ricavati dall'apprendimento automatico, dall'intelligenza artificiale e dalla statistica. I metodi disponibili nella palette Modelli consentono di ricavare nuove informazioni dai dati e di sviluppare modelli predittivi. Ogni metodo ha determinati punti di forza e si presta meglio per particolari tipi di problemi.

SPSS Modeler può essere acquistato come prodotto autonomo oppure utilizzato come client in combinazione con SPSS Modeler Server. È inoltre disponibile una serie di opzioni, come illustrato nelle sezioni seguenti. Per ulteriori informazioni, consultare <http://www.ibm.com/software/analytics/spss/products/modeler/>.

Prodotti IBM SPSS Modeler

La famiglia di prodotti IBM SPSS Modeler e del software associato comprende quanto segue.

- IBM SPSS Modeler
- IBM SPSS Modeler Server
- IBM SPSS Modeler Administration Console
- IBM SPSS Modeler Batch
- IBM SPSS Modeler Solution Publisher
- Adattatori IBM SPSS Modeler Server per IBM SPSS Collaboration and Deployment Services

IBM SPSS Modeler

SPSS Modeler è una versione del prodotto con funzionalità complete che viene installata ed eseguita sul proprio PC. È possibile eseguire SPSS Modeler in modalità locale come prodotto autonomo oppure in modalità distribuita assieme a IBM SPSS Modeler Server per ottenere una migliore performance su insiemi di dati di grandi dimensioni.

Grazie a SPSS Modeler si possono creare, in modo veloce e intuitivo, modelli predittivi accurati senza ricorrere alla programmazione. La sua avanzata interfaccia visiva permette di visualizzare con facilità il processo di data mining. Grazie alle funzionalità di analisi avanzate incorporate nel prodotto, l'utente potrà rilevare la presenza di pattern e tendenze, che altrimenti rimarrebbero occulti, all'interno dei dati. La modellazione dei risultati e la comprensione dei fattori che li influenzano consente di beneficiare di maggiori opportunità di business e, al contempo, di ridurre i rischi.

SPSS Modeler è disponibile in due edizioni: SPSS Modeler Professional e SPSS Modeler Premium. Per ulteriori informazioni, consultare l'argomento "Edizioni di IBM SPSS Modeler" a pagina 2.

IBM SPSS Modeler Server

SPSS Modeler utilizza un'architettura client/server per distribuire le richieste di operazioni che utilizzano molte risorse a potenti componenti software server, con un conseguente miglioramento della performance su insiemi di dati di grandi dimensioni.

SPSS Modeler Server è un prodotto con licenza separata che viene eseguito continuamente in modalità di analisi distribuita su un host server insieme a una o più installazioni IBM SPSS Modeler. Una configurazione di questo tipo consente a SPSS Modeler Server di ottenere prestazioni migliori quando si lavora su insiemi di dati di grandi dimensioni, in quanto le operazioni che richiedono un utilizzo consistente della memoria possono essere eseguite sul server senza scaricare i dati sul computer client. IBM SPSS Modeler Server offre inoltre il supporto delle funzionalità di ottimizzazione SQL e di modellazione nel database, garantendo ulteriori benefici dal punto di vista delle prestazioni e del livello di automazione.

IBM SPSS Modeler Administration Console

Modeler Administration Console è un'applicazione grafica per la gestione di molte delle opzioni di configurazione di SPSS Modeler Server, la cui configurazione può avvenire, inoltre, mediante un file delle opzioni. L'applicazione fornisce un'interfaccia utente di console per monitorare e configurare le installazioni di SPSS Modeler Server ed è disponibile gratuitamente per i clienti esistenti di SPSS Modeler Server. L'applicazione può essere installata solo sui computer Windows; tuttavia, può gestire un server installato su qualsiasi piattaforma supportata.

IBM SPSS Modeler Batch

Nonostante il data mining sia generalmente un processo di tipo interattivo, è possibile eseguire SPSS Modeler da una riga di comando senza il bisogno di ricorrere all'interfaccia utente grafica. Poniamo, ad esempio, che si debbano svolgere varie attività laboriose e ripetitive che non richiedono l'intervento di un utente. SPSS Modeler Batch è una versione speciale del prodotto che supporta l'intera gamma di funzionalità analitiche di SPSS Modeler senza richiedere l'accesso all'interfaccia utente normale. Per utilizzare SPSS Modeler Batch, è necessario disporre di una licenza SPSS Modeler Server.

IBM SPSS Modeler Solution Publisher

SPSS Modeler Solution Publisher è uno strumento che consente di creare una versione a pacchetto di un flusso SPSS Modeler che potrà essere eseguito da un motore di runtime esterno oppure incorporato in una applicazione esterna. Questo permette di pubblicare e sottoporre a deployment stream SPSS Modeler completi in ambienti in cui SPSS Modeler non è installato. SPSS Modeler Solution Publisher è distribuito come parte del servizio IBM SPSS Collaboration and Deployment Services - Scoring, per cui è necessario procurarsi una licenza separata. Insieme alla licenza, si riceve SPSS Modeler Solution Publisher Runtime, che consente di eseguire i flussi pubblicati.

Adattatori IBM SPSS Modeler Server per IBM SPSS Collaboration and Deployment Services

È disponibile una serie di adattatori per IBM SPSS Collaboration and Deployment Services che abilitano l'interazione di SPSS Modeler e SPSS Modeler Server con un repository IBM SPSS Collaboration and Deployment Services. In questo modo, un flusso SPSS Modeler sottoposto a deployment sul repository potrà essere condiviso da più utenti oppure risulterà accessibile dall'applicazione thin client IBM SPSS Modeler Advantage. L'adattatore va installato sul sistema che ospita il repository.

Edizioni di IBM SPSS Modeler

SPSS Modeler è disponibile nelle edizioni seguenti.

SPSS Modeler Professional

SPSS Modeler Professional contiene tutti gli strumenti necessari per utilizzare la maggior parte dei tipi di dati strutturati, quali comportamenti e interazioni registrati in sistemi CRM, dati demografici, dati sulle vendite e sul comportamento d'acquisto.

SPSS Modeler Premium

SPSS Modeler Premium è un prodotto con licenza separata che amplia l'ambito di utilizzo di SPSS Modeler Professional aggiungendo il supporto di dati speciali, quali quelli usati per l'analisi delle entità o dei social network, e di dati di testo non strutturati. SPSS Modeler Premium comprende i seguenti componenti.

IBM SPSS Modeler Entity Analytics aggiunge una dimensione supplementare alle analisi predittive di IBM SPSS Modeler . Se l'analisi predittiva tenta di prevedere il comportamento futuro sulla base di dati precedenti, l'analisi dell'entità si concentra sul miglioramento della coerenza dei dati correnti risolvendo i conflitti tra gli stessi record. Un'identità può essere di un individuo, un'organizzazione, un oggetto o qualsiasi altra entità per cui possa esistere ambiguità. La risoluzione dell'identità può essere essenziale in diversi campi, tra cui la gestione delle relazioni con i clienti, il rilevamento di frodi, il riciclaggio di denaro e la sicurezza nazionale e internazionale.

IBM SPSS Modeler Social Network Analysis trasforma le informazioni sulle relazioni in campi che caratterizzano il comportamento sociale di individui e gruppi. Facendo leva sui dati che descrivono le relazioni esistenti nelle reti sociali, IBM SPSS Modeler Social Network Analysis riesce a individuare i leader in grado di influenzare il comportamento degli altri membri della rete. Consente inoltre di stabilire quali individui della rete sono maggiormente influenzati dagli altri membri. La combinazione di questi risultati ad altre misurazioni permette di delineare profili complessi degli individui su cui basare dei modelli predittivi. I modelli che contengono informazioni sociali generano risultati più accurati rispetto agli altri.

IBM SPSS Modeler Text Analytics utilizza tecnologie linguistiche avanzate e di NLP (Natural Language Processing) per elaborare rapidamente una grande varietà di dati di testo non strutturati, estrarre ed organizzare i concetti chiave e raggruppare tali concetti in categorie. È quindi possibile combinare i concetti e le categorie estratti con dati strutturati esistenti, per esempio dati demografici, e applicarli alla modellazione utilizzando la suite completa degli strumenti di data mining di IBM SPSS Modeler per prendere decisioni migliori e più mirate.

Documentazione di IBM SPSS Modeler

La documentazione nel formato guida in linea è disponibile nel menu Aiuto di SPSS Modeler. Sono incluse la documentazione per SPSS Modeler, SPSS Modeler Server e SPSS Modeler Solution Publisher, nonché la Guida alle applicazioni e altro materiale di supporto.

La documentazione completa in formato PDF dei singoli prodotti, istruzioni di installazione comprese, è disponibile nella cartella *\Documentation* del DVD di ciascun prodotto. I documenti per l'installazione possono anche essere scaricati dal Web, all'indirizzo <http://www-01.ibm.com/support/docview.wss?uid=swg27038316>.

La documentazione in entrambi i formati è inoltre disponibile presso il Centro informazioni SPSS Modeler all'indirizzo <http://publib.boulder.ibm.com/infocenter/spssmodl/v16r0m0/>.

Documentazione di SPSS Modeler Professional

La documentazione completa di SPSS Modeler Professional, escluse le istruzioni di installazione, è la seguente.

- **IBM SPSS Modeler - Guida per l'utente.** Introduzione generale all'utilizzo di SPSS Modeler che illustra come creare flussi di dati, gestire valori mancanti, generare espressioni CLEM, utilizzare progetti e report e assemblare flussi per la distribuzione tramite IBM SPSS Collaboration and Deployment Services, le applicazioni predittive o IBM SPSS Modeler Advantage.
- **IBM SPSS Modeler - Nodi origine, elaborazione e output.** Descrizioni di tutti i nodi utilizzati per leggere, elaborare e generare dati di output in vari formati, ovvero di nodi ad eccezione dei nodi Modelli.

- **IBM SPSS Modeler - Nodi di modellazione.** Descrizioni di tutti i nodi utilizzati per creare modelli data mining. IBM SPSS Modeler offre numerosi metodi di modellazione ricavati dall'apprendimento automatico, dall'intelligenza artificiale e dalla statistica.
- **IBM SPSS Modeler Algorithms Guide.** Descrizione dei fondamenti di matematica per i metodi di modellazione utilizzati in IBM SPSS Modeler. Questa guida è disponibile solo in formato PDF.
- **IBM SPSS Modeler Applications Guide.** Gli esempi inclusi in questa guida forniscono indicazioni mirate e sintetiche su specifici metodi e tecniche di modellazione. Una versione in linea di questa guida è inoltre disponibile dal menu Aiuto. Per ulteriori informazioni, consultare l'argomento "Esempi di applicazioni" a pagina 5.
- **IBM SPSS Modeler - Guida per script e automazione.** Informazioni sulle modalità di automazione del sistema tramite script, incluse le proprietà che è possibile utilizzare per manipolare nodi e flussi.
- **IBM SPSS Modeler - Guida alla distribuzione.** Informazioni sull'esecuzione di flussi e scenari IBM SPSS Modeler come fasi dell'elaborazione di lavori in IBM SPSS Collaboration and Deployment Services Deployment Manager.
- **IBM SPSS Modeler - Guida per lo sviluppatore CLEF.** CLEF consente di integrare programmi di terze parti come routine di elaborazione dei dati o algoritmi di modellazione come nodi in IBM SPSS Modeler.
- **IBM SPSS Modeler - Guida al mining nel database.** Informazioni sulle modalità per utilizzare al meglio la potenza del database in uso al fine di ottenere prestazioni migliori ed estendere la gamma di funzionalità analitiche tramite algoritmi di terze parti.
- **IBM SPSS Modeler Server - Guida della performance e amministrazione.** Informazioni su come configurare e amministrare IBM SPSS Modeler Server.
- **IBM SPSS Modeler Administration Console - Guida per l'utente.** Informazioni sull'installazione e l'utilizzo dell'interfaccia utente della console per il monitoraggio e la configurazione di IBM SPSS Modeler Server. La console viene implementata come plug-in dell'applicazione Deployment Manager.
- **IBM SPSS Modeler - Guida CRISP-DM.** Guida passo a passo al data mining tramite la metodologia CRISP-DM con SPSS Modeler.
- **IBM SPSS Modeler Batch - Guida per l'utente.** Guida completa all'utilizzo di IBM SPSS Modeler in modalità batch, contenente dettagli per l'esecuzione della modalità batch e gli argomenti della riga di comando. Questa guida è disponibile solo in formato PDF.

Documentazione di SPSS Modeler Premium

La documentazione completa di SPSS Modeler Premium, escluse le istruzioni di installazione, è la seguente.

- **IBM SPSS Modeler Entity Analytics User Guide.** Contiene informazioni per l'utilizzo dell'analisi delle entità con SPSS Modeler; descrive l'installazione e la configurazione di repository, i nodi Entity Analytics e le attività amministrative.
- **IBM SPSS Modeler Social Network Analysis User Guide.** Guida che spiega come eseguire l'analisi dei social network con SPSS Modeler; comprende l'analisi di gruppo e l'analisi di diffusione.
- **SPSS Modeler Text Analytics - Guida per l'utente.** Contiene informazioni per l'utilizzo di analisi di testo con SPSS Modeler; descrive i nodi di text mining, il workbench interattivo, i modelli e altre risorse.
- **IBM SPSS Modeler Text Analytics Administration Console - Guida per l'utente.** Informazioni sull'installazione e l'utilizzo dell'interfaccia utente della console per il monitoraggio e la configurazione di IBM SPSS Modeler Server per l'utilizzo con SPSS Modeler Text Analytics. La console viene implementata come plug-in dell'applicazione Deployment Manager.

Esempi di applicazioni

Mentre gli strumenti per il data mining di SPSS Modeler consentono di risolvere un'ampia gamma di problemi a livello di business e organizzativo, gli esempi di applicazioni forniscono indicazioni mirate e sintetiche su specifici metodi e tecniche di modellazione. Gli insiemi di dati utilizzati negli esempi hanno dimensioni molto più limitate rispetto agli enormi archivi di dati gestiti da alcuni data miner, ma i concetti e i metodi coinvolti sono rapportabili alle applicazioni del mondo reale.

È possibile accedere agli esempi facendo clic su **Esempi di applicazioni** nel menu Aiuto di SPSS Modeler. I file di dati e i flussi di esempio sono installati nella cartella *Demos* nella directory di installazione del prodotto. Per ulteriori informazioni, consultare l'argomento "Cartella Demos".

Esempi di modellazione del database. Vedere gli esempi nella *IBM SPSS Modeler Guida al mining nel database*.

Esempi di script. Vedere gli esempi nella *IBM SPSS Modeler Guida per script e automazione*.

Cartella Demos

I file di dati e i flussi di esempio utilizzati negli esempi di applicazioni sono installati nella cartella *Demos* nella directory di installazione del prodotto. È possibile accedere a questa cartella anche dal gruppo di programmi IBM SPSS Modeler nel menu Start di Windows oppure facendo clic su *Demos* nell'elenco delle directory recenti nella finestra di dialogo Apri file.

Capitolo 2. Nuove funzionalità di IBM SPSS Modeler 16

IBM SPSS Modeler aggiunge le seguenti funzioni in questa release.

Simulazione Monte Carlo. Un nuovo nodo di Origine della simulazione consente di generare da zero ed in modo semplice dati sintetici, utilizzando una vasta selezione di distribuzioni statistiche. In alternativa, un nuovo nodo di Adattamento può creare automaticamente un nodo di origine preconfigurato che riflette le distribuzioni delle relazioni tra le variabili cronologiche. Il nodo di Valutazione della simulazione è un nodo terminare progettato per valutare i campi risultanti da un flusso di analisi simulato, e fornisce degli utili grafici di distribuzione e correlazione.

Script Python. Gli script in IBM SPSS Modeler, utilizzati per l'automazione dei processi nell'interfaccia dell'utente, possono ora utilizzare il linguaggio Python, fermo restando il supporto della modalità di script precedente. Python è un linguaggio ben noto e popolare che fornisce una vasta gamma di funzioni, tra le quali una sintassi ricca e concisa, la gestione degli errori e potenti moduli di add-on. **Nota:** La scheda **Script** in **Strumenti > Proprietà flusso** è ora denominata **Esecuzione**.

Esecuzione condizionale e di cicli. Queste nuove opzioni consentono semplici esecuzioni condizionali e di cicli di flussi senza l'utilizzo di script di codice. Queste nuove opzioni sono raggiungibili da **Strumenti > Proprietà flusso > Esecuzioni** o facendo clic con il tasto destro del mouse su un nodo all'interno di un flusso e selezionando l'opzione **Esecuzione ciclica/condizionale**.

Nodo STB (Space-Time-Boxes). Crea bin di dati di ubicazione e data/ora per il supporto di analisi più sofisticate. In modalità hangout, questo nodo identifica i momenti e le posizioni in cui risiedono le entità. Altre funzioni del builder di espressioni supportano l'estrazione di baricentri STB e dello geohashing.

Miglioramenti all'analisi entità. Oltre alla risoluzione di entità individuali ora riesce ad identificare relazioni di grado n tra le entità. Viene fornito ulteriore supporto per l'aggiornamento in tempo reale attraverso il nodo di Streaming, per le entità risolte mediante livellamento (nodo Elimina duplicati), e per l'anonimizzazione dei dati al momento dell'inserimento in un repository di un'entità. **Nota:** il database locale precedente SolidDB è stato sostituito con DB2.

Nuovo tipo di grafico del nodo di Valutazione ROC (Receiver Operating Characteristic) e metriche AUC (Area Under the Curve) e Gini nel nodo di analisi. Supporta obiettivi binari.

Nuova opzioni del nodo Elimina duplicati per la creazione di un record composito. Consente di specificare il metodo di aggregazione di ciascun campo raggruppato (primo valore, ultimo valore, valori concatenati e così via) piuttosto che eliminare record duplicati. Mentre il nodo di aggregazione viene tipicamente utilizzato per il riepilogo dei dati ad un livello più alto, questa nuova opzione viene utilizzata per il livellamento dei duplicati, ad esempio quelli identificati mediante la risoluzione dell'entità.

Nodi di esportazione e origine di TM1. Consente di accedere alle Viste del cubo di TM1 attraverso il nodo di origine TM1, e di assegnare un punteggio ai dati sul nodo TM1 esistente utilizzando il nodo di esportazione TM1.

Espressioni di aggregazione e funzioni di aggregazione della finestra. È possibile creare espressioni di aggregazione personalizzate nel nodo di aggregazione, incorporando le funzioni aggregate integrate (MEAN, SUM e così via) e/o Funzioni definite dall'utente aggregate del database. Nelle espressioni derivate è possibile derivare i campi che richiedono funzioni di aggregazione a finestra (come ad esempio le medie mobili). Sono disponibili funzioni aggregate a finestra incorporate e fornite dal database.

Miglioramenti al mining nel database di IBM Netezza Analytics. Nuovo algoritmo Two step di Netezza, ulteriori opzioni dell'applicazione di supporto per gestire (eliminare, rinominare e così via) i modelli analitici di Netezza, supporto per i visualizzatori del modello per le strutture ad albero di regressione, per le strutture ad albero delle decisioni, per le K-medie e per TwoStep. **Nota:** Le nuove funzioni richiedono INZA 3.0.

Nuovi nodi R e Builder di finestre di dialogo personalizzate per R. In aggiunta al nugget del modello ed al Nodo di costruzione modello R introdotti in SPSS Modeler 15 Fix Pack 2, questa release aggiunge 2 nuovi nodi - Elaborazione R e Output R. Con il nodo Elaborazione R, è possibile prelevare i dati da un flusso SPSS Modeler ed applicarvi le trasformazioni utilizzando lo script R. Con il nodo Output R, è possibile utilizzare i propri script R personalizzati per eseguire l'analisi dei dati e per riepilogare i risultati del calcolo del punteggio del modello.

È possibile produrre un output grafico e di testo delle proprie analisi. È quindi possibile reindirizzarlo ad un file o visualizzarlo nel browser dell'output del nodo Output R. Il Builder di finestre di dialogo personalizzate consente di definire Creazioni del modello personalizzate, tipi di nodi di Output e di Elaborazione e nugget del modello, compresi selettori di campo, testo, numeri, pulsanti di opzioni, e controlli delle finestre di dialogo secondari per abilitare l'astrazione e la parametrizzazione dei programmi R. È possibile selezionare il tipo di nodo, la tavolozza di destinazione e l'icona del nodo prima dell'installazione o della condivisione dello stesso affinché possa essere utilizzato da altri utenti di SPSS Modeler. Il Builder di finestre di dialogo personalizzate viene avviato dal menu Strumenti. **Nota:** Per utilizzare questa funzione, è necessario installare SPSS Modeler - Essentials for R.

R nel database. Supporto dei push back SQL dei nodi R, per Netezza, SAP Hana e Oracle utilizzando il relativo supporto R. **Nota:** I database devono avere installate le appropriate estensioni R del fornitore.

Nodo elaborazione delle serie temporali del flusso. Crea ed assegna il punteggio ai modelli di serie temporali in un'unica fase per consentire la distribuzione in tempo reale mediante IBM InfoSphere Streams, il Servizio di calcolo del punteggio di IBM SPSS Collaboration and Deployment Services o IBM SPSS Modeler Solution Publisher.

Pulsante Anteprima. Quando utilizzato in combinazione con un'origine database l'anteprima abilita la visualizzazione dei push back SQL. Quando si utilizza in combinazione con l'origine dati Analytic Server, utilizzare il pulsante Anteprima per ricevere informazioni su eventuali spostamenti di grandi quantità di dati.

Nuove opzioni di Analytic Server nei nodi Classificatore automatico, Numerico automatico e Cluster automatico. Quando si esegue un flusso su IBM SPSS Analytic Server, è possibile decidere tra un'esecuzione con le Suddivisioni abilitate (se si utilizza la funzione modello di suddivisione, utilizzare questa opzione) o l'opzione Dati di grandi dimensioni (le suddivisioni vengono ignorate e gli obiettivi di modellazione vengono impostati per la gestione di Dati di grandi dimensioni). Il calcolo del punteggio è ora per i modelli Classificatore automatico, Numerico automatico e Cluster automatico che possono essere integrati in SPSS Modeler Server.

Supporto migliorato dell'adattatore di calcolo del punteggio. Supporto per il calcolo del punteggio dei modelli di Text Mining mediante Adattatori di calcolo del punteggio del database. Sono presenti nuovi adattatori di calcolo del punteggio del database per DB2 LUW e Oracle.

Nodo origine della Vista dati. Il nodo Vista dati consente di leggere i dati da una Vista dati analitica in un flusso SPSS Modeler. La Vista dati analitica rappresenta un nuovo modo di creare una Vista dati unificata in IBM SPSS Collaboration and Deployment Services 6.

Opzioni varie ed altre migliorie.

- Suddivisioni in sezioni della query Teradata delle preimpostazioni della connessione al Database.
- Opzioni del nodo GLMM aggiornate.

- Supporto per file .sav compressi e crittografati.
- Supporto per il SSO (single sign-on) di SPSS Modeler Server senza IBM SPSS Collaboration and Deployment Services.
- Supporto a livello database del SSO (Single Sign-On).
- Pushback SQL per il nodo campione per zDB2.
- I nodi R vengono ora installati come parte di un'installazione di base di SPSS Modeler.
- Gli adattatori Modeler per IBM SPSS Collaboration and Deployment Services vengono installati mediante IBM Installation Manager.
- Aggiornamento del modello di suddivisione.
- Supporto della localizzazione in Russo.

Capitolo 3. Panoramica su IBM SPSS Modeler

Guida Introduttiva

IBM SPSS Modeler è un'applicazione di data mining che offre un approccio strategico al rilevamento di relazioni utili all'interno di insiemi di dati di grandi dimensioni. Rispetto ai metodi statistici più tradizionali, non è necessario conoscere fin dall'inizio che cosa cercare esattamente. È possibile esplorare i dati, adattarvi modelli diversi e analizzare le diverse relazioni, fino a individuare informazioni utili.

Avvio di IBM SPSS Modeler

Per avviare l'applicazione, fare clic su

Start > [Tutti i] Programmi > IBM SPSS Modeler 16 > IBM SPSS Modeler 16

Dopo qualche secondo viene visualizzata la finestra principale.

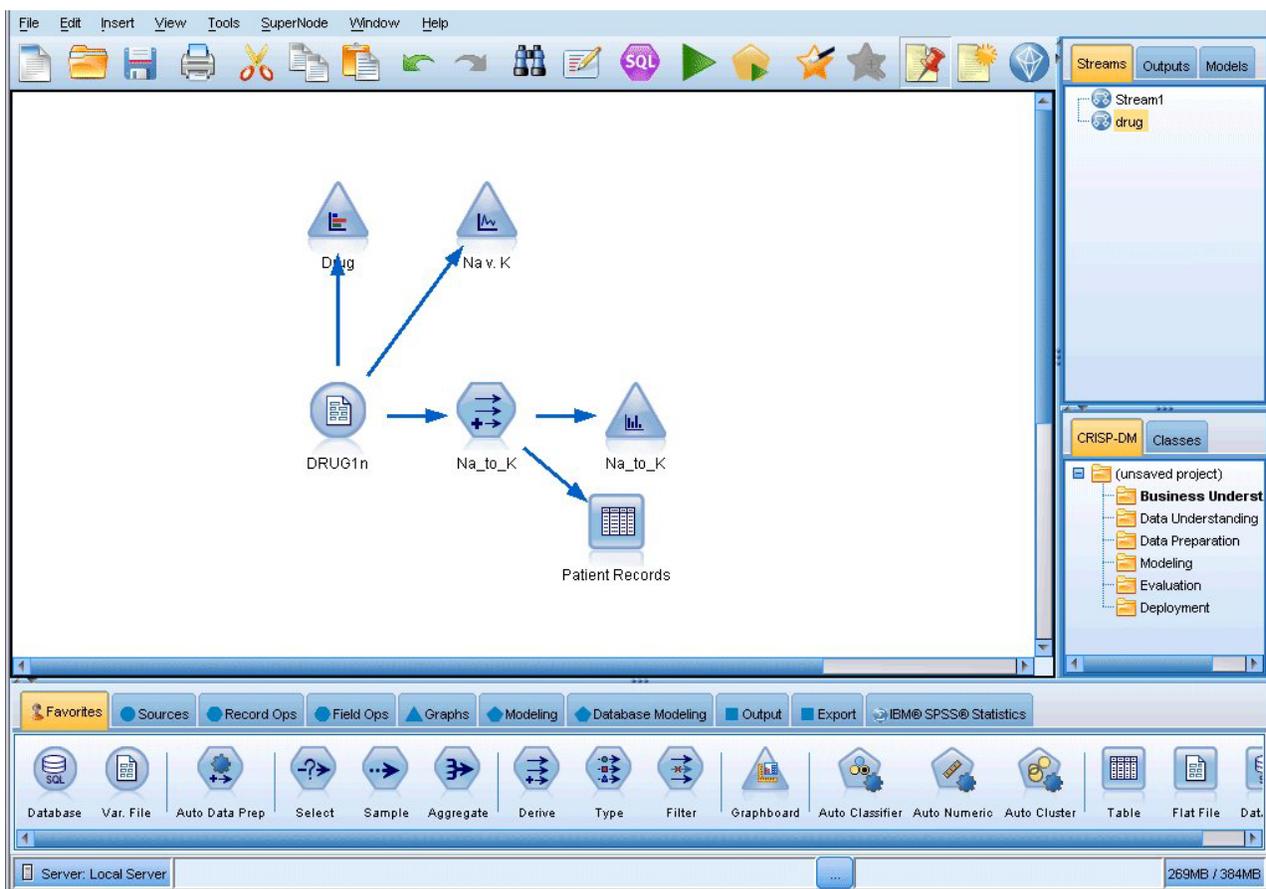


Figura 1. Finestra principale di IBM SPSS Modeler

Avvio dalla riga di comando

È possibile utilizzare la riga di comando del sistema operativo per avviare IBM SPSS Modeler:

1. Sul computer in cui è installato IBM SPSS Modeler, aprire una finestra DOS (prompt dei comandi).

2. Per avviare l'interfaccia di IBM SPSS Modeler in modalità interattiva, digitare il comando `modelerclient` seguito dagli argomenti richiesti, ad esempio:

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

Gli argomenti disponibili (flag) consentono di connettersi a un server, caricare stream, eseguire script o specificare altri parametri.

Connessione a IBM SPSS Modeler Server

È possibile eseguire IBM SPSS Modeler come applicazione autonoma oppure come client connesso direttamente a IBM SPSS Modeler Server o a un cluster di server tramite il plug-in Coordinator of Processes disponibile in IBM SPSS Collaboration and Deployment Services. Lo stato di connessione corrente è visualizzato nella parte inferiore sinistra della finestra di IBM SPSS Modeler.

Per connettersi a un server, è possibile immettere manualmente il nome del server al quale connettersi oppure selezionare un nome definito in precedenza. Tuttavia, se si dispone di IBM SPSS Collaboration and Deployment Services, è possibile cercare nell'elenco di server o cluster di server disponibili nella finestra di dialogo Login server. Il plug-in Coordinator of Processes consente di spostarsi tra i servizi Statistics in esecuzione in una rete.

Per connettersi a un server

1. Nel menu Strumenti, fare clic su **Login server**. Viene visualizzata la finestra di dialogo Login server. In alternativa, fare doppio clic sull'area dello stato di connessione nella finestra di IBM SPSS Modeler.
2. Nella finestra di dialogo specificare le opzioni per la connessione al server locale oppure selezionare una connessione dalla tabella.
 - Fare clic su **Aggiungi** o **Modifica** per aggiungere o modificare una connessione. Consultare l'argomento "Aggiunta e modifica della connessione di IBM SPSS Modeler Server" a pagina 13 per ulteriori informazioni.
 - Fare clic su **Cerca** per accedere al server o al cluster di server in Coordinator of Processes. Consultare l'argomento "Ricerca di server in IBM SPSS Collaboration and Deployment Services" a pagina 13 per ulteriori informazioni.

Tabella server. Questa tabella contiene l'insieme di connessioni server definite. Nella tabella sono visualizzati la connessione di default, il nome del server, la descrizione e il numero di porta. È possibile aggiungere manualmente una nuova connessione, nonché selezionare o cercare una connessione esistente. Per impostare un determinato server come connessione di default, selezionare la casella di controllo nella colonna Default della tabella per la connessione desiderata.

Percorso dati di default. Specificare il percorso utilizzato per i dati nel server. Fare clic sul pulsante con i puntini di sospensione (...) per passare alla posizione richiesta.

Imposta credenziali. Lasciare questa casella deselezionata per abilitare la funzione di **Single Sign-On (SSO)**, che cercherà di far accedere l'utente al server utilizzando il nome utente e la password del computer locale. Se la funzione di Single Sign-On (SSO) non è disponibile oppure se si seleziona la casella per disabilitare Single Sign-On (SSO) (per esempio, per accedere a un account amministratore), vengono attivati i seguenti campi per l'inserimento delle credenziali.

ID utente. Immettere il nome utente con il quale accedere al server.

Password. Immettere la password associata al nome utente specificato.

Dominio. Specificare il dominio utilizzato per accedere al server. È necessario specificare un nome di dominio solo quando il computer server risiede in un dominio Windows diverso da quello del computer client.

3. Fare clic su **OK** per completare la connessione.

Per disconnettersi da un server

1. Nel menu Strumenti, fare clic su **Login server**. Viene visualizzata la finestra di dialogo Login server.
In alternativa, fare doppio clic sull'area dello stato di connessione nella finestra di IBM SPSS Modeler.
2. Nella finestra di dialogo, selezionare il Server locale e fare clic su **OK**.

Aggiunta e modifica della connessione di IBM SPSS Modeler Server

È possibile modificare o aggiungere manualmente una connessione server nella finestra di dialogo Login server. Facendo clic su **Aggiungi**, è possibile accedere a una finestra di dialogo **Aggiungi/Modifica server** vuota in cui è possibile immettere i dettagli di connessione al server. Dopo aver selezionato una connessione esistente e aver fatto clic su **Modifica** nella finestra di dialogo Login server, viene visualizzata la finestra di dialogo **Aggiungi/Modifica server** con i dettagli della connessione da modificare.

Nota: non è possibile modificare una connessione server che è stata aggiunta da IBM SPSS Collaboration and Deployment Services, poiché il nome, la porta e altri dettagli sono definiti in IBM SPSS Collaboration and Deployment Services.

Per aggiungere connessioni server

1. Nel menu Strumenti, fare clic su **Login server**. Viene visualizzata la finestra di dialogo Login server.
2. Nella finestra di dialogo, fare clic su **Aggiungi**. Viene visualizzata la finestra di dialogo Login server: **Aggiungi/Modifica server**.
3. Immettere i dettagli della connessione server e fare clic su **OK** per salvare la connessione e ritornare alla finestra di dialogo Login server.
 - **Server.** Specificare uno dei server disponibili o selezionarne uno dall'elenco. È possibile identificare il computer server tramite un nome alfanumerico, per esempio *serverpersonale*, oppure tramite l'indirizzo IP assegnato al computer server, per esempio 202.123.456.78.
 - **Porta.** Indica il numero di porta su cui il server rimane in ascolto. Se l'impostazione di default non funziona, chiedere il numero di porta corretto all'amministratore di sistema.
 - **Descrizione.** Immettere una descrizione facoltativa della connessione al server.
 - **Verificare la connessione sicura (utilizza SSL).** Specifica se è necessario utilizzare una connessione SSL (**Secure Sockets Layer**). SSL è un protocollo molto diffuso per la gestione della sicurezza dei dati trasmessi in rete. Per utilizzare questa funzione è necessario che il protocollo SSL sia stato attivato sul server di hosting di IBM SPSS Modeler Server. Se sono necessarie ulteriori informazioni, rivolgersi all'amministratore di sistema.

Per modificare connessioni server

1. Nel menu Strumenti, fare clic su **Login server**. Viene visualizzata la finestra di dialogo Login server.
2. Nella finestra di dialogo, selezionare la connessione da modificare, quindi fare clic su **Modifica**. Viene visualizzata la finestra di dialogo Login server: **Aggiungi/Modifica server**.
3. Modificare i dettagli della connessione server e fare clic su **OK** per salvare le modifiche e ritornare alla finestra di dialogo Login server.

Ricerca di server in IBM SPSS Collaboration and Deployment Services

Anziché immettere una connessione server manualmente, è possibile selezionare un server o un cluster di server della rete tramite Coordinator of Processes, disponibile in IBM SPSS Collaboration and Deployment Services. Un cluster di server è un gruppo di server tra i quali Coordinator of Processes stabilisce quello più adatto a rispondere a una richiesta di elaborazione.

Sebbene nella finestra di dialogo Login server sia possibile aggiungere server manualmente, cercando i server disponibili è possibile connettersi agli stessi senza doverne conoscere il nome e il numero di porta. Queste informazioni sono infatti fornite automaticamente. È in ogni caso necessario disporre delle informazioni di accesso corrette, quali nome utente, dominio e password.

Nota: se non si dispone della funzionalità Coordinator of Processes, è comunque possibile immettere manualmente il nome del server a cui connettersi oppure selezionare un nome definito in precedenza. Consultare l'argomento "Aggiunta e modifica della connessione di IBM SPSS Modeler Server" a pagina 13 per ulteriori informazioni.

Per cercare server e cluster

1. Nel menu Strumenti, fare clic su **Login server**. Viene visualizzata la finestra di dialogo Login server.
2. Nella finestra di dialogo, fare clic su **Cerca** per aprire la finestra di dialogo Cerca server. Se non si è ancora connessi a IBM SPSS Collaboration and Deployment Services quando si tenta di sfogliare Coordinator of Processes, viene richiesto di connettersi.
3. Selezionare un server o un cluster di server dall'elenco.
4. Fare clic su **OK** per chiudere la finestra di dialogo e aggiungere la connessione alla tabella nella finestra di dialogo Login server.

Modifica della directory Temp

Alcune operazioni eseguite da IBM SPSS Modeler Server potrebbero richiedere la creazione di file temporanei. Per default, per la creazione di file temporanei in IBM SPSS Modeler viene utilizzata la directory temporanea di sistema. È possibile modificare il percorso della directory temporanea, attenendosi alla procedura seguente.

1. Creare una nuova directory con nome *spss* e una sottodirectory con nome *servertemp*.
2. Modificare il file *options.cfg*, disponibile nella directory */config* dell'installazione di IBM SPSS Modeler in uso. Modificare il parametro *temp_directory* in tale file come segue: *temp_directory*, "C:/spss/servertemp".
3. Dopo avere eseguito tale operazione, è necessario riavviare il servizio IBM SPSS Modeler Server, facendo clic sulla scheda **Servizi** nel Pannello di controllo di Windows. Arrestare il servizio, quindi avviarlo per attivare le modifiche apportate. È possibile riavviare il servizio anche riavviando il computer.

Tutti i file temporanei verranno quindi scritti nella nuova directory.

Nota: l'errore più comune in questo tipo di operazione è l'utilizzo del tipo di barra non corretto, utilizzare la barra normale.

Avvio di più sessioni di IBM SPSS Modeler

Per avviare più sessioni di IBM SPSS Modeler contemporaneamente occorre apportare alcune modifiche alle impostazioni di IBM SPSS Modeler e di Windows. Questo può essere necessario, per esempio, se si dispone di due licenze server separate e si desidera eseguire due stream su due server diversi dallo stesso computer client.

Per attivare sessioni multiple di IBM SPSS Modeler:

1. Fare clic su
Start > [Tutti i] Programmi > IBM SPSS Modeler 16
2. Con il pulsante destro del mouse, fare clic sul collegamento IBM SPSS Modeler 16 (quello con l'icona) e selezionare **Proprietà**.
3. Nella casella di testo **Destinazione**, aggiungere *-noshare* alla fine della stringa.
4. In Esplora risorse, selezionare:
Strumenti > Opzioni cartella...
5. Nella scheda Tipi di file, selezionare l'opzione Flusso di IBM SPSS Modeler e fare clic su **Avanzate**.
6. Nella finestra di dialogo Modifica tipo file, selezionare Apri con IBM SPSS Modeler e fare clic su **Edit**.
7. Nella casella di testo **Applicazione utilizzata per eseguire l'operazione**, aggiungere *-noshare* prima dell'argomento **-stream**.

Nozioni di base sull'interfaccia di IBM SPSS Modeler

In ogni fase del processo di data mining, l'interfaccia visiva di IBM SPSS Modeler sollecita le capacità di business specifiche dell'utente. Gli algoritmi di creazione di modelli, quali quelli di previsione, classificazione, segmentazione e individuazione delle associazioni, assicurano modelli potenti e precisi. I modelli possono essere facilmente implementati e letti nei database, nell'IBM SPSS Statistics e in molte altre applicazioni.

Il processo di elaborazione dei dati in IBM SPSS Modeler è articolato in tre passaggi.

- In primo luogo, si leggono i dati in IBM SPSS Modeler.
- Quindi li si sottopone a una serie di manipolazioni.
- Infine, li si invia a una destinazione.

Questa sequenza di operazioni è nota come **flusso di dati** perché i dati fluiscono, record dopo record, dalla sorgente attraverso ogni manipolazione e infine arrivano alla destinazione, che può essere un modello o un tipo di output di dati.

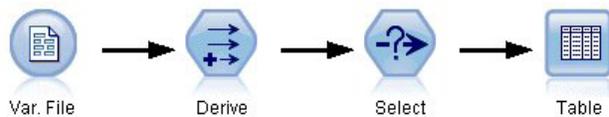


Figura 2. Un flusso semplice

Area del flusso di IBM SPSS Modeler

L'area del flusso è l'area di maggiori dimensioni della finestra di IBM SPSS Modeler. In quest'area vengono costruiti e manipolati i flussi di dati.

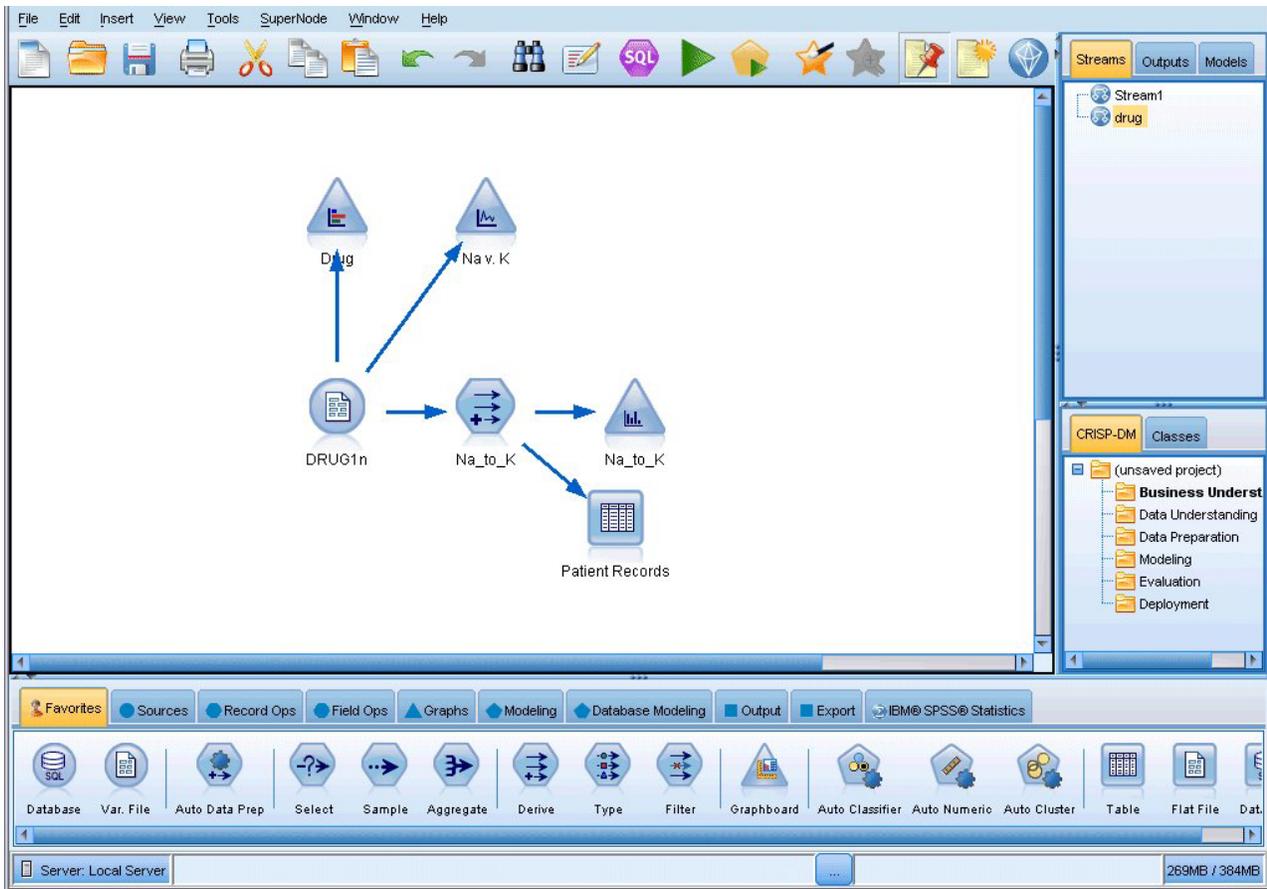


Figura 3. Spazio di lavoro di IBM SPSS Modeler (visualizzazione di default)

I flussi vengono creati disegnando diagrammi delle operazioni sui dati pertinenti per la propria azienda nell'area principale dell'interfaccia. Ogni operazione è rappresentata da un'icona o **nodo** e i nodi sono collegati insieme in uno **stream** che rappresenta il flusso dei dati attraverso ciascuna operazione.

Con IBM SPSS Modeler è possibile utilizzare più flussi contemporaneamente, nella stessa area o aprendo una nuova area del flusso. Nel corso di una sessione, i flussi vengono archiviati nel relativo Manager, nella parte superiore destra della finestra di IBM SPSS Modeler.

Palette nodi

La maggior parte dei dati e degli strumenti di modellazione di IBM SPSS Modeler si trovano nella **palette dei nodi**, visualizzata nella sezione sottostante l'area del flusso.

Per esempio, la scheda Oper su record della palette contiene i nodi utilizzabili per eseguire operazioni sui **record** di dati quali la selezione, l'unione e l'accodamento.

Per aggiungere nodi all'area, fare doppio clic sulle icone dalla palette dei nodi oppure trascinare e rilasciare le icone nell'area. Connettere quindi le icone per creare uno **stream** che rappresenti il flusso di dati.



Figura 4. Scheda Oper su record della tavolozza dei nodi

Ogni scheda della palette contiene una raccolta di nodi correlati che viene utilizzata per le diverse fasi delle operazioni dei flussi, per esempio:

- **Origini.** Nodi che consentono di inserire i dati in IBM SPSS Modeler.
- **Operazioni record.** Nodi utilizzati per le operazioni sui **record** di dati, per esempio la selezione, l'unione e l'accodamento.
- **Operazioni campi.** Nodi utilizzati per le operazioni sui **campi** di dati, per esempio l'applicazione di filtri, la derivazione di nuovi campi e la definizione del livello di misurazione per campi specifici.
- **Grafici.** Nodi utilizzati per rappresentare graficamente i dati prima e dopo la fase di modellazione. I grafici includono i plot, gli istogrammi, i nodi Web e i grafici di valutazione.
- **Modellazione.** Nodi che utilizzano gli algoritmi di creazione di modelli disponibili in IBM SPSS Modeler, per esempio le reti neurali, le strutture ad albero delle decisioni, gli algoritmi cluster e la sequenzializzazione dei dati.
- **Modellazione del database.** Nodi che utilizzano gli algoritmi di modellazione disponibili nei database di Microsoft SQL Server, IBM DB2, Oracle e Netezza.
- **Output.** Nodi che generano output di svariati tipi per dati, grafici e risultati di modelli, visualizzabili in IBM SPSS Modeler.
- **Esporta.** Nodi che generano vari tipi di output, visualizzabili in applicazioni esterne quali IBM SPSS Data Collection o Excel.
- **IBM SPSS Statistics.** Nodi utilizzati per importare o esportare dati da IBM SPSS Statistics e per eseguire le procedure di IBM SPSS Statistics.

Una volta acquisita una certa dimestichezza con le funzioni di IBM SPSS Modeler, sarà possibile personalizzare il contenuto della palette a piacere.

Il riquadro Report, posto sotto la palette dei nodi, consente di controllare l'andamento delle varie operazioni, per esempio in fase di lettura dei dati nel flusso di dati. Anche il riquadro dello stato si trova sotto la palette dei nodi. Esso fornisce informazioni sulle operazioni in corso nell'applicazione e offre indicazioni qualora sia richiesto il feedback dell'utente.

Manager di IBM SPSS Modeler

Il riquadro dei manager si trova nella parte superiore destra della finestra ed è formato da tre schede, utilizzate per gestire i flussi, l'output e i modelli.

La scheda Flussi può essere utilizzata per aprire, rinominare, salvare ed eliminare i flussi creati in una sessione.



Figura 5. Scheda Flussi



Figura 6. Scheda Output

La scheda Output contiene tutti i file creati (quali grafici e tabelle) tramite le operazioni su stream in IBM SPSS Modeler. I grafici, le tabelle e i report elencati in questa scheda possono essere visualizzati, salvati, rinominati e chiusi.



Figura 7. Scheda Modelli contenente dei nugget del modello

La scheda Modelli è la scheda più importante della finestra dei manager. Essa contiene tutti i **nugget del modello**, che contengono i modelli generati in IBM SPSS Modeler per la sessione corrente. I modelli possono essere visualizzati direttamente dalla scheda Modelli oppure possono essere aggiunti al flusso nell'area.

Progetti IBM SPSS Modeler

Nella parte inferiore destra della finestra viene visualizzato il riquadro dei progetti, utilizzato per creare e gestire **progetti** di data mining (gruppi di file relativi a un'attività di data mining). Sono disponibili due modalità di visualizzazione per i progetti creati in IBM SPSS Modeler: le viste Classi e CRISP-DM.



Figura 8. Vista CRISP-DM

La scheda CRISP-DM consente di organizzare i progetti in base allo standard CRISP-DM (Cross-Industry Standard Process for Data Mining), una metodologia del settore consolidata e non proprietaria. L'utilizzo dello strumento CRISP-DM consente ai data miner esperti e principianti di organizzare e comunicare al meglio le operazioni eseguite.

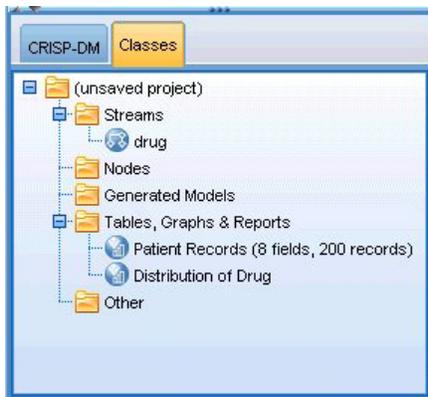


Figura 9. Vista Classi

La scheda Classi consente di organizzare il proprio lavoro in IBM SPSS Modeler per categorie, in base ai tipi di oggetti creati. Si tratta di una visualizzazione utile per l'inventario di dati, stream e modelli.

IBM SPSS Modeler Barra degli strumenti

Nella parte superiore della finestra di IBM SPSS Modeler vi è una barra degli strumenti a icone che offre diverse funzioni utili. Di seguito vengono illustrati i pulsanti della barra degli strumenti e le relative funzioni.



Crea un nuovo stream



Apri flusso

	Salva stream		Stampa stream corrente
	Taglia e sposta negli Appunti		Copia negli Appunti
	Incollare la selezione		Annulla l'ultima operazione
	Ripeti		Cerca nodi
	Modifica proprietà del flusso		Anteprima generazione SQL
	Esegui stream corrente		Esegui selezione di stream
	Interrompi stream (attivato solo durante l'esecuzione del flusso)		Aggiunge un Supernodo
	Zoom avanti (solo Supernodi)		Zoom indietro (solo Supernodi)
	No markup in stream		Inserisci commento
	Nascondi markup stream (se presente)		Mostra markup del flusso nascosto
	Apri flusso in IBM SPSS Modeler Advantage		

Il markup del flusso comprende i commenti ai flussi, i collegamenti dei modelli e l'indicazione dei rami di calcolo del punteggio.

I collegamenti dei modelli sono descritti nel manuale *Nodi Modelli in IBM SPSS*.

Personalizzazione della barra degli strumenti

È possibile modificare vari aspetti della barra degli strumenti, per esempio:

- Se è visualizzata o meno
- Se viene visualizzato il testo di suggerimento delle icone
- Se utilizzare le icone grandi o piccole

Per attivare o disattivare la visualizzazione della barra degli strumenti:

1. Nel menu principale, fare clic su:
Visualizza > Barra degli strumenti > Visualizzazione

Per modificare l'impostazione del testo di suggerimento o delle dimensioni delle icone:

1. Nel menu principale, fare clic su:
Visualizza > Barra degli strumenti > Personalizza

Fare clic su **Mostra descrizioni** o **Pulsanti grandi** a seconda dei casi.

Personalizzazione della finestra di IBM SPSS Modeler

I divisori tra le varie parti dell'interfaccia di IBM SPSS Modeler consentono di ridimensionare o chiudere a piacere gli strumenti. Per esempio, per un flusso di grandi dimensioni è possibile utilizzare le piccole frecce poste su ciascun divisore per chiudere la palette dei nodi, il riquadro dei manager e il riquadro dei progetti. In questo modo l'area viene ingrandita e consente di utilizzare più flussi o stream di grandi dimensioni.

In alternativa, nel menu **Visualizza** fare clic su **Palette nodi**, **Manager** o **Progetto** per attivare o disattivare la visualizzazione di questi elementi.

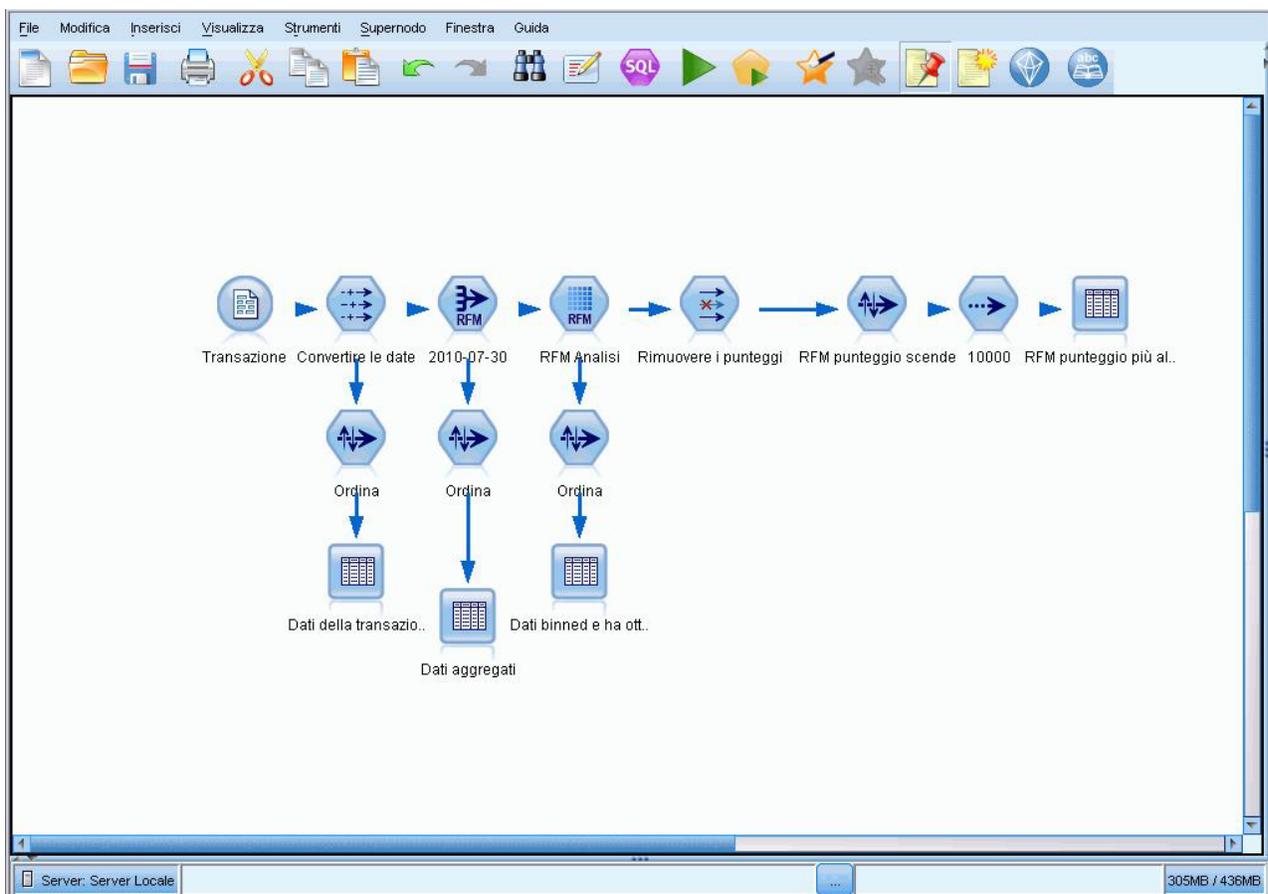


Figura 10. Area del flusso ingrandita

In alternativa alla chiusura della palette dei nodi e dei riquadri dei manager e dei progetti è possibile utilizzare l'area come pagina scorrevole spostandosi in verticale e in orizzontale tramite le barre di scorrimento poste di fianco e in fondo alla finestra di IBM SPSS Modeler.

È possibile anche controllare la visualizzazione del markup, costituito dai commenti ai flussi, dai collegamenti dei modelli e dall'indicazione dei rami di calcolo del punteggio. Per attivare o disattivare la visualizzazione, fare clic su:

Visualizza > Markup flusso

Modifica della dimensione dell'icona di un flusso

È possibile modificare le dimensioni delle icone dei flussi nei modi seguenti.

- Attraverso l'impostazione di una proprietà del flusso
- Attraverso un menu di scelta rapida del flusso
- Utilizzando la tastiera

È possibile ridimensionare l'intera visualizzazione del flusso su una scala compresa tra l'8% e il 200% della dimensione dell'icona standard.

Per ridimensionare l'intero stream (attraverso le proprietà del flusso)

1. Dal menu principale, scegliere **Strumenti > Proprietà flusso > Opzioni > Layout**.
2. Scegliere la dimensione desiderata dal menu Dimensione icona.
3. Fare clic su **Applica** per visualizzare il risultato.
4. Fare clic su **OK** per salvare la modifica.

Per ridimensionare l'intero stream (attraverso il menu)

1. Fare clic con il pulsante destro del mouse sullo sfondo del flusso nell'area.
2. Scegliere **Dimensione icona** e selezionare la dimensione desiderata.

Per ridimensionare l'intero stream (attraverso la tastiera)

1. Premere Ctrl + [-] sulla tastiera principale per eseguire lo zoom indietro alla successiva dimensione più piccola.
2. Premere Ctrl + Maiusc + [+] sulla tastiera principale per eseguire lo zoom avanti alla successiva dimensione più grande.

Questa funzione risulta particolarmente utile per ottenere una vista generale di un flusso complesso. La funzione consente inoltre di ridurre al minimo il numero di pagine necessario per la stampa di un flusso.

Utilizzo del mouse in IBM SPSS Modeler

Di seguito sono riportati gli utilizzi più comuni del mouse in IBM SPSS Modeler:

- **Clic singolo.** Utilizzare il pulsante destro o sinistro del mouse per scegliere opzioni dai menu, aprire menu di scelta rapida e accedere a numerosi altri controlli e opzioni standard. Fare clic e tenere premuto il pulsante per spostare e trascinare i nodi.
- **Doppio clic.** Fare doppio clic con il pulsante sinistro del mouse per posizionare i nodi nell'area del flusso e modificare i nodi esistenti.
- **Clic con il pulsante centrale.** Fare clic con il pulsante centrale del mouse e trascinare il cursore per connettere i nodi nell'area del flusso. Fare doppio clic con il pulsante centrale del mouse per disconnettere un nodo. Se non si dispone di un mouse con tre pulsanti, è possibile simulare questa funzionalità tenendo premuto Alt mentre si seleziona e si trascina l'elemento desiderato.

Utilizzo dei tasti di scelta rapida

In IBM SPSS Modeler, numerose operazioni di programmazione visuale sono associate a tasti di scelta rapida. Per esempio, per eliminare un nodo è possibile fare clic su di esso e premere il tasto Canc.

Analogamente, per salvare rapidamente un flusso tenere premuto il tasto Ctrl e premere contemporaneamente il tasto S. I comandi di controllo di questo tipo sono indicati da una combinazione di Ctrl più un altro tasto, per esempio Ctrl+S.

Nelle operazioni standard di Windows sono utilizzati diversi tasti di scelta rapida, per esempio Ctrl+X per tagliare. Questi tasti di scelta rapida sono supportati in IBM SPSS Modeler insieme ai tasti di scelta rapida seguenti specifici per l'applicazione.

Nota: in alcuni casi, i precedenti tasti di scelta rapida utilizzati in IBM SPSS Modeler sono in conflitto con quelli standard di Windows. I precedenti tasti di scelta rapida sono supportati con l'aggiunta del tasto Alt. Per esempio, Ctrl+Alt+C può essere utilizzato per attivare/disattivare la cache.

Tabella 1. Tasti di scelta rapida supportati

Tasto di scelta rapida	Funzione
Ctrl+A	Seleziona tutto
Ctrl+X	Taglia
Ctrl+N	Nuovo stream
Ctrl+O	Apri flusso
Ctrl+P	Stampa
Ctrl+C	Copia
Ctrl+V	Incolla
Ctrl+Z	Annulla
Ctrl+Q	Seleziona tutti i nodi downstream del nodo selezionato
Ctrl+W	Deseleziona tutti i nodi downstream del nodo selezionato (si alterna con Ctrl+Q)
Ctrl+E	Esegui a partire dal nodo selezionato
Ctrl+S	Salva stream corrente
Alt+tasti freccia	Sposta i nodi selezionati nell'area del flusso in direzione della freccia utilizzata
Maiusc+F10	Apri il menu di scelta rapida per il nodo selezionato

Tabella 2. Scelte rapide supportate per i precedenti tasti di scelta rapida

Tasto di scelta rapida	Funzione
Ctrl+Alt+D	Duplica il nodo
Ctrl+Alt+L	Carica nodo
Ctrl+Alt+R	Rinomina il nodo
Ctrl+Alt+U	Crea il nodo input utente
Ctrl+Alt+C	Attiva/disattiva la cache
Ctrl+Alt+F	Svuota cache
Ctrl+Alt+X	Espandi Supernodo
Ctrl+Alt+Z	Zoom avanti/indietro
Canc	Elimina il nodo o la connessione

Stampa

In IBM SPSS Modeler è possibile stampare i seguenti oggetti:

- Diagrammi di flusso
- Grafici
- Tabelle
- Report (dal nodo Report e dai report del progetto)
- Script (dalle finestre di dialogo Proprietà stream, Script autonomo o Script Supernodo)
- Modelli (browser Modello, schede di finestre di dialogo con stato attivo corrente, visualizzatori della struttura ad albero)
- Annotazioni (utilizzando la scheda Annotazioni per l'output)

Per stampare un oggetto:

- Per stampare senza anteprima, fare clic sul pulsante Stampa sulla barra degli strumenti.
- Per impostare la pagina prima della stampa, scegliere **Imposta pagina** dal menu File.
- Per richiamare l'anteprima prima della stampa, scegliere **Anteprima di stampa** dal menu File.
- Per visualizzare la finestra di dialogo di stampa standard con le opzioni di selezione delle stampanti e specificare le opzioni dell'aspetto, scegliere **Stampa** dal menu File.

Automazione di IBM SPSS Modeler

Il data mining può essere un processo complesso e talvolta lungo. Per questo motivo IBM SPSS Modeler comprende diversi tipi di supporto per la codifica e l'automazione.

- **Control Language for Expression Manipulation (CLEM)** è un linguaggio per l'analisi e la manipolazione dei flussi di dati nei flussi di IBM SPSS Modeler. I data miner fanno ampio ricorso al linguaggio CLEM nelle operazioni stream per eseguire attività sia semplici, come la derivazione dei profitti dai dati relativi ai costi e alle entrate, sia complesse, come la trasformazione dei dati dei registri Web in un insieme di campi e record contenenti informazioni utilizzabili.
- **Gli script** sono un potente strumento per automatizzare i processi nell'interfaccia utente. Possono eseguire lo stesso tipo di azioni compiute dagli utenti con un mouse o una tastiera. Inoltre è possibile specificare l'output e manipolare i modelli generati.

Capitolo 4. Nozioni sul data mining

Panoramica del data mining

Grazie a una vasta gamma di tecniche, il **data mining** consente di identificare nugget di informazioni utili all'interno di corpi di dati. Con il data mining, le informazioni vengono estratte per essere utilizzate in settori quali il supporto decisionale e l'elaborazione di previsioni e stime. Ingenti quantità di dati si rivelano in genere di scarsa utilità quando si presentano sotto forma di dati non elaborati. Ad avere valore sono infatti le informazioni nascoste all'interno dei dati.

Nel data mining, il successo è garantito da una corretta combinazione della conoscenza dei dati di cui dispongono gli utenti o gli esperti aziendali con tecniche avanzate di analisi attiva, che consentono di individuare le relazioni e le caratteristiche sottostanti i dati. Mediante il processo di data mining vengono generati modelli basati su dati cronologici, che vengono successivamente utilizzati per l'elaborazione di previsioni, il rilevamento di schemi e così via. Le tecniche per la creazione di questi modelli vengono definite di **apprendimento automatico** o **modellazione**.

Tecniche di modellazione

In IBM SPSS Modeler sono disponibili numerose tecniche di apprendimento automatico e di modellazione, raggruppabili in base alla tipologia dei problemi che sono in grado di risolvere.

- I metodi di modellazione predittiva includono le strutture ad albero delle decisioni, le reti neurali e i modelli statistici.
- I modelli di raggruppamento tramite cluster sono incentrati sull'identificazione di gruppi di record simili e sull'assegnazione di etichette ai record in base al gruppo di appartenenza. Tali metodi includono i metodi Kohonen, *K*-medie e TwoStep.
- Le regole di associazione consentono di associare una conclusione specifica, per esempio l'acquisto di un particolare prodotto, a un insieme di condizioni, come l'acquisto di numerosi altri prodotti.
- I modelli di screening possono essere utilizzati per sottoporre a screening i dati, in modo da individuare i campi e i record di maggior interesse ai fini della modellazione e identificare i valori anomali che potrebbero non rientrare in schemi già noti. I metodi disponibili includono la selezione delle funzioni e il rilevamento di anomalie.

Manipolazione e rilevamento dei dati

IBM SPSS Modeler include inoltre numerose funzionalità per l'applicazione delle proprie conoscenze ai dati:

- **Manipolazione dei dati.** Consente di creare nuovi elementi di dati derivanti da quelli esistenti e di suddividere i dati in sottoinsiemi significativi. È possibile unire e filtrare dati da numerose origini.
- **Esplorazione e visualizzazione.** È possibile visualizzare aspetti diversi dei dati utilizzando il nodo Esplora, che consente di analizzare i dati e visualizzare grafici e statistiche. Fra le opzioni avanzate vi è la possibilità di creare grafici interattivi, esportabili nei report dei progetti.
- **Statistiche.** Consentono di trovare conferma delle presunte relazioni tra le variabili all'interno dei dati. In IBM SPSS Modeler possono essere utilizzate anche statistiche fornite da IBM SPSS Statistics.
- **Verifica delle ipotesi.** Consentono di creare modelli del comportamento dei dati e di verificare tali modelli.

Queste funzionalità vengono in genere utilizzate per identificare un insieme interessante di attributi all'interno dei dati. Questi attributi possono quindi essere utilizzati nell'ambito delle tecniche di creazione di modelli per identificare regole e relazioni sottostanti.

Applicazioni comuni

Tra le applicazioni più comuni delle tecniche di data mining sono incluse le seguenti:

Direct mailing. È possibile stabilire quali gruppi demografici presentano il tasso di risposta più elevato e utilizzare queste informazioni per massimizzare la risposta ai messaggi futuri.

Indice di affidabilità creditizia. È possibile basare una decisione relativa a un credito sulla storia personale dei crediti di un singolo.

Risorse umane. È possibile analizzare le procedure di assunzione precedenti e creare regole decisionali che razionalizzano il processo di assunzione.

Ricerca medica. È possibile creare regole decisionali in grado di suggerire le procedure appropriate in base alla documentazione medica.

Analisi di mercato. È possibile stabilire quali sono le variabili associate alle vendite, per esempio gli aspetti geografici, il prezzo e le caratteristiche dei clienti.

Controllo qualità. È possibile analizzare i dati a partire dalla produzione e identificare le variabili che determinano i difetti dei prodotti.

Studio di linee guida. È possibile utilizzare i dati di indagine per formulare linee guida applicando regole decisionali per selezionare le variabili più importanti.

Servizi sanitari. È possibile combinare le indagini e i dati clinici per individuare le variabili che contribuiscono allo stato di buona salute.

Terminologia

I termini **attributo**, **campo** e **variabile** fanno riferimento a un singolo elemento di dati comune a tutti i casi considerati. Una raccolta di valori di attributo che fa riferimento a un caso specifico viene definita **record**, **esempio** o **caso**.

Valutazione dei dati

Il data mining si rivela utile solo se i dati che si desidera utilizzare soddisfano criteri specifici. Nelle sezioni riportate di seguito vengono illustrati alcuni aspetti dei dati e della loro applicazione che è opportuno considerare.

Assicurarsi che i dati siano disponibili

Sebbene questo possa apparire ovvio, è importante tenere in considerazione che i dati eventualmente disponibili potrebbero non essere in un formato facilmente utilizzabile. IBM SPSS Modeler supporta l'importazione di dati da database (tramite ODBC) o da file. È tuttavia possibile che i dati siano memorizzati in un altro formato su un computer non accessibile direttamente. In questo caso, prima dell'utilizzo è necessario eseguirne il download o il dump in un formato appropriato. Se sono distribuiti in origini e database diversi, i dati dovranno inoltre essere raccolti. È infine possibile che non siano disponibili in linea. Se esistono solo in formato cartaceo, prima di avviare il data mining è necessario immettere i dati.

Controllare se i dati coprono gli attributi pertinenti

Poiché l'obiettivo del data mining consiste nell'identificazione di attributi pertinenti, includere questo controllo può apparire singolare. È tuttavia estremamente utile esaminare i dati disponibili e tentare di identificare i fattori probabilmente pertinenti non registrati. Se si tenta di prevedere le vendite di gelati,

per esempio, è possibile che si disponga di ingenti quantità di informazioni sui punti vendita o sulla cronologia delle vendite, ma non di informazioni meteorologiche o sulle temperature, che avranno probabilmente un ruolo significativo. Sebbene non impediscano di ottenere risultati utili con il data mining, gli attributi mancanti possono limitare la precisione delle risultanti previsioni.

Un modo rapido per valutare la situazione consiste nell'eseguire un controllo completo dei dati. Prima di proseguire, è opportuno valutare l'opportunità di associare un nodo Esplora all'origine dati e generare un report completo.

Attenzione ai dati rumorosi

I dati contengono spesso errori o possono comprendere giudizi soggettivi e pertanto variabili. Questi fenomeni vengono definiti collettivamente **rumore**. La presenza di rumore nei dati è talvolta normale. Le regole sottostanti non possono infatti controllare il 100% dei casi.

In genere, la quantità di rumore presente nei dati è direttamente proporzionale alla difficoltà di ottenere risultati precisi. I metodi di apprendimento automatico di IBM SPSS Modeler, tuttavia, sono in grado di gestire dati di questo tipo e hanno fornito esito positivo su insiemi di dati con quasi il 50% di rumore.

Assicurarsi che i dati a disposizione siano sufficienti

Nell'ambito del data mining, le dimensioni di un insieme di dati non rivestono necessariamente un ruolo importante e risulta nettamente più significativa la *rappresentatività* dell'insieme di dati, così come la copertura dei risultati e delle combinazioni di variabili possibili.

In genere, per fornire una copertura rappresentativa sono necessari tanti più record quanti più attributi vengono presi in considerazione.

In presenza di dati rappresentativi e di regole sottostanti generali, è possibile che un campione di dati di alcune migliaia o addirittura centinaia di record fornisca risultati altrettanto validi di quelli garantiti da un milione di record, restituendo i risultati con maggiore rapidità.

Confrontarsi con gli esperti sui dati

In numerosi casi si utilizzano dati propri e si dispone quindi di un'elevata familiarità con il relativo contenuto e significato. Se tuttavia si gestiscono dati per un altro reparto aziendale o per un cliente, è preferibile potersi avvalere di esperti con una conoscenza approfondita dei dati, che possano semplificare l'identificazione degli attributi significativi e l'interpretazione dei risultati del data mining, distinguendo gli effettivi nugget di informazioni da dati falsati a causa di anomalie negli insiemi di dati.

Strategia per il data mining

Come la maggior parte delle attività di business, il data mining si rivela più efficiente se eseguito in modo sistematico e pianificato. Anche con strumenti di data mining all'avanguardia come IBM SPSS Modeler, per la maggior parte delle operazioni di data mining è richiesto un analista di business competente, in grado di gestire correttamente il processo. Per completare la pianificazione è necessario individuare quanto segue:

- Il problema di base che si desidera risolvere
- Le origini dati disponibili e le parti dei dati pertinenti per il problema corrente
- Le operazioni di pre-elaborazione e di pulitura dei dati da eseguire prima dell'avvio del data mining
- Le tecniche di data mining da utilizzare
- La modalità di valutazione dei risultati dell'analisi di data mining
- Le possibilità di utilizzo delle informazioni ottenute dal data mining

Un tipico processo di data mining può spesso risultare complicato. È infatti necessario tenere traccia di complessi problemi di business, di più origini dati, della diversa qualità dei dati nelle diverse origini, di numerose tecniche di data mining, di metodi differenti di valutazione dell'esito del data mining e così via.

La gestione di tutti questi aspetti può essere semplificata da un modello di elaborazione esplicitamente definito per il data mining. Il modello di elaborazione consente di rispondere alle domande elencate in precedenza in questa sezione e di assicurarsi che siano affrontati i punti importanti. Inoltre, tale modello guida l'utente nel processo di data mining in modo da superare la complessità dei dati.

Il processo di data mining consigliato per l'uso con SPSS Modeler è rappresentato dal modello CRISP-DM (Cross-Industry Standard Process for Data Mining), progettato come modello generale applicabile a una vasta gamma di settori e problemi di business.

Modello di elaborazione CRISP-DM

Il modello di processo generale CRISP-DM comprende sei fasi relative alle principali problematiche del data mining. Le sei fasi costituiscono complessivamente un processo ciclico, studiato per integrare il data mining nell'attività di business globale.

Di seguito vengono descritte queste sei fasi.

- **Business understanding.** Costituisce probabilmente la fase più importante del data mining. Comprende l'individuazione degli obiettivi di business, la valutazione della situazione, la definizione degli scopi del data mining e la creazione di un piano di progetto.
- **Fase di analisi dei dati.** I dati costituiscono la "materia prima" del data mining. Questa fase, in cui viene soddisfatta l'esigenza di analizzare le risorse di dati disponibili e le caratteristiche di tali risorse, comprende la raccolta dei dati iniziali, la descrizione e l'esplorazione dei dati e la verifica della relativa qualità. Il nodo Esplora, disponibile nella palette di nodi Output, è uno strumento indispensabile per l'analisi dei dati.
- **Preparazione dei dati.** Dopo la catalogazione delle risorse di dati è necessario preparare i dati per il data mining. Le operazioni di preparazione comprendono la selezione, la pulitura, la creazione, l'integrazione e la formattazione.
- **Modellazione.** Costituisce la parte più accattivante del data mining, in cui vengono utilizzati avanzati metodi di analisi per estrarre le informazioni dai dati. Questa fase comprende la selezione di tecniche di creazione di modelli, la generazione di strutture di test e la creazione e la valutazione di modelli.
- **Valutazione.** Una volta selezionati i modelli, è possibile valutare i vantaggi che può offrire il data mining ai fini del raggiungimento degli obiettivi di business. Gli elementi costitutivi di questa fase comprendono la valutazione dei risultati, la revisione del processo di data mining e l'individuazione dei passaggi successivi.
- **Distribuzione.** A questo punto è possibile usufruire dei vantaggi. Questa fase è incentrata sull'integrazione delle nuove conoscenze nei comuni processi di business allo scopo di risolvere il problema originale. Comprende la distribuzione, il monitoraggio e la gestione del piano, la creazione di un report finale e la revisione del progetto.

Questo modello di elaborazione è caratterizzato da alcuni punti principali. Innanzitutto, sebbene generalmente le fasi del processo seguano l'ordine illustrato nei paragrafi precedenti, esistono vari punti in cui si influenzano reciprocamente in modo non lineare. La fase di preparazione dei dati, per esempio, precede in genere la fase di modellazione. Tuttavia, le decisioni definite e le informazioni raccolte durante la fase di modellazione possono spesso indurre a riconsiderare alcune parti della fase di preparazione dei dati, che possono a loro volta presentare nuove problematiche a livello di modellazione. Le due fasi continuano a influenzarsi reciprocamente finché entrambe non vengono risolte in modo appropriato. Analogamente, la fase Valutazione può determinare una rivalutazione della fase Business Understanding

iniziale, e questo potrebbe portare alla scoperta che si sta tentando di ottenere una risposta alla domanda sbagliata. A questo punto, si può rivedere la fase Business Understanding e proseguire nuovamente con il resto del processo con un obiettivo più appropriato.

Il secondo punto principale è costituito dal carattere iterativo del data mining. Raramente il processo si esaurisce con la pianificazione di un progetto di data mining, il relativo completamento e l'ottenimento dei dati. L'utilizzo del data mining per soddisfare le richieste dei clienti costituisce un'attività costante. Le conoscenze acquisite con un ciclo di data mining generano solitamente nuove domande, nuove problematiche e nuove opportunità di individuare e soddisfare le esigenze dei clienti. Queste nuove domande, problematiche e opportunità possono in genere essere gestite con un nuovo processo di data mining. Questo processo di estrazione e identificazione delle nuove opportunità deve diventare parte dell'approccio allo svolgimento delle attività di business e un elemento di base della strategia di business complessiva.

Nella presente introduzione viene fornita solo una breve panoramica del modello di elaborazione CRISP-DM. Per informazioni dettagliate sull'utilizzo del modello, consultare le seguenti risorse:

- La *CRISP-DM Guide*, accessibile insieme ad altra documentazione dalla cartella `\Documentation` nel disco di installazione.
- Il sistema di guida di CRISP-DM, accessibile dal menu Start o facendo clic su **Guida in linea di CRISP-DM** nel menu Aiuto di IBM SPSS Modeler.

Tipi di modelli

IBM SPSS Modeler offre numerosi metodi di modellazione ricavati dall'apprendimento automatico, dall'intelligenza artificiale e dalla statistica. I metodi disponibili nella palette Modelli consentono di ricavare nuove informazioni dai dati e di sviluppare modelli predittivi. Ogni metodo ha determinati punti di forza e si presta meglio per particolari tipi di problemi.

Il documento *IBM SPSS Modeler Guida alle applicazioni* fornisce numerosi esempi di questi metodi, oltre a un'introduzione generale sul processo di modellazione. Tale guida è disponibile come esercitazione online e in formato PDF. Per ulteriori informazioni, consultare "Esempi di applicazioni" a pagina 5.

I metodi di modellazione sono suddivisi in tre categorie:

- Classificazione
- Associazione
- Segmentazione

Modelli Classificazione

I *Modelli Classificazione* utilizzano i valori di uno o più campi di **input** per prevedere il valore di uno o più campi di **output** o **obiettivo**. Alcuni esempi di queste tecniche sono: strutture ad albero delle decisioni (C&R Tree, QUEST, CHAID e algoritmi C5.0), regressione (lineare, logistica, lineare generalizzata e algoritmo di regressione di Cox), reti neurali, SVM (support vector machines) e reti bayesiane.

I modelli Classificazione consentono di prevedere un risultato noto, per esempio se un cliente effettuerà l'acquisto o vi rinuncerà oppure se una transazione rientra in un modello noto di comportamento fraudolento. Le tecniche di modellazione includono l'apprendimento automatico, l'induzione di regole, l'identificazione dei sottogruppi, i metodi statistici e la creazione di più modelli.

Nodi di classificazione



Il nodo Classificatore automatico crea e confronta svariati tipi di modelli per risultati binari (sì o no, abbandono oppure no e così via), consentendo di scegliere l'approccio migliore per una determinata analisi. Sono supportati numerosi algoritmi di modellazione ed è possibile selezionare i metodi da utilizzare, le opzioni specifiche per ognuno di essi e i criteri per confrontare i risultati. Il nodo genera un insieme di modelli basato sulle opzioni specificate e classifica i candidati migliori in base ai criteri indicati.



Il nodo Numerico automatico stima e confronta i modelli per i risultati di intervalli numerici continui utilizzando svariati metodi. Il nodo funziona in modo analogo al nodo Classificatore automatico e consente di scegliere gli algoritmi da utilizzare e di sperimentare più combinazioni di opzioni in un singolo passaggio di modellazione. Gli algoritmi supportati includono reti neurali, C&R Tree, CHAID, regressione lineare, regressione lineare generalizzata e SVM (Support Vector Machine). I modelli si possono confrontare in base a correlazione, errore relativo o numero di variabili utilizzato.



Il nodo Struttura ad albero di classificazione e regressione (C&R) genera una struttura ad albero delle decisioni che consente di prevedere o classificare osservazioni future. Il metodo utilizza partizionamento ricorsivo per suddividere i record di addestramento in segmenti, riducendo l'impurità ad ogni passaggio. Un nodo della struttura ad albero è considerato "puro" quando il 100% dei casi nel nodo fa parte di una categoria specifica del campo obiettivo. I campi obiettivo e di input possono essere intervalli numerici o categoriali (nominali, ordinali o flag); tutte le suddivisioni sono binarie (solo due sottogruppi).



Il nodo QUEST offre un metodo di classificazione binario per la creazione di strutture ad albero delle decisioni, progettato per ridurre i tempi di elaborazione necessari per le analisi C&R Tree più complesse, riducendo inoltre la tendenza dei metodi per le strutture ad albero di classificazione a favorire gli input che consentono un numero maggiore di suddivisioni. I campi di input possono essere intervalli numerici (continui), ma il campo obiettivo deve essere categoriale. Tutte le suddivisioni sono binarie.



Il nodo CHAID genera una struttura ad albero delle decisioni utilizzando statistiche chi-quadrato per identificare suddivisioni ottimali. A differenza dei nodi C&R Tree e QUEST, il nodo CHAID può generare strutture ad albero non binarie e pertanto alcune suddivisioni possono avere più di due rami. I campi obiettivo e di input possono essere intervallo numerico (continui) o categoriali. Un CHAID completo è una modificazione di CHAID che esegue operazioni avanzate per l'analisi di tutte le suddivisioni possibili, ma richiede tempi di elaborazione maggiori.



Il nodo C5.0 crea una struttura ad albero delle decisioni o un insieme di regole. Il modello suddivide il campione in base al campo che fornisce il massimo guadagno di informazioni a ogni livello. Il campo obiettivo deve essere categoriale. Sono consentite suddivisioni multiple in più di due sottogruppi.



Il nodo Elenco di decisioni identifica i sottogruppi o i segmenti che mostrano una probabilità maggiore o minore che si verifichi un determinato risultato binario rispetto alla popolazione globale. Per esempio, è possibile che si cerchino i clienti non a rischio di abbandono o quelli che più probabilmente rispondano in modo favorevole a una campagna. È possibile incorporare le proprie conoscenze di business nel modello aggiungendo propri segmenti personalizzati e visualizzando in anteprima modelli alternativi uno accanto all'altro per confrontarne i risultati. I modelli Elenco di decisioni consistono in un elenco di regole in cui ogni regola ha una condizione e un risultato. Le regole vengono applicate in ordine e la prima regola corrispondente determina il risultato.



I modelli di regressione lineare prevedono un target continuo basato sulle relazioni lineari tra l'obiettivo e uno o più predittori.



Il nodo fattoriale/PCA offre potenti tecniche di riduzione dei dati che consentono di diminuirne la complessità. L'analisi dei componenti principali (PCA, Principal Components Analysis) trova le combinazioni lineari dei campi di input che catturano meglio la varianza nell'intero insieme di campi, dove i componenti sono ortogonali (perpendicolari) l'uno rispetto all'altro. L'analisi fattoriale tenta di identificare i concetti sottostanti, o fattori, che spiegano lo schema delle correlazioni all'interno dell'insieme di campi osservati. Entrambi gli approcci mirano a trovare un numero ridotto di campi derivati che riassumono in modo efficace le informazioni presenti nell'insieme originale di campi.



Il nodo Selezione funzioni effettua lo screening dei campi di input, rimuovendoli in base a un insieme di criteri quali la percentuale di valori mancanti. Classifica quindi gli input restanti in ordine di importanza rispetto a un determinato obiettivo. Per esempio, dato un insieme di dati con centinaia di input potenziali, quali sono quelli con la maggiore probabilità di essere utili nella modellazione di risultati clinici?



L'analisi discriminante prevede presupposti più rigidi rispetto alla regressione logistica, ma può essere una valida alternativa o un complemento dell'analisi di regressione logistica quando vengono soddisfatti tali presupposti.



La regressione logistica, una tecnica statistica che consente di classificare i record in base ai valori dei campi di input, è analoga alla regressione lineare ma, al posto di un intervallo numerico, prende un campo obiettivo categoriale.



Il modello Lineare generalizzato amplia il modello lineare generale in modo che la variabile dipendente venga linearmente correlata ai fattori e alle covariate tramite una funzione di collegamento specifica. Inoltre, il modello consente alla variabile dipendente di avere una distribuzione non normale. Copre la funzionalità di un grande numero di modelli statistici, inclusi modelli di regressione lineare, modelli di regressione logistica, modelli loglineari per dati dei conteggi e modelli di sopravvivenza censurati per intervallo.



Un modello misto lineare generalizzato (GLMM) estende il modello lineare in modo che l'obiettivo possa avere una distribuzione non normale, sia linearmente correlato ai fattori e alle covariate tramite una funzione di collegamento specifica e in modo che le osservazioni possano essere correlate. I modelli misti lineari generalizzati includono un'ampia gamma di modelli, dalla regressione lineare semplice ai modelli multilivello complessi per i dati longitudinali non normali.



Il nodo Regressione di Cox consente di generare un modello di sopravvivenza per i dati della relazione tempo-evento in presenza di record censurati. Il modello produce una funzione di sopravvivenza che prevede la probabilità che l'evento di interesse si sia verificato a una determinata ora (t) per i valori dati delle variabili di input.



Il nodo SVM (Support Vector Machine) consente di classificare i dati in uno di due gruppi senza sovradattamento. Il nodo SVM è particolarmente indicato per l'utilizzo con insiemi di dati di grandi dimensioni, cioè quelli con un elevato numero di campi di input.



Il nodo Rete bayesiana consente di generare un modello di probabilità combinando elementi osservati e registrati con conoscenze del mondo reale per stabilire la probabilità di occorrenze. Il nodo si concentra sulle reti TAN (Tree Augmented Naïve Bayes) e coperta di Markov, che sono prevalentemente utilizzate a scopo di classificazione.



Il nodo Modello risposta autoapprendimento consente di creare un modello in cui è possibile utilizzare un unico nuovo caso oppure un numero limitato di nuovi casi per eseguire una nuova stima del modello senza doverlo riaddestrare con tutti i dati.



Il nodo Serie temporali stima i modelli di livellamento esponenziale, i modelli ARIMA (Autoregressive Integrated Moving Average, autoregressivi integrati a media mobile) univariati e ARIMA (o a funzione di trasferimento) multivariati per i dati di serie temporali e genera previsioni di prestazioni future. Il nodo Serie temporali deve sempre essere preceduto da un nodo Intervalli di tempo.



Il nodo Elemento vicino più prossimo K (KNN) associa un nuovo caso alla categoria o valore degli oggetti K più vicini ad esso nello spazio predittore, dove K è un numero intero. I casi simili sono vicini gli uni agli altri, mentre i casi dissimili sono distanti gli uni dagli altri.

Modelli Associazione

I *Modelli Associazione* trovano nei dati gli schemi nei quali una o più entità (eventi, acquisti o attributi) vengono associate a un'altra o a diverse altre entità. I modelli creano insiemi di regole che definiscono queste relazioni. I campi all'interno dei dati possono essere sia di input che obiettivo. Queste associazioni possono essere trovate manualmente, ma gli algoritmi delle regole di associazione funzionano più velocemente e possono esplorare schemi più complessi. I modelli Apriori e Carma sono esempi dell'utilizzo di questi algoritmi. Un altro tipo di modello di associazione è un modello di rilevamento di sequenze, che trova gli schemi sequenziali nei dati a struttura temporale.

I modelli Associazione sono utili soprattutto nella previsione di risultati multipli; per esempio, i clienti che hanno acquistato il prodotto X hanno acquistato anche Y e Z . I modelli Associazione associano una particolare conclusione (per esempio la decisione di acquistare qualcosa) a un insieme di condizioni. Rispetto agli algoritmi della struttura ad albero delle decisioni più tradizionali (C5.0 e C&RT), quelli delle regole di associazione presentano il vantaggio di poter definire associazioni tra qualsiasi tipo di attributo. Mentre un algoritmo della struttura ad albero delle decisioni genera regole con un'unica conclusione, gli algoritmi di associazione tentano di individuare più regole, ciascuna delle quali può fornire una diversa conclusione.

Nodi di associazione



Il nodo Apriori estrae un insieme di regole dai dati, estrapolando le regole con il più alto contenuto di informazioni. Apriori offre cinque diversi metodi per la selezione delle regole e utilizza uno schema di indicizzazione sofisticato per elaborare in modo efficiente insiemi di dati di grandi dimensioni. In caso di problemi complessi, l'addestramento di Apriori è in genere più rapido. Apriori non ha un limite arbitrario per quanto riguarda il numero di regole che possono essere mantenute e può gestire regole con un massimo di 32 precondizioni. Apriori richiede che tutti i campi di input e output siano categoriali ma garantisce prestazioni migliori perché è ottimizzato per questo tipo di dati.



Il modello CARMA estrae un insieme di regole dai dati senza che venga richiesto all'utente di specificare i campi di input o obiettivo. A differenza di Apriori, il nodo CARMA fornisce le impostazioni di creazione per il supporto delle regole (sia per l'antecedente che per il conseguente) anziché solo per il supporto antecedente. Pertanto, le regole generate possono essere utilizzate per una gamma più vasta di applicazioni, ad esempio per trovare un elenco di prodotti o di servizi (antecedenti) il cui conseguente è rappresentato dall'articolo che si desidera promuovere per le festività correnti.



Il nodo Sequenza consente di scoprire le regole di associazione nei dati sequenziali o basati su valori temporali. Per sequenza si intende un elenco di serie di elementi che tendono a ricorrere secondo un ordine prevedibile. Ad esempio, un cliente che acquista un rasoio e la lozione dopobarba potrebbe in seguito acquistare la schiuma da barba. Il nodo Sequenza si basa sull'algoritmo delle regole di associazione CARMA, che utilizza un metodo efficiente in due passaggi per trovare le sequenze.

Modelli di segmentazione

I *Modelli di segmentazione* suddividono i dati in segmenti, o cluster, di record con schemi simili di campi di input. I modelli di segmentazione si concentrano soltanto sui campi di input e quindi non tengono in considerazione i campi di output o obiettivo. Tra gli esempi di modelli di segmentazione citiamo le reti di Kohonen, il raggruppamento in cluster Medie K, il raggruppamento in cluster TwoStep e il rilevamento delle anomalie.

I modelli di segmentazione (denominati anche "modelli di cluster") sono utili nei casi in cui il risultato specifico non è conosciuto, per esempio nell'identificazione di nuovi modelli di comportamento fraudolento oppure nell'identificazione di gruppi di interesse nella base clienti. I modelli di raggruppamento tramite cluster sono incentrati sull'identificazione di gruppi di record simili e sull'assegnazione di etichette ai record in base al gruppo di appartenenza. Ciò avviene senza che siano necessarie previe conoscenze relative ai gruppi e alle loro caratteristiche e distingue i modelli di raggruppamento tramite cluster da altre tecniche di modellazione - non esiste un campo di destinazione o un output predefinito per il modello da prevedere. Per questi modelli non esistono risposte corrette o errate. Il valore è determinato dalla capacità di acquisire gruppi significativi all'interno dei dati e di fornire descrizioni utili di tali raggruppamenti. Spesso i modelli di raggruppamento tramite cluster vengono utilizzati per creare cluster o segmenti poi utilizzati come input nelle analisi successive, per esempio segmentando i potenziali clienti in sottogruppi omogenei.

Nodi di segmentazione



Il nodo Cluster automatico stima e confronta i modelli di cluster che identificano gruppi di record con caratteristiche simili. Il nodo funziona in modo analogo ad altri nodi Modelli automatici e consente di sperimentare varie combinazioni di opzioni in un singolo passaggio di modellazione. I modelli si possono confrontare utilizzando misure di base con cui tentare di filtrare e classificare l'utilità dei modelli di cluster e fornire una misura in base all'importanza di determinati campi.



Il nodo Medie K raggruppa l'insieme di dati in gruppi distinti (o cluster). Il metodo definisce un numero fisso di cluster, esegue un'assegnazione iterativa dei record ai cluster e modifica i centri di cluster finché un'ulteriore ridefinizione non consente più un miglioramento del modello. Invece di tentare di prevedere un risultato, il nodo K-medie utilizza un processo denominato apprendimento non supervisionato per scoprire gli schemi nell'insieme di campi di input.



Il nodo Kohonen genera un tipo di rete neurale che può essere utilizzato per raggruppare l'insieme di dati in gruppi distinti. Al termine dell'apprendimento della rete, i record analoghi dovranno essere vicini nella mappa di output, mentre i record diversi saranno a notevole distanza. Per identificare le unità forti, è possibile controllare il numero di osservazioni catturate da ciascuna unità nel nugget del modello. In questo modo è possibile avere un'idea del numero appropriato di cluster.



Il nodo TwoStep è un metodo di raggruppamento tramite cluster in due fasi. La prima fase esegue un singolo passaggio nei dati per comprimere i dati di input non elaborati in un insieme gestibile di cluster secondari. Nella seconda fase viene utilizzato un metodo di raggruppamento tramite cluster gerarchico per unire progressivamente i cluster secondari in cluster sempre più grandi. Il nodo TwoStep offre il vantaggio di stimare automaticamente il numero ottimale di cluster per i dati di addestramento. Può gestire in modo efficiente tipi di campo misti e insiemi di dati di grandi dimensioni.



Il nodo Rilevamento anomalie identifica casi insoliti, o valori anomali, non conformi a schemi di dati "normali". Con questo nodo è possibile identificare valori anomali anche se questi non rientrano in schemi precedentemente conosciuti e anche se l'utente non sa esattamente ciò che sta cercando.

Modelli mining nel database

IBM SPSS Modeler supporta l'integrazione con gli strumenti di data mining e di modellazione offerti dai fornitori di database, quali Oracle Data Miner, IBM DB2 InfoSphere Warehouse e Microsoft Analysis Services. Operando all'interno dell'applicazione IBM SPSS Modeler è infatti possibile sia creare modelli che calcolarne il punteggio e archivarli nel database. Per ulteriori informazioni, consultare la *IBM SPSS Modeler Guida al mining nel database* disponibile nel DVD del prodotto.

Modelli IBM SPSS Statistics

Se si dispone di una copia di IBM SPSS Statistics installata con licenza sul proprio computer, è possibile accedere a determinate routine IBM SPSS Statistics ed eseguirle dall'interno di IBM SPSS Modeler per creare e determinare il punteggio dei modelli.

Ulteriori informazioni

È inoltre disponibile una documentazione esaustiva sugli algoritmi di modellazione. Per ulteriori informazioni, vedere la *IBM SPSS Modeler Algorithms Guide* disponibile sul DVD del prodotto.

Esempi di Oracle Data Mining

Il metodo migliore per apprendere i concetti pratici del data mining consiste nel partire da un esempio. Numerosi esempi di applicazioni sono disponibili in *IBM SPSS Modeler Guida alle applicazioni*, che fornisce brevi indicazioni mirate su specifici metodi e tecniche di modellazione. Consultare l'argomento "Esempi di applicazioni" a pagina 5 per ulteriori informazioni.

Capitolo 5. Creazione di flussi

Panoramica sulla creazione dei flussi

Il data mining di IBM SPSS Modeler è incentrato sul processo di elaborazione dei dati attraverso una serie di nodi, denominato **stream**. La serie di nodi rappresenta le operazioni che verranno eseguite sui dati, mentre i collegamenti tra i nodi indicano la direzione del flusso di dati. Con un flusso di dati, è in genere possibile leggere i dati di IBM SPSS Modeler, eseguire una serie di manipolazioni e quindi inviarli a una destinazione, per esempio a una tabella o a un visualizzatore.

Si supponga, per esempio, che sia necessario aprire una sorgente dati, aggiungere un nuovo campo, selezionare i record in base ai valori del nuovo campo e quindi visualizzare i risultati in una tabella. In questo caso, il flusso di dati includerà quattro nodi:



Un nodo Testo File, che è possibile impostare per leggere i dati dalla sorgente dati.



Un nodo Ricava, mediante il quale è possibile aggiungere il nuovo campo calcolato all'insieme di dati.



Un nodo Seleziona, che consente di impostare i criteri di selezione per l'esclusione di record dal flusso di dati.



Un nodo Tabella, che consente di visualizzare sullo schermo i risultati delle manipolazioni.

Creazione dei flussi di dati

L'interfaccia di IBM SPSS Modeler consente di eseguire il data mining visuale grazie all'utilizzo dei diagrammi dei flussi di dati. La creazione dei flussi di dati prevede i seguenti passaggi fondamentali:

- Aggiunta di nodi nell'area del flusso.
- Connessione dei nodi per creare un flusso.
- Definizione delle opzioni dei nodi o del flusso.
- Esecuzione del flusso.

In questa sezione vengono fornite informazioni dettagliate sull'utilizzo dei nodi per la creazione di stream di dati più complessi e vengono inoltre illustrate le opzioni e le impostazioni relative ai nodi e i flussi. Per esempi di procedure dettagliate relative alla creazione di flussi mediante i dati disponibili in IBM SPSS Modeler (cartella Demos del programma di installazione), vedere "Esempi di applicazioni" a pagina 5.

Utilizzo dei nodi

I nodi vengono utilizzati in IBM SPSS Modeler per semplificare l'analisi dei dati. I vari nodi presenti nello spazio di lavoro rappresentano oggetti e operazioni diversi. La palette disponibile nella parte inferiore della finestra di IBM SPSS Modeler contiene tutti i possibili nodi utilizzati per la creazione dei flussi.

Sono disponibili vari tipi di nodi. I **nodi origine** importano i dati nel flusso e si trovano nella scheda Input della palette dei nodi. I **nodi di elaborazione** eseguono operazioni su singoli campi e record di dati e sono disponibili nelle schede Oper su record e Oper su campi della palette. I **nodi di output** generano vari tipi di output per dati, grafici e risultati di modelli e sono inclusi nelle schede Grafici, Output ed Esporta della palette dei nodi. I **nodi Modelli** utilizzano algoritmi statistici per creare nugget del modello e sono contenuti nella scheda Modelli e (se attivata) nella scheda Modelli database della palette dei nodi. Consultare l'argomento "Palette nodi" a pagina 16 per ulteriori informazioni.

È possibile connettere i nodi per creare i flussi che, una volta eseguiti, consentono di visualizzare le relazioni e di trarre le conclusioni appropriate. I flussi sono simili agli script, è infatti possibile salvarli e riutilizzarli con file di dati diversi.

Un nodo eseguibile che elabora i dati di un flusso è detto **nodo terminale**. Un nodo Modelli o di output è un nodo terminale se compare alla fine di un flusso o di un ramo dello stesso. Ai nodi terminali non è possibile collegare altri nodi.

Nota: è possibile personalizzare la tavolozza Nodi. Consultare l'argomento "Personalizzazione della palette dei nodi" a pagina 154 per ulteriori informazioni.

Aggiunta di nodi a un flusso

È possibile aggiungere nodi a un flusso dalla palette dei nodi in vari modi:

- Fare doppio clic su un nodo nella tavolozza. *Nota:* facendovi doppio clic sopra, il nodo verrà connesso automaticamente al flusso corrente. Consultare l'argomento "Connessione di nodi in un flusso" per ulteriori informazioni.
- Selezionare un nodo e trascinarlo dalla palette nell'area di disegno del flusso.
- Fare clic su un nodo nella palette e quindi sull'area di disegno del flusso.
- Selezionare l'opzione desiderata dal menu Inserisci di IBM SPSS Modeler.

Dopo avere aggiunto un nodo nell'area di disegno del flusso, farvi doppio clic sopra per aprire la relativa finestra di dialogo. Le opzioni disponibili nella finestra di dialogo variano in base al tipo di nodo da aggiungere. Per informazioni sui controlli specifici della finestra di dialogo, fare clic sul pulsante **Aiuto**.

Rimozione di nodi

Per rimuovere un nodo dal flusso di dati, fare clic sul nodo e premere il tasto CANC oppure farvi clic sopra con il pulsante destro del mouse e scegliere **Elimina** dal menu.

Connessione di nodi in un flusso

I nodi aggiunti nell'area del flusso non formano un flusso di dati fino a quando non vengono connessi. Le connessioni tra i nodi indicano la direzione del flusso di dati da un'operazione alla successiva. Esistono diversi modi per connettere i nodi in modo da formare un flusso: doppio clic, utilizzando il pulsante centrale del mouse, o manualmente.

Per aggiungere e connettere i nodi mediante il doppio clic

Il modo più semplice per creare un flusso consiste nel fare doppio clic sui nodi nella palette. Il nuovo nodo verrà connesso automaticamente al nodo selezionato nell'area del flusso. Se, per esempio, l'area contiene un nodo Database, è possibile selezionare questo nodo e fare doppio clic sul nodo successivo nella palette, quale un nodo Ricava. Il nodo Ricava verrà connesso automaticamente al nodo Database esistente. È possibile ripetere questa operazione fino a quando si raggiunge un nodo terminale, per esempio un nodo Istogramma o Tabella. A questo punto, i nuovi nodi verranno connessi all'ultimo nodo a monte diverso da un nodo terminale.

Per connettere i nodi utilizzando il pulsante centrale del mouse

Nell'area del flusso, è possibile fare clic e trascinare da un nodo a un altro utilizzando il pulsante centrale del mouse. (Se il mouse in uso non dispone di un pulsante centrale, è possibile simularlo tenendo premuto il tasto Alt mentre si trascina con il mouse da un nodo a un altro).

Per connettere i nodi manualmente

Se non è disponibile un pulsante centrale del mouse e si preferisce connettere manualmente i nodi, utilizzare il menu di scelta rapida di un nodo per connetterlo a un altro nodo già presente nell'area.

1. Fare clic con il pulsante destro del mouse sul nodo da cui si desidera iniziare la connessione. Viene visualizzato il menu del nodo.
2. Fare clic su **Connetti** nel menu di scelta rapida.
3. Sul nodo iniziale e in prossimità del cursore viene visualizzata l'icona di una connessione. Fare clic su un secondo nodo per connettere i due nodi.

Per la connessione dei nodi, esistono numerose indicazioni da seguire. Verrà visualizzato un messaggio di errore se si tenta di eseguire i seguenti tipi di connessione:

- Una connessione verso un nodo origine
- Una connessione da un nodo finale
- Una connessione relativa a un nodo per il quale è stato superato il numero massimo di connessioni di input consentite
- Una connessione di due nodi già connessi
- Circolarità (i dati tornano a un nodo dal quale sono già passati)

Ignorare i nodi di un flusso

Quando si ignora un nodo del flusso di dati, tutte le relative connessioni di input e output vengono sostituite da connessioni dirette dai nodi di input ai nodi di output. Se il nodo non dispone delle connessioni di input e di output, tutte le relative connessioni vengono eliminate anziché essere reindirizzate.

Si consideri, per esempio, un flusso che deriva un nuovo campo, filtra i campi e quindi analizza i risultati in un istogramma e in una tabella. Se si desidera visualizzare lo stesso grafico e la stessa tabella di dati anche *prima* che i campi vengano filtrati, è possibile aggiungere nuovi nodi Istogramma e Tabella al flusso oppure ignorare il nodo Filtro. Se si ignora il campo Filtro, le connessioni al grafico e alla tabella vengono trasferite direttamente dal nodo Deriva e il nodo Filtro viene disconnesso dal flusso.

Per ignorare un nodo

1. Nell'area di disegno del flusso, fare doppio clic con il pulsante centrale del mouse sul nodo che si desidera ignorare. In alternativa, è possibile utilizzare Alt+doppio clic.

Nota: per annullare l'operazione è possibile fare clic su **Annulla** nel menu Modifica oppure premere Ctrl+Z.

Disattivazione di nodi in un flusso

I nodi di elaborazione con un unico input all'interno dei flussi possono essere disattivati, con il risultato che il nodo viene ignorato durante l'esecuzione. In questo modo non è necessario rimuovere o ignorare il nodo e questo può restare connesso agli altri nodi. Le impostazioni del nodo possono ancora essere aperte e modificate; tuttavia, le modifiche non verranno applicate finché il nodo non viene nuovamente attivato.

Per esempio, si supponga di avere un flusso che filtra diversi campi e in seguito crea i modelli con l'insieme di dati ridotto. Se si desidera creare gli stessi modelli anche *senza* che i campi vengano filtrati, per verificare se i risultati migliorano, è possibile disattivare il nodo Filtro. Quando si disattiva il nodo Filtro, le connessioni ai nodi Modelli vengono trasferite direttamente dal nodo Ricava al nodo Tipo.

Per disattivare un nodo

1. Sull'area del flusso, fare clic con il pulsante destro del mouse sul nodo che si desidera disattivare.
2. Fare clic su **Disattiva nodo** nel menu di scelta rapida.

In alternativa, fare clic su **Nodo > Disattiva nodo** nel menu Modifica. Quando si desidera includere nuovamente il nodo nel flusso, fare clic su **Attiva nodo** allo stesso modo.

Nota: per annullare l'operazione è possibile fare clic su **Annulla** nel menu Modifica oppure premere Ctrl+Z.

Aggiunta di nodi a connessioni esistenti

È possibile aggiungere un nuovo nodo tra due nodi già connessi trascinando la freccia che connette i due nodi.

1. Fare clic con il pulsante centrale del mouse sulla freccia della connessione in cui si desidera inserire il nuovo nodo. In alternativa, tenere premuto il tasto Alt in sostituzione del pulsante centrale del mouse.
2. Trascinare la connessione sul nodo che si desidera aggiungere e rilasciare il pulsante del mouse.

Note: per rimuovere le nuove connessioni dal nodo e ripristinare le originali, è possibile **ignorare** il nodo.

Eliminazione delle connessioni tra nodi

Per eliminare la connessione tra due nodi:

1. Fare clic con il pulsante destro del mouse sulla freccia della connessione.
2. Nel menu, fare clic su **Elimina connessione**.

Per eliminare tutte le connessioni di un nodo, procedere in uno dei seguenti modi:

- Selezionare il nodo e premere F3.
- Selezionare il nodo e nel menu principale fare clic su:

Modifica > Nodo > Disconnetti

Impostazione delle opzioni per i nodi

Dopo che i nodi sono stati creati e connessi, è possibile personalizzarli in diversi modi. Fare clic con il pulsante destro del mouse su un nodo e scegliere una delle opzioni del menu.

- Fare clic su **Modifica** per aprire la finestra di dialogo relativa al nodo selezionato.
- Fare clic su **Connetti** per connettere manualmente un nodo a un altro nodo.
- Fare clic su **Disconnetti** per eliminare tutte le connessioni a e dal nodo.
- Fare clic su **Rinomina e annota** per visualizzare la scheda Annotazioni della finestra di dialogo di modifica.
- Fare clic su **Nuovo commento** per aggiungere un commento relativo al nodo. Consultare l'argomento "Aggiunta di commenti e annotazioni a nodi e flussi" a pagina 55 per ulteriori informazioni.
- Fare clic su **Disattiva nodo** per "nascondere" il nodo durante l'elaborazione. Per rendere il nodo nuovamente visibile per l'elaborazione, fare clic su **Attiva nodo**. Consultare l'argomento "Disattivazione di nodi in un flusso" a pagina 37 per ulteriori informazioni.
- Fare clic su **Taglia** o **Elimina** per rimuovere il nodo o i nodi selezionati dall'area del flusso. *Nota:* se si fa clic su **Taglia** è possibile incollare i nodi, mentre **Elimina** non lo consente.
- Fare clic su **Copia nodo** per creare una copia del nodo senza connessioni. La copia potrà essere aggiunta a un nuovo flusso o a un flusso esistente.
- Fare clic su **Carica nodo** per aprire un nodo salvato in precedenza e caricare le relative opzioni nel nodo corrente selezionato. *Nota:* i nodi devono essere dello stesso tipo.
- Fare clic su **Recupera nodo** per recuperare un nodo da un IBM SPSS Collaboration and Deployment Services Repository connesso.

- Fare clic su **Salva nodo** per salvare i dettagli del nodo in un file. È possibile caricare i dettagli del nodo solo in un altro nodo dello stesso tipo.
- Fare clic su **Archivia nodo** per archiviare il nodo selezionato in un IBM SPSS Collaboration and Deployment Services Repository connesso.
- Fare clic su **Cache** per espandere il menu e visualizzare le opzioni relative alla memorizzazione nella cache per il nodo selezionato.
- Fare clic su **Mappatura dati** per espandere il menu e visualizzare le opzioni per la mappatura dei dati a una nuova sorgente o per la specifica dei campi obbligatori.
- Fare clic su **Crea Supernodo** per espandere il menu e visualizzare le opzioni per la creazione di un Supernodo nel flusso corrente.
- Fare clic su **Genera Nodo input utente** per sostituire il nodo selezionato. Gli esempi generati dal nodo includeranno gli stessi campi del nodo corrente.
- Fare clic su **Esegui da qui** per eseguire tutti i nodi terminali downstream del nodo selezionato.

Opzioni di memorizzazione dei dati nella cache per i nodi

Per ottimizzare l'esecuzione del flusso, è possibile impostare una **cache** in qualsiasi nodo diverso da un nodo terminale. La cache configurata per un nodo viene riempita con i dati che passano attraverso il nodo alla successiva esecuzione del flusso di dati. Da questo momento in poi, i dati vengono letti dalla cache, che è archiviata in una directory temporanea, anziché dalla sorgente dati.

La memorizzazione dei dati nella cache è utile soprattutto dopo un'operazione che richiede tempo, per esempio un ordinamento, un'unione o un'aggregazione. Si supponga, per esempio, che esista un nodo origine impostato in modo da leggere i dati relativi alle vendite da un database e un nodo Aggregazione che riepiloga le vendite in base alla posizione. È possibile impostare una cache nel nodo Aggregazione anziché nel nodo origine, in modo tale da memorizzarvi i dati aggregati anziché l'intero insieme di dati.

Nota: l'impostazione di una cache nei nodi origine, che consente semplicemente di archiviare una copia dei dati originali non appena vengono letti in IBM SPSS Modeler, nella maggior parte dei casi non migliorerà le prestazioni.

I nodi con la cache attivata sono indicati da una piccola icona di documento nell'angolo in alto a destra. Quando i dati vengono memorizzati nella cache del nodo, l'icona del documento diventa di colore verde.

Per attivare una cache:

1. Nell'area del flusso, fare clic con il pulsante destro del mouse sul nodo e fare clic su **Cache** nel menu.
2. Fare clic su **Attiva** nel sottomenu.
3. Per disattivare la cache, fare clic con il pulsante destro del mouse sul nodo e fare clic su **Disattiva** nel sottomenu.

Memorizzazione nella cache di nodi nel database

Per i flussi eseguiti nel database, i dati possono essere memorizzati nella cache midstream in una tabella temporanea nel database anziché nel file system. Insieme all'ottimizzazione SQL, ciò consente di ottenere significativi vantaggi in termini di prestazioni. Per esempio, l'output di un flusso che unisce più tabelle per creare una visualizzazione data mining può essere memorizzato nella cache e riutilizzato in un secondo tempo in base alle esigenze. Generando automaticamente codice SQL per tutti i nodi downstream, è possibile migliorare ulteriormente le prestazioni.

Se si memorizzano dati nella cache del database con stringhe di lunghezza superiore a 255 caratteri, assicurarsi che esista un nodo tipologia a monte del nodo di cache e che i valori di campo vengano letti oppure impostare la lunghezza della stringa utilizzando il parametro `default_sql_string_length` nel file `options.cfg`. In questo modo la colonna corrispondente nella tabella temporanea viene impostata sulla larghezza corretta per adattarsi alle stringhe.

Per sfruttare la memorizzazione nella cache del database, è necessario che siano attivate l'ottimizzazione SQL e la memorizzazione nella cache del database. Si noti che le impostazioni di ottimizzazione del Server prevalgono su quelle del Client. Consultare l'argomento "Impostazione delle opzioni di ottimizzazione per i flussi" a pagina 44 per ulteriori informazioni.

Con la memorizzazione nella cache del database attivata, è sufficiente fare clic con il pulsante destro del mouse su qualsiasi nodo non terminale per memorizzare i dati nella cache in quel punto e la cache verrà creata automaticamente direttamente nel database alla successiva esecuzione del flusso. Se la memorizzazione nella cache del database o l'ottimizzazione SQL non sono attivate, la cache verrà scritta nel file system.

Nota: i seguenti database supportano tabelle temporanee per la memorizzazione di dati nella cache: DB2, Netezza, Oracle, SQL Server e and Teradata. Altri database utilizzano una tabella normale per la memorizzazione di dati nella cache. È possibile personalizzare il codice SQL per database specifici - contattare l'assistenza.

Per svuotare una cache

Un'icona del documento di colore bianco su un nodo indica che la relativa cache è vuota. Quando la cache è piena, l'icona del documento diventa di colore verde. Per sostituire il contenuto della cache, è necessario innanzitutto svuotare la cache e quindi rieseguire il flusso di dati per riempirla di nuovo.

1. Nell'area di disegno del flusso, fare clic con il pulsante destro del mouse sul nodo e fare clic su **Cache** nel menu.
2. Fare clic su **Svuota** nel sottomenu.

Per salvare una cache

È possibile salvare il contenuto di una cache come file di dati IBM SPSS Statistics (*.sav). e quindi ricaricare il file come una cache oppure impostare un nodo che utilizza il file cache come la propria sorgente dati. È inoltre possibile caricare una cache salvata da un altro progetto.

1. Nell'area di disegno del flusso, fare clic con il pulsante destro del mouse sul nodo e fare clic su **Cache** nel menu.
2. Fare clic su **Salva cache** nel sottomenu.
3. Nella finestra di dialogo Salva cache passare alla posizione in cui si desidera salvare il file cache.
4. Digitare un nome nella casella di testo Nome file.
5. Verificare che nell'elenco dei tipi di file sia selezionato *.sav e fare clic su **Salva**.

Per caricare una cache

I file cache che sono stati salvati prima di essere rimossi dal nodo possono essere ricaricati.

1. Nell'area di disegno del flusso, fare clic con il pulsante destro del mouse sul nodo e fare clic su **Cache** nel menu.
2. Fare clic su **Carica cache** nel sottomenu.
3. Nella finestra di dialogo Carica cache passare alla posizione del file cache, selezionarlo e fare clic su **Carica**.

Anteprima dei dati nei nodi

Per fare in modo che i dati vengano modificati secondo le aspettative durante la creazione di un flusso, è possibile elaborare i dati stessi mediante un nodo Tabella a ogni punto significativo del processo. In alternativa, è possibile generare da ciascun nodo un'anteprima di un campione dei dati che verranno creati, riducendo così il tempo necessario per creare ciascun nodo.

Per i nodi a monte di un nugget del modello, l'anteprima mostra i campi di input; per un nugget del modello o per i nodi downstream dell'insieme (ad eccezione dei nodi terminali) l'anteprima mostra i campi di input e generati.

Per impostazione predefinita, vengono visualizzate dieci righe. È tuttavia possibile modificare il valore nelle proprietà del flusso. Consultare l'argomento "Impostazione delle opzioni generali per i flussi" a pagina 42 per ulteriori informazioni.

Dal menu Genera è possibile creare diversi tipi di nodi.

Nodi di blocco

Per impedire ad altri utenti di modificare le impostazioni di uno o più nodi di un flusso è possibile incapsulare i nodi in uno speciale tipo di nodo detto Supernodo e quindi bloccare quest'ultimo proteggendolo con una password.

Utilizzo dei flussi

La connessione dei nodi origine, di processo e finali determina la creazione di un flusso. I flussi non sono altro che una raccolta di nodi e possono pertanto essere salvati, annotati e aggiunti ai progetti. È inoltre possibile impostare numerose opzioni per i flussi, per esempio le opzioni per l'ottimizzazione, le impostazioni di data e ora, i parametri e gli script. Queste proprietà verranno illustrate nelle sezioni seguenti.

In IBM SPSS Modeler è possibile utilizzare e modificare più flussi di dati nella stessa sessione di IBM SPSS Modeler. Nella parte destra della finestra principale è disponibile il riquadro dei manager, che semplifica gli spostamenti nei flussi, negli output e nei modelli aperti. Se il riquadro dei manager non è visibile, fare clic su **Manager** nel menu Visualizza e fare clic sulla scheda **Flusso**.

Da questa finestra è possibile:

- Accedere ai flussi.
- Salvare i flussi.
- Salvare i flussi nel progetto corrente.
- Chiudere i flussi.
- Aprire nuovi stream.
- Archiviare e recuperare stream da un repository di IBM SPSS Collaboration and Deployment Services (se disponibile sul sito). Consultare l'argomento "Informazioni su IBM SPSS Collaboration and Deployment Services Repository" a pagina 117 per ulteriori informazioni.

Fare clic con il pulsante destro del mouse su un flusso nella scheda Flusso per accedere a queste opzioni.

Impostazione delle opzioni per i flussi

È possibile specificare un numero di opzioni da applicare al flusso corrente. Le opzioni possono inoltre essere salvate come default da applicare a tutti i flussi. Le opzioni sono le seguenti:

- **Generale.** Opzioni varie come simboli e codifica del testo da utilizzare nel flusso. Consultare l'argomento "Impostazione delle opzioni generali per i flussi" a pagina 42 per ulteriori informazioni.
- **Data/Ora.** Opzioni relative al formato delle espressioni di data e ora. Consultare l'argomento "Impostazione delle opzioni di data e ora per i flussi" a pagina 43 per ulteriori informazioni.
- **Formati numero.** Opzioni che controllano il formato delle espressioni numeriche. Consultare l'argomento "Impostazione delle opzioni relative al formato numerico per i flussi" a pagina 44 per ulteriori informazioni.
- **Ottimizzazione.** Opzioni relative all'ottimizzazione delle performance del flusso. Consultare l'argomento "Impostazione delle opzioni di ottimizzazione per i flussi" a pagina 44 per ulteriori informazioni.

- **Registrazione e stato.** Opzioni che controllano la registrazione SQL e lo stato dei record. Consultare l'argomento "Impostazione delle opzioni relative a registrazione SQL e stato dei record per i flussi" a pagina 46 per ulteriori informazioni.
- **Layout.** Opzioni relative al layout del flusso nell'area. Consultare l'argomento "Impostazione delle opzioni di layout per i flussi" a pagina 46 per ulteriori informazioni.

Per impostare le opzioni dei flussi

1. Nel menu File fare clic su **Proprietà flusso**. In alternativa, selezionare il flusso dalla scheda Flusso del riquadro dei manager, quindi fare clic con il pulsante destro del mouse e scegliere **Proprietà stream** dal menu di scelta rapida).
2. Fare clic sulla scheda **Opzioni**.

In alternativa, nel menu Strumenti, fare clic su:

Proprietà flusso > Opzioni

Impostazione delle opzioni generali per i flussi: Le opzioni generali sono un insieme di opzioni varie che vengono applicate a diversi aspetti del flusso corrente.

Simbolo decimale. Selezionare la virgola (,) o il punto (.) come separatore decimale.

Simbolo di raggruppamento. Per i formati di visualizzazione numerici, selezionare il simbolo utilizzato per raggruppare valori, per esempio, il punto 3.000. Le opzioni disponibili sono Nessuno, Punto, Virgola, Spazio e Definito in base alle opzioni internazionali (in questo caso verrà utilizzato il valore di default per le opzioni internazionali correnti).

Codifica. Specificare il metodo di default del flusso per la codifica di testo. (*Nota:* Si applica a Var. Nodo origine Testo variabile e al nodo di esportazione File flat. L'impostazione non viene utilizzata da altri nodi. La maggior parte dei file di dati includono informazioni di codifica incorporate). È possibile scegliere l'impostazione di default del sistema o UTF-8. L'impostazione di default del sistema è specificata nel Pannello di controllo di Windows oppure, se si esegue il programma in modalità distribuita, sul computer server. Consultare l'argomento "Supporto Unicode in IBM SPSS Modeler" a pagina 177 per ulteriori informazioni.

Valutazione insieme di regole. Specifica la modalità di valutazione dei modelli di insiemi di regole. Per default, gli insiemi di regole utilizzano l'impostazione **Confronto** per combinare le previsioni delle singole regole e definire la previsione finale. Per assicurarsi che gli insiemi di regole utilizzino per default la regola del primo risultato, selezionare **Primo risultato**. Si noti che questa opzione non si applica ai modelli Elenco di decisioni, che utilizzano sempre il primo risultato definito dall'algoritmo.

Numero massimo di righe visualizzate nell'anteprima dei dati. Specificare il numero di righe da visualizzare quando viene richiesta un'anteprima dei dati di un nodo. Consultare l'argomento "Anteprima dei dati nei nodi" a pagina 40 per ulteriori informazioni.

Numero di membri massimo per campi nominali. Selezionare questa opzione per specificare il numero massimo di membri per i campi nominali (insieme) dopo il quale il tipo di campo diventa **Senza tipo**. Tale opzione è utile quando si lavora con campi nominali di grandi dimensioni. *Nota:* quando il livello di misurazione di un campo è impostato su **Senza tipo**, il suo ruolo viene impostato automaticamente su **Nessuno**. Ciò implica che i campi non sono disponibili per la fase di modeling.

Limita dimensione insieme per la modellazione, Kohonen e Medie K. Selezionare questa opzione per specificare un numero massimo di membri per i campi nominali utilizzati nelle modellazioni di reti Kohonen e K-medie. La dimensione di default è 20, dopo la quale il campo verrà ignorato e verrà visualizzato un avviso che fornisce informazioni sul campo.

Tenere presente che, per la compatibilità, questa opzione è valida anche per il vecchio nodo Rete neurale che è stato sostituito nella versione 14 di IBM SPSS Modeler; alcuni flussi legacy potrebbero ancora contenere questo nodo.

Aggiorna nodi origine durante l'esecuzione. Selezionare questa opzione per aggiornare automaticamente tutti i nodi origine durante l'esecuzione del flusso corrente. Questa operazione è analoga alla selezione del pulsante **Aggiorna** su un nodo origine, ma consente di aggiornare automaticamente tutti i nodi, esclusi i nodi Input utente, per il flusso corrente.

Nota: se si seleziona questa opzione, la cache dei nodi downstream verrà svuotata, anche se i dati non sono stati modificati. Durante l'esecuzione del flusso viene tuttavia eseguito un solo svuotamento. Sarà quindi possibile utilizzare comunque le cache a valle come archivio temporaneo per una singola esecuzione. Per esempio, si supponga che sia stata impostata una cache midstream dopo una complessa operazione di creazione di un nuovo campo e che downstream rispetto al nodo Ricava siano disponibili svariati grafici e report. Durante l'esecuzione del flusso, la cache del nodo Ricava verrà svuotata e riempita nuovamente, ma solo per il primo grafico o report. I nodi finali successivi leggeranno i dati dalla cache del nodo Ricava.

Visualizza etichette di valori e campi nell'output. Visualizza etichette di valori e campi in tabelle, grafici e altre forme di output. Se non esistono etichette, vengono visualizzati i nomi dei campi e i valori dei dati. Le etichette sono disattivate per default; tuttavia, è possibile attivarle o disattivarle singolarmente altrove in IBM SPSS Modeler. È inoltre possibile scegliere di visualizzare etichette nella finestra di output utilizzando un pulsante di attivazione disponibile sulla barra degli strumenti.



Figura 11. Icona della barra degli strumenti utilizzata per attivare le etichette di campi e valori

Visualizza tempi di esecuzione. Dopo l'esecuzione del flusso, visualizza i tempi di esecuzione dei singoli nodi del flusso nella tabella Tempi di esecuzione. Consultare l'argomento "Visualizzazione dei tempi di esecuzione del nodo" a pagina 48 per ulteriori informazioni.

Salva come default. Le opzioni specificate vengono applicate solo al flusso corrente. Fare clic su questo pulsante per impostare queste opzioni come default per tutti i flussi.

Impostazione delle opzioni di data e ora per i flussi: Queste opzioni consentono di specificare il formato da utilizzare per diverse espressioni di data e ora nel flusso corrente.

Importa data/ora come. Specificare se i campi di data/ora verranno memorizzati nel formato di data/ora o se verranno importati come variabili stringa.

Formato data. Selezionare un formato di data da utilizzare per i campi di archiviazione delle date o per le stringhe interpretate come date dalle funzioni di data CLEM.

Formato ora. Selezionare un formato di ora da utilizzare per i campi di archiviazione dell'ora o per le stringhe interpretate come orari dalle funzioni di ora CLEM.

Cambia giorni/min. Per i formati di ora, specificare se le differenze di ora negative devono essere interpretate come riferite al giorno o all'ora precedenti.

Data di riferimento (1 gen). Selezionare gli anni di riferimento (sempre 1 gen) che verranno utilizzati dalle funzioni di data CLEM che operano con una singola data.

Date di 2 cifre a partire da. Specificare l'anno di riferimento per l'aggiunta delle cifre del secolo negli anni a due cifre. Per esempio, se si specifica 1930 come anno di riferimento, si presuppone che il

05/11/02 sia nell'anno 2002. La medesima impostazione utilizzerà il 20esimo secolo per le date successive all'anno 1930. Di conseguenza, si presupporrà che il 05/11/73 sia nell'anno 1973.

Salva come default. Le opzioni specificate vengono applicate solo al flusso corrente. Fare clic su questo pulsante per impostare queste opzioni come default per tutti i flussi.

Impostazione delle opzioni relative al formato numerico per i flussi: Queste opzioni consentono di specificare il formato da utilizzare per diverse espressioni numeriche nel flusso corrente.

Formato di visualizzazione numeri. È possibile scegliere il formato di visualizzazione standard (####.###), scientifico (#.###E+##) o valuta (\$###.###).

Numero di decimali (standard, scientifico, valuta). Per i formati di visualizzazione numerici, specifica il numero di decimali utilizzati per la visualizzazione o la stampa di numeri reali. Questa opzione viene specificata separatamente per ogni formato di visualizzazione.

Calcoli in. Selezionare **Radiani** o **Gradi** per specificare l'unità di misura da utilizzare nelle espressioni CLEM trigonometriche. Consultare l'argomento "Funzioni trigonometriche" a pagina 97 per ulteriori informazioni.

Salva come default. Le opzioni specificate vengono applicate solo al flusso corrente. Fare clic su questo pulsante per impostare queste opzioni come default per tutti i flussi.

Impostazione delle opzioni di ottimizzazione per i flussi: Per ottimizzare le prestazioni dei flussi sono disponibili opzioni di ottimizzazione. Si noti che le impostazioni relative alle prestazioni e all'ottimizzazione in IBM SPSS Modeler Server (se utilizzato) prevalgono sulle impostazioni equivalenti nel client.

Nota: le funzionalità di modellazione nel database e ottimizzazione SQL richiedono che la connettività IBM SPSS Modeler Server venga abilitata sul computer IBM SPSS Modeler. Con questa impostazione attivata, è possibile accedere agli algoritmi di database, restituire codice SQL direttamente da IBM SPSS Modeler, e accedere a IBM SPSS Modeler Server. Per verificare lo stato attuale della licenza, scegliere le seguenti opzioni dal menu IBM SPSS Modeler.

Guida > Informazioni su > Dettagli aggiuntivi

Se la connettività è abilitata, l'opzione **Abilitazione server** viene visualizzata nella scheda Stato della licenza.

Per ulteriori informazioni, consultare l'argomento "Connessione a IBM SPSS Modeler Server" a pagina 12.

Nota: il supporto del push back e dell'ottimizzazione SQL dipende dal tipo di database in uso. Per le ultime informazioni sui database e i driver ODBC che sono supportati e testati per l'uso con IBM SPSS Modeler 16, vedere il sito aziendale dedicato al supporto tecnico all'indirizzo <http://www.ibm.com/support>.

Consenti riscrittura flusso. Selezionare questa opzione per consentire la riscrittura dei flussi in IBM SPSS Modeler. Sono disponibili quattro tipi di riscrittura ed è possibile selezionarne uno o entrambi. La riscrittura dei flussi riordina in background i nodi in un flusso per migliorarne il funzionamento, senza alterare la semantica dei flussi.

- **Ottimizza generazione SQL.** Questa opzione consente di riordinare i nodi nel flusso in modo da permettere di eseguire il push back di un maggior numero di operazioni tramite la generazione SQL per l'esecuzione nel database. Quando trova un nodo che non può essere interpretato in codice SQL, l'ottimizzatore controlla più avanti l'eventuale presenza di nodi downstream che è possibile interpretare in SQL e spostare in modo sicuro davanti al nodo del problema senza compromettere la semantica del flusso. Non solo il database può eseguire le operazioni in modo più efficiente di IBM SPSS Modeler,

ma questi pushback o "rinvii" hanno anche la possibilità di ridurre la dimensione dell'insieme di dati restituito a IBM SPSS Modeler per l'elaborazione. Di conseguenza, si riduce il traffico di rete e si sveltiscono le operazioni stream. Si noti che è necessario selezionare la casella di controllo **Genera SQL** per consentire l'ottimizzazione SQL.

- **Ottimizza espressione CLEM.** Questa opzione abilita l'ottimizzatore per ricercare le espressioni CLEM che possono essere pre-elaborate prima che il flusso sia eseguito in modo da incrementare la velocità di elaborazione. Come semplice esempio, se si ha un'espressione come $\log(\text{salary})$, l'ottimizzatore dovrebbe calcolare il valore effettivo del salario e passarlo per l'elaborazione. Questo può essere utilizzato sia per migliorare il push back SQL che le prestazioni IBM SPSS Modeler Server.
- **Ottimizza esecuzione sintassi.** Questo metodo di riscrittura dei flussi aumenta l'efficienza delle operazioni che incorporano più di un nodo contenente la sintassi IBM SPSS Statistics. L'ottimizzazione si ottiene combinando i comandi di sintassi all'interno di un'unica operazione, invece di eseguirli singolarmente come operazioni separate.
- **Ottimizza altra esecuzione.** Questo metodo di riscrittura dei flussi aumenta l'efficienza delle operazioni (quelle che non possono essere delegate al database). L'ottimizzazione viene raggiunta riducendo la quantità dei dati nel flusso non appena possibile. Pur mantenendo l'integrità dei dati, il flusso viene riscritto per spingere le operazioni più vicino alla sorgente dati, riducendo in tal modo i dati a valle per operazioni dispendiose quali le unioni.

Abilitare elaborazione parallela. Quando l'esecuzione avviene su un computer con più processori, questa opzione consente al sistema di distribuire il carico di lavoro su più processori, con un conseguente miglioramento delle prestazioni. L'utilizzo di più nodi o l'utilizzo dei singoli nodi seguenti possono trarre vantaggio dall'elaborazione parallela: C5.0, Unione (per chiave), Ordina, Raccogli (metodi di rango e tile) e Aggregazione (utilizzando uno o più campi chiave).

Genera SQL. Selezionare questa opzione per attivare la generazione SQL, consentendo di rinviare operazioni del flusso al database tramite codice SQL per generare processi di esecuzione in grado di migliorare le prestazioni. Per migliorare ulteriormente le prestazioni, è inoltre possibile selezionare l'opzione **Ottimizza generazione SQL** per aumentare al massimo il numero di operazioni rinviate al database. Dopo che le operazioni per un nodo sono state rinviate al database, il nodo viene evidenziato dal colore porpora durante l'esecuzione del flusso.

- **Memorizzazione in cache del database.** Per i flussi che generano SQL da eseguire nel database, i dati possono essere memorizzati nella cache midstream in una tabella temporanea nel database anziché nel file system. Insieme all'ottimizzazione SQL, ciò consente di ottenere significativi vantaggi in termini di prestazioni. Per esempio, l'output di un flusso che unisce più tabelle per creare una visualizzazione data mining può essere memorizzato nella cache e riutilizzato in un secondo tempo in base alle esigenze. Con la memorizzazione nella cache del database attivata, è sufficiente fare clic con il pulsante destro del mouse su qualsiasi nodo non terminale per memorizzare i dati nella cache in quel punto e la cache viene creata automaticamente direttamente nel database alla successiva esecuzione del flusso. Ciò consente la generazione di SQL per i nodi downstream, migliorando ulteriormente le prestazioni. In alternativa, se lo si desidera, è possibile disattivare questa opzione, per esempio se i criteri o le autorizzazioni precludono la scrittura dei dati nel database. Se la memorizzazione nella cache del database o l'ottimizzazione SQL non sono attivate, la cache verrà scritta nel file system. Per ulteriori informazioni, consultare "Opzioni di memorizzazione dei dati nella cache per i nodi" a pagina 39.
- **Utilizza conversione di tipo relaxed.** Questa opzione consente la conversione dei dati da stringhe a numeri o da numeri a stringhe, purché siano archiviati in un formato adatto. Ad esempio, se i dati sono conservati nel database sotto forma di stringa, ma contengono un numero significativo, possono essere convertiti per essere utilizzati quando si esegue il push back.

Nota: a causa di piccole differenze di implementazione del codice SQL, i flussi eseguiti in un database potrebbero restituire risultati leggermente diversi da quelli che restituiscono quando vengono eseguiti in IBM SPSS Modeler. Per ragioni analoghe, tali differenze possono variare anche in base al fornitore di database.

Salva come default. Le opzioni specificate vengono applicate solo al flusso corrente. Fare clic su questo pulsante per impostare queste opzioni come default per tutti i flussi.

Impostazione delle opzioni relative a registrazione SQL e stato dei record per i flussi: Queste impostazioni includono alcune opzioni che controllano la visualizzazione delle istruzioni SQL generate dal flusso e la visualizzazione del numero di record elaborati dal flusso.

Visualizza SQL nel registro dei messaggi durante l'esecuzione del flusso. Specifica se l'SQL generato durante l'esecuzione del flusso viene passato al registro del messaggio.

Visualizza SQL nel registro dei messaggi durante la preparazione dei flussi. Durante l'anteprima dei flussi, specifica se un'anteprima di SQL da generare viene passata al registro dei messaggi.

Visualizza SQL. Specifica se l'SQL visualizzato nel registro debba contenere funzioni SQL native o funzioni ODBC standard nel formato {fn FUNC()}, generato da IBM SPSS Modeler. Le prime si basano sulla funzionalità dei driver ODBC che non è possibile implementare. Per esempio, questo controllo non ha effetto su SQL Server.

Riformattare SQL per una migliore leggibilità. Specifica se l'SQL visualizzato nel registro debba essere formattato per migliorarne la leggibilità.

Mostra stato dei record. Specifica quando devono essere registrati i record al loro arrivo ai nodi terminali. Specificare il numero utilizzato per aggiornare lo stato ogni *N* record.

Salva come default. Le opzioni specificate vengono applicate solo al flusso corrente. Fare clic su questo pulsante per impostare queste opzioni come default per tutti i flussi.

Impostazione delle opzioni di layout per i flussi: Queste impostazioni forniscono alcune opzioni relative alla visualizzazione e all'uso dell'area del flusso.

Larghezza minima area flusso. Specificare la larghezza minima in pixel dell'area del flusso.

Altezza minima area flusso. Specificare l'altezza minima in pixel dell'area del flusso.

Velocità di scorrimento flusso. Specificare la velocità di scorrimento dell'area del flusso per controllare la velocità a cui il riquadro dell'area del flusso scorre quando un nodo viene trascinato da un punto all'altro dell'area. Più alti sono i valori, maggiore è la velocità di scorrimento.

Massima lunghezza nome icona. Specificare un numero massimo di caratteri per i nomi dei nodi dell'area del flusso.

Dimensione icona. Selezionare un'opzione per ridimensionare l'intera visualizzazione del flusso su una scala compresa tra l'8% e il 200% della dimensione dell'icona standard.

Dimensione celle griglia. Selezionare una dimensione delle celle della griglia dall'elenco. Il valore specificato verrà utilizzato per l'allineamento dei nodi nell'area del flusso mediante una griglia invisibile. La dimensione di default per le celle della griglia è 0,25.

Blocca sulla griglia. Selezionare questa opzione per allineare le icone su una griglia invisibile (opzione di default).

Posizione icona generata. Scegliere il punto dell'area in cui posizionare le icone relative ai nodi generati dai nugget del modello. L'impostazione predefinita è in alto a sinistra.

Salva come default. Le opzioni specificate vengono applicate solo al flusso corrente. Fare clic su questo pulsante per impostare queste opzioni come default per tutti i flussi.

Proprietà del flusso Analytic Server: Tali impostazioni forniscono un numero di opzioni per l'utilizzo di Analytic Server.

Numero massimo di record da elaborare all'esterno di Analytic Server

Specifica il numero massimo di record da importare nel server SPSS Modeler da un'origine dati Analytic Server.

Notifica quando un nodo non può essere elaborato in Analytic Server

Questa impostazione determina cosa accade quando un flusso da inoltrare a Analytic Server contiene un nodo che non può essere elaborato in Analytic Server. Specifica se inviare un avviso e proseguire l'elaborazione del flusso o se inviare un errore ed arrestarla.

Impostazioni di archiviazione del modello di suddivisione

Archivia i modelli di suddivisione in base al riferimento in Analytic Server quando la dimensione (MB) del modello è eccessiva

I nugget del modello vengono solitamente archiviati come parte del flusso. I modelli di suddivisione con molte suddivisioni possono dar luogo a nugget di grandi dimensioni, ed il loro ripetuto spostamento tra il flusso e Analytic Server, potrebbe avere un impatto sulle prestazioni. Come soluzione, quando un modello supera una dimensione specificata, esso viene archiviato nell'Analytic Server, ed il nugget in SPSS Modeler contiene un riferimento al modello.

Cartella di default per l'archiviazione di modelli in base al riferimento su Analytic Server una volta completata l'esecuzione.

Specifica il percorso di default in cui si desidera archiviare i modelli di suddivisione su Analytic Server. Il percorso deve iniziare con un nome di progetto Analytic Server valido.

Cartella per archiviare i modelli promossi

Specifica il percorso predefinito in cui archiviare i modelli "promossi". Un modello promosso non viene ripulito quando si conclude la sessione SPSS Modeler.

Visualizzazione dei messaggi relativi alle operazioni del flusso

I messaggi relativi alle operazioni del flusso, per esempio l'esecuzione, l'ottimizzazione e il tempo richiesto dalla creazione del modello e dalla valutazione possono essere facilmente visualizzati mediante la scheda Messaggi della finestra di dialogo delle proprietà del flusso. I messaggi di errore vengono inoltre riportati nella tabella.

Per visualizzare i messaggi relativi al flusso

1. Nel menu File fare clic su **Proprietà flusso**. In alternativa, selezionare il flusso dalla scheda Flusso del riquadro dei manager, quindi fare clic con il pulsante destro del mouse e scegliere **Proprietà stream** dal menu di scelta rapida).
2. Fare clic sulla scheda **Messaggi**.

In alternativa, nel menu Strumenti, fare clic su:

Proprietà flusso > Messaggi

Oltre ai messaggi relativi alle operazioni di sistema, nella scheda sono visualizzati anche i messaggi di errore. Se l'esecuzione del flusso viene interrotta a causa di un errore, verrà visualizzata la scheda Messaggi di questa finestra di dialogo con il messaggio di errore specifico. Inoltre, il nodo con gli errori sarà evidenziato in rosso nell'area del flusso.

Se le opzioni di ottimizzazione SQL e registrazione sono attivate nella finestra di dialogo Opzioni utente, vengono visualizzate anche le informazioni sul codice SQL generato. Consultare l'argomento "Impostazione delle opzioni di ottimizzazione per i flussi" a pagina 44 per ulteriori informazioni.

È possibile salvare i messaggi relativi a un flusso facendo clic su **Salva messaggi** nell'elenco a discesa del pulsante di salvataggio (a sinistra, subito sotto la scheda Messaggi). È inoltre possibile cancellare tutti i messaggi per un flusso specifico facendo clic su **Cancella tutti i messaggi** nello stesso elenco.

Visualizzazione dei tempi di esecuzione del nodo

Nella scheda Messaggi è possibile scegliere di visualizzare i tempi di esecuzione, laddove è possibile visualizzare singoli tempi di esecuzione per tutti i nodi nel flusso che sono in esecuzione in IBM SPSS Modeler Server. considerare che i tempi potrebbero non essere precisi per l'esecuzione dei flussi in altre aree come ad esempio R o Analytic Server.

Nota: affinché questa funzione possa operare, la casella **Visualizza tempi di esecuzione** deve essere selezionata nella sezione **Generale** della scheda **Opzioni**.

La tabella dei tempi di esecuzione dei nodi contiene le colonne seguenti. Fare clic sull'intestazione di una colonna per ordinarne il contenuto in modo crescente o decrescente (ad esempio, per vedere quali nodi presentano i tempi di esecuzione più lunghi).

Nodo terminale. L'identificatore del ramo a cui appartiene il nodo. L'identificatore è il nome del nodo finale alla fine del ramo.

Etichetta nodo. Nome del nodo a cui si riferisce il tempo di esecuzione.

id nodo. Identificatore univoco del nodo a cui si riferisce il tempo di esecuzione. L'identificatore è generato dal sistema al momento della creazione del nodo.

Tempi di esecuzione. Il tempo, espresso in secondi, che ha richiesto l'esecuzione di questo nodo.

Impostazione di parametri di flusso e di sessione

È possibile definire parametri da utilizzare in espressioni CLEM e negli script. Si tratta a tutti gli effetti di variabili definite dall'utente che vengono salvate e mantenute con il flusso, la sessione o il Supernodo corrente e a cui è possibile accedere dall'interfaccia utente nonché attraverso script. Per esempio, durante il salvataggio di un flusso, qualsiasi parametro impostato per tale stream viene salvato contemporaneamente. In questo modo i parametri vengono distinti dalle variabili di script locali, utilizzabili solo nello script in cui sono dichiarate. I parametri vengono spesso utilizzati nello script per controllare il comportamento dello script, fornendo informazioni sui campi e i valori che non necessitano di essere attentamente codificati nello script.

L'ambito di un parametro dipende dalla posizione in cui è impostato:

- i parametri impostati per un flusso possono essere impostati in uno script del flusso o nella finestra di dialogo delle proprietà del flusso e sono disponibili a tutti i nodi del flusso. Vengono visualizzati nell'elenco Parametri nel generatore di espressioni.
- I parametri di sessione possono essere impostati in uno script autonomo oppure nella finestra di dialogo dei parametri di sessione e sono disponibili per tutti i flussi utilizzati nella sessione corrente, ovvero per tutti i flussi elencati nella scheda Flusso del riquadro dei manager.

È inoltre possibile impostare parametri per Supernodi. In questo caso, sono visibili solo ai nodi incapsulati all'interno del Supernodo.

Per impostare parametri di stream e di sessione tramite l'interfaccia utente

1. Per impostare i parametri dei flussi, nel menu principale fare clic su:

Strumenti > Proprietà flusso > Parametri

2. Per impostare parametri di sessione, fare clic su **Imposta parametri di sessione** nel menu Strumenti.

Prompt?. Selezionare questa casella per far sì che il sistema chieda all'utente di inserire il valore di questo parametro al momento dell'esecuzione.

Nome. In questa colonna sono visualizzati i nomi dei parametri. È possibile creare un nuovo parametro inserendo nel campo un nome, Per esempio, per creare un parametro per la temperatura minima, è possibile digitare `valminimo`. Non includere il prefisso `$P-` che indica un parametro nelle espressioni CLEM. Tale nome è inoltre quello visualizzato nel Generatore espressioni CLEM.

Nome lungo. Elenca il nome descrittivo di ogni parametro creato.

Archiviazione. Selezionare un tipo di archiviazione dall'elenco. La classe archiviazione indica come sono stati archiviati i valori dei dati nel parametro. Per esempio, quando si utilizzano valori contenenti zero iniziali che si desidera mantenere (per esempio 008), è necessario selezionare **Stringa** come tipo di archiviazione. In caso contrario, gli zeri verranno rimossi dal valore. I tipi di archiviazione disponibili sono stringa, intero, reale, ora, data e timestamp. Per i parametri data, tenere presente che è necessario specificare i valori con la notazione standard ISO, come illustrato di seguito.

Valore. Elenca il valore corrente di ogni parametro. Modificare il parametro in base alle esigenze specifiche. Si noti che per specificare i parametri data è necessario utilizzare la notazione standard ISO (ovvero, AAAA-MM-GG). Le date specificate in altri formati non sono accettate.

Tipo (facoltativo). Se si pianifica la distribuzione del flusso in un'applicazione esterna, selezionare un livello di misurazione dall'elenco. In caso contrario, si consiglia di non modificare la colonna *Tipo*. Se si desidera specificare dei vincoli per il valore del parametro quale un limite superiore e inferiore per un intervallo numerico, selezionare **Specificata** dall'elenco.

Si noti che è possibile impostare le opzioni nome lungo, classe archiviazione e tipo per i parametri solo tramite l'interfaccia utente e non tramite script.

Fare clic sulle frecce a destra per spostare il parametro selezionato più in alto o in basso nell'elenco dei parametri. Per rimuovere il parametro selezionato, utilizzare il pulsante di eliminazione contrassegnato da una X.

Definizione dei prompt di runtime per i valori dei parametri

In presenza di stream in cui è necessario inserire valori diversi per lo stesso parametro a seconda dei casi, è possibile specificare dei prompt di runtime per uno o più valori di parametri di stream o di sessione.

Parametri. (Facoltativo) Inserire un valore per il parametro o lasciare il valore di default se già indicato.

Disattiva questi prompt. Selezionare questa casella per non visualizzare questi prompt quando si esegue il flusso. È possibile fare in modo che vengano visualizzati nuovamente selezionando la casella di controllo **Prompt?** nella finestra di dialogo delle proprietà del flusso o della sessione in cui sono stati definiti i parametri. Consultare l'argomento "Impostazione di parametri di flusso e di sessione" a pagina 48 per ulteriori informazioni.

Definizione di vincoli per i valori di un tipo di parametro

È possibile rendere disponibili dei vincoli per i valori di un parametro durante la distribuzione dei flussi a un'applicazione esterna che legge i flussi di modellazione dei dati. Questa finestra di dialogo consente di specificare i valori disponibili per un utente esterno che esegue il flusso. A seconda del tipo di dati, i vincoli relativi ai valori cambiano dinamicamente nella finestra di dialogo. Le opzioni illustrate sono identiche a quelle disponibili per i valori relativi al nodo Tipo.

Tipo. Visualizza il livello di misurazione selezionato. Questo valore si può modificare per indicare il modo in cui si intende utilizzare il parametro in IBM SPSS Modeler.

Archiviazione. Visualizza il tipo di archiviazione, se noto. I tipi di archiviazione non sono influenzati dal livello di misurazione (continuo, nominale o flag) scelto in IBM SPSS Modeler. È possibile modificare il tipo di archiviazione nella scheda Parametri principale.

La parte inferiore della finestra di dialogo viene modificata dinamicamente, a seconda del livello di misurazione selezionato nel campo **Tipo**.

Livelli di misurazione continui

Ultimo. Specificare un limite inferiore per i valori dei parametri.

Superiore. Specificare un limite superiore per i valori dei parametri.

Etichette. È possibile definire le etichette per qualsiasi valore di un campo intervallo. Fare clic sul pulsante **Etichette** per aprire una finestra di dialogo nella quale è possibile definire le etichette dei valori.

Livelli di misurazione nominali

Valori. Questa opzione consente di specificare i valori di un parametro che verrà utilizzato come campo nominale. I valori non verranno forzati nel flusso di IBM SPSS Modeler, ma verranno utilizzati in un elenco a discesa per le applicazioni di distribuzione esterne. I pulsanti freccia e quello contrassegnato da una X consentono di modificare l'ordine dei valori esistenti o di eliminarli.

Livelli di misurazione flag

Vero. Specificare un valore flag per il parametro quando la condizione viene soddisfatta.

Falso. Specificare un valore flag per il parametro quando la condizione non viene soddisfatta.

Etichette. È possibile definire le etichette per i valori di un campo flag.

Opzioni di distribuzione dei flussi

La scheda Distribuzione della finestra di dialogo delle proprietà del flusso consente di specificare le opzioni per la distribuzione del flusso come scenario in IBM SPSS Collaboration and Deployment Services ai fini dell'aggiornamento dei modelli, della pianificazione automatica dei lavori o di ulteriore uso da parte di IBM Analytical Decision Management. Per poter essere sottoposti a deployment, tutti i flussi necessitano di un ramo di calcolo del punteggio designato; gli altri requisiti e opzioni dipendono dal tipo di distribuzione. Consultare l'argomento "Archiviazione e deployment di oggetti di repository" a pagina 118 per ulteriori informazioni.

Esecuzione dei cicli dei flussi

Mediante la scheda Esecuzione della finestra di dialogo delle proprietà dello stream, è possibile impostare le condizioni di loop per automatizzare le attività ripetitive nello stream corrente.

Una volta definite tali condizioni è possibile utilizzarla come un'introduzione agli script durante il popolamento della finestra dello script con script di base per il flusso che è possibile modificare, ad esempio per utilizzarlo come base di creazione di script migliori. Consultare l'argomento "Funzioni globali" a pagina 113 per ulteriori informazioni.

Istruzioni per l'impostazione dell'esecuzione di cicli per un flusso

1. Nel menu File fare clic su **Proprietà flusso**. In alternativa, selezionare il flusso dalla scheda Flussi del riquadro dei manager, quindi fare clic con il pulsante destro del mouse e scegliere **Proprietà stream** dal menu di scelta rapida).
2. Fare clic sulla scheda **Esecuzione**.
3. Selezionare la modalità di esecuzione **Esecuzione ciclica/condizionale**.

4. Fare clic sulla scheda **Ciclo**.

In alternativa, nel menu Strumenti, fare clic su:

Proprietà flusso > Esecuzione

Come ulteriore alternativa, fare clic con il tasto destro del mouse sul nodo e dal menu di scelta rapida, fare clic su:

Esecuzione ciclica/condizionale > Modifica impostazioni cicliche

Iterazione. Non è possibile modificare questo valore di numero di righe, ma è possibile aggiungere, eliminare o spostare un'iterazione in alto o in basso utilizzando i pulsanti sulla destra della tabella.

Intestazioni tabella. Questi riportano la chiave di iterazione e qualsiasi variabile di iterazione che è stata creata quando è stato impostato il loop.

Visualizzazione dei valori globali dei flussi

Mediante la scheda Valori globali della finestra di dialogo delle proprietà del flusso è possibile visualizzare i valori globali impostati per il flusso corrente. Per creare i valori globali viene utilizzato un nodo Calcola globali, che consente di determinare le statistiche quali la media, la somma o la deviazione standard per i campi selezionati.

Dopo avere eseguito il nodo Calcola globali, sarà possibile utilizzare i valori ottenuti in molti modi diversi nelle operazioni del flusso. Consultare l'argomento "Funzioni globali" a pagina 113 per ulteriori informazioni.

Per visualizzare i valori globali per un flusso

1. Nel menu File fare clic su **Proprietà flusso**. In alternativa, selezionare il flusso dalla scheda Flusso del riquadro dei manager, quindi fare clic con il pulsante destro del mouse e scegliere **Proprietà flusso** dal menu di scelta rapida).
2. Fare clic sulla scheda **Valori globali**.

In alternativa, nel menu Strumenti, fare clic su:

Proprietà flusso > Valori globali

Valori globali disponibili. Nella tabella sono elencati i valori globali disponibili. Non è possibile modificare i valori globali qui, ma è possibile cancellare tutti i valori globali per un flusso mediante il pulsante Cancella tutti i valori disponibile a destra della tabella.

Ricerca di nodi in un flusso

È possibile effettuare ricerche di nodi in un flusso specificando una serie di criteri di ricerca, per esempio nome, categoria e identificatore del nodo. Questa funzione risulta particolarmente utile per stream complessi contenenti un numero elevato di nodi.

Per effettuare ricerche di nodi in un flusso

1. Nel menu File fare clic su **Proprietà flusso**. In alternativa, selezionare il flusso dalla scheda Flusso del riquadro dei manager, quindi fare clic con il pulsante destro del mouse e scegliere **Proprietà stream** dal menu di scelta rapida).
2. Fare clic sulla scheda **Ricerca**.

In alternativa, nel menu Strumenti, fare clic su:

Proprietà flusso > Ricerca

È possibile specificare più di un'opzione per limitare la ricerca. Utilizzando come criterio di ricerca l'ID del nodo (campo **ID uguale a**), tuttavia, vengono escluse le altre opzioni.

L'etichetta del nodo contiene. Selezionare questa casella di controllo e immettere l'etichetta di un nodo (completa o parte di essa) per eseguire la ricerca di quel determinato nodo. Le ricerche non distinguono i caratteri maiuscoli/minuscoli, mentre le parole multiple vengono considerate come un testo unico.

Categoria nodo. Selezionare questa casella di controllo e fare clic su una categoria nell'elenco per cercare un determinato tipo di nodo. **Nodo di elaborazione** indica un nodo della scheda Oper su record o Oper su campi della palette dei nodi; **Applica nodo modello** fa riferimento a un nugget del modello.

Le parole chiave comprendono. Selezionare questa casella di controllo e immettere una o più parole chiave complete per cercare i nodi che presentano quel testo nel campo Parole chiave nella scheda Annotazioni della finestra di dialogo del nodo. Il testo delle parole chiave inserite deve essere una corrispondenza esatta. Separare le parole chiave multiple con punto e virgola in modo da cercare alternative (per esempio digitando proton;neutron si troveranno tutti i nodi con una di queste due parole chiave). Consultare l'argomento "Annotazioni" a pagina 59 per ulteriori informazioni.

L'annotazione contiene. Selezionare questa casella di controllo e immettere una o più parole per cercare i nodi che contengono il testo in questione nell'area di testo principale della scheda Annotazioni della finestra di dialogo del nodo. Le ricerche non distinguono i caratteri maiuscoli/minuscoli, mentre le parole multiple vengono considerate come un testo unico. Consultare l'argomento "Annotazioni" a pagina 59 per ulteriori informazioni.

Genera un campo denominato. Selezionare questa casella di controllo e immettere il nome di un campo generato (ad esempio, \$C-Cura). Questa opzione consente di cercare nodi Modelli che generano uno specifico campo. Immettere solo un nome di campo, che deve essere una corrispondenza esatta.

ID uguale a. Selezionare questa casella di controllo e immettere l'ID di un nodo per cercare uno specifico nodo con tale identificatore (selezionando questa opzione vengono deselectionate tutte quelle precedenti). Gli ID dei nodi vengono assegnati dal sistema alla creazione del nodo e possono essere utilizzati come riferimenti al nodo per operazioni legate agli script o di automazione. Immettere solo un ID di nodo, che deve essere una corrispondenza esatta. Consultare l'argomento "Annotazioni" a pagina 59 per ulteriori informazioni.

Cerca nei supernodi. Questa casella di controllo è selezionata di default. La ricerca viene pertanto effettuata nei nodi sia all'interno sia all'esterno dei Supernodi. Deselezionare la casella di controllo se si desidera che la ricerca venga effettuata solo nei nodi al di fuori dei Supernodi, al livello più elevato del flusso.

Trova. Dopo aver specificato tutte le operazioni desiderate, fare clic su questo pulsante per avviare la ricerca.

I nodi che corrispondono alle opzioni specificate vengono elencati nella parte inferiore della finestra di dialogo. Selezionare un nodo dell'elenco per evidenziarlo nell'area del flusso.

Rinomina dei flussi

Nella scheda Annotazioni della finestra di dialogo delle proprietà del flusso è possibile aggiungere annotazioni descrittive e creare un nome personalizzato per il flusso. Queste opzioni sono particolarmente utili durante la creazione di report per i flussi aggiunti al riquadro dei progetti. Consultare l'argomento "Annotazioni" a pagina 59 per ulteriori informazioni.

Descrizioni flusso

Per ogni stream che viene creato, IBM SPSS Modeler produce una descrizione contenente informazioni sul contenuto del flusso. Questa opzione può essere utile se si sta cercando di capire quali funzioni

vengono svolte da un flusso ma non è stato installato IBM SPSS Modeler; per esempio, quando si accede a un flusso mediante IBM SPSS Collaboration and Deployment Services.

La descrizione del flusso viene visualizzata sotto forma di documento HTML formato da una serie di sezioni.

Informazioni generali sul flusso

Questa sezione contiene il nome del flusso e i dettagli relativi alla creazione e all'ultimo salvataggio del flusso.

Descrizione e commenti

Questa sezione include qualsiasi:

- Annotazioni flusso (vedere “Annotazioni” a pagina 59)
- Commenti non collegati a nodi specifici
- Commenti collegati a nodi dei rami di modellazione e calcolo del punteggio del flusso

Informazioni sul calcolo del punteggio

Questa sezione contiene informazioni, raggruppate sotto diverse intestazioni, relative al ramo del calcolo del punteggio del flusso.

- **Commenti.** Include i commenti legati solo ai nodi nel ramo di calcolo del punteggio.
- **Input.** Elenca i campi di input con i relativi tipi di archiviazione (per esempio, stringa, numero intero, reale, e così via).
- **Output.** Elenca i campi di output, inclusi i campi aggiuntivi generati dal nodo Modelli, nonché i relativi tipi di archiviazione.
- **Parametri.** Elenca i parametri relativi al ramo di calcolo del punteggio del flusso, che possono essere visualizzati o modificati ogni volta che viene calcolato il punteggio del modello. Questi parametri vengono identificati quando si fa clic sul pulsante **Parametri di calcolo del punteggio** nella scheda **Deployment** della finestra di dialogo Proprietà stream.
- **Nodo modellazione.** Mostra il nome e il tipo di modello (per esempio Rete neurale, C&R Tree e così via). Questo è il nugget del modello selezionato per il campo **nodo Modello** nella scheda **Deployment** della finestra di dialogo delle proprietà del flusso.
- **Dettagli del modello.** Mostra i dettagli del nugget del modello identificato sotto l'intestazione precedente. Dove possibile, vengono inclusi l'importanza dei predittori e i grafici di valutazione relativi al modello.

Informazioni sul modello

Contiene informazioni relative al ramo di modellazione del flusso.

- **Commenti.** Elenca gli eventuali commenti o annotazioni collegati ai nodi nel ramo di modellazione.
- **Input.** Elenca i campi di input e i relativi ruoli nel ramo di modellazione (sotto forma di valore del ruolo dei campi, per esempio Input, Obiettivo, Suddivisione e così via).
- **Parametri.** Elenca i parametri relativi al ramo di modellazione del flusso, che possono essere visualizzati o modificati ogni volta che il modello viene aggiornato. Questi parametri vengono identificati quando si fa clic sul pulsante **Parametri di creazione del modello** nella scheda **Deployment** della finestra di dialogo delle proprietà del flusso.
- **Nodo modellazione.** Mostra il nome e il tipo del nodo Modelli utilizzato per generare o aggiornare il modello.

Anteprima delle descrizioni dei flussi

Per visualizzare il contenuto della descrizione di un flusso in un browser Web, è possibile fare clic su un'opzione nella finestra di dialogo delle proprietà del flusso. Il contenuto della descrizione dipende dalle opzioni specificate nella scheda Distribuzione della finestra di dialogo. Consultare l'argomento "Opzioni di distribuzione dei flussi" a pagina 131 per ulteriori informazioni.

Per visualizzare la descrizione di un flusso:

1. Nel menu principale di IBM SPSS Modeler, fare clic su:
Strumenti > Proprietà flusso > Distribuzione
2. Impostare il tipo di distribuzione, il nodo designato per il calcolo del punteggio e i parametri di calcolo del punteggio.
3. Se il tipo di distribuzione è Aggiornamento modelli, è possibile, se lo si desidera, selezionare un:
 - Nodo Modelli e i parametri di creazione del modello
 - Nugget del modello sul ramo di calcolo del punteggio del flusso
4. Fare clic sul pulsante **Anteprima descrizione flusso**.

Esportazione di descrizioni flusso

Il contenuto della descrizione del flusso può essere esportato in un file HTML.

Per esportare la descrizione di un flusso:

1. Nel menu principale, fare clic su:
File > Esporta descrizione flusso
2. Immettere un nome per il file HTML e fare clic su **Salva**.

Esecuzione dei flussi

Dopo avere specificato le opzioni richieste per i flussi e avere connesso i nodi, è possibile eseguire il flusso mediante l'elaborazione dei dati attraverso i nodi del flusso. Esistono diversi modi per eseguire un flusso all'interno di IBM SPSS Modeler. È possibile:

- Fare clic su **Esegui** nel menu Strumenti.
- Fare clic su uno dei pulsanti **Esegui...** sulla barra degli strumenti. Questi pulsanti consentono di eseguire l'intero stream o semplicemente il nodo terminale selezionato. Consultare l'argomento "IBM SPSS Modeler Barra degli strumenti" a pagina 19 per ulteriori informazioni.
- Eseguire un singolo flusso di dati facendo clic con il pulsante destro del mouse su un nodo terminale e scegliendo **Esegui** dal menu di scelta rapida.
- Eseguire parte di un flusso di dati facendo clic con il pulsante destro del mouse su un nodo non terminale e scegliendo **Esegui da qui** dal menu di scelta rapida. Questa operazione determina l'esecuzione soltanto delle operazioni successive al nodo selezionato.

Per interrompere l'esecuzione di un flusso, è possibile fare clic sul pulsante di colore rosso Interrompi disponibile sulla barra degli strumenti oppure fare clic su **Interrompi esecuzione** nel menu Strumenti.

Se l'esecuzione di un flusso richiede oltre 3 secondi, viene visualizzata la finestra di dialogo Feedback di esecuzione che mostra l'avanzamento dell'esecuzione.

Per alcuni nodi vengono visualizzate informazioni aggiuntive sull'esecuzione del flusso. Per visualizzarle, è necessario selezionare la riga corrispondente nella finestra di dialogo. La prima riga è selezionata automaticamente.

Utilizzo dei modelli

Se in un flusso è presente un nodo Modelli (cioè un nodo contenuto nella scheda Modelli o Modelli database della palette dei nodi) quando si esegue il flusso viene creato un **nugget del modello**. Un

nugget del modello è un contenitore per un **modello**, ovvero per l'insieme di regole, formule o equazioni che consentono di generare previsioni mediante i dati di input e che costituiscono il fulcro dell'analisi predittiva.



Figura 12. Nugget del modello

Quando si esegue un nodo Modelli, nell'area del flusso viene inserito un nugget del modello corrispondente, rappresentato da un'icona che raffigura un diamante dorato. È possibile aprire il nugget per esaminarne il contenuto e visualizzare i dettagli del modello. Per visualizzare le previsioni, collegare ed eseguire uno o più nodi terminali, il cui output presenta le previsioni in formato leggibile.

Un tipico stream di modellazione è composto da due rami. Il **ramo di modellazione** contiene il nodo Modelli con i nodi origine e di elaborazione che lo precedono. Il **ramo di calcolo del punteggio** viene creato quando si esegue il nodo di modellazione e contiene il nugget del modello e il nodo o i nodi terminali da utilizzare per visualizzare le previsioni.

Per ulteriori informazioni, fare riferimento al documento *IBM SPSS Modeler Modeling Nodes guide*.

Aggiunta di commenti e annotazioni a nodi e flussi

Può essere necessario descrivere un flusso ad altri utenti della propria azienda. A questo scopo è possibile allegare commenti esplicativi a stream, nodi e nugget del modello.

Gli altri utenti possono visualizzare questi commenti sullo schermo oppure è possibile stampare un'immagine del flusso che contenga i commenti.

È possibile elencare tutti i commenti relativi a un flusso o a un Supernodo, cambiare l'ordine dei commenti nell'elenco, modificare il testo di commento e cambiare il colore di primo piano o di sfondo di un commento. Consultare l'argomento "Elenco di commenti su flussi" a pagina 58 per ulteriori informazioni.

È inoltre possibile aggiungere note sotto forma di annotazioni di testo a stream, nodi e insiemi, per mezzo della scheda Annotazioni della finestra di dialogo delle proprietà di un flusso o di un nodo o da una finestra di nugget del modello. Queste note sono visibili solo quando la scheda Annotazioni è aperta; tuttavia, le annotazioni ai flussi possono anche essere visualizzate come commenti su schermo. Consultare l'argomento "Annotazioni" a pagina 59 per ulteriori informazioni.

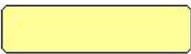
Commenti

I commenti hanno la forma di caselle di testo in cui è possibile inserire testo e commenti senza limitazioni. Un commento può essere indipendente (non collegato a nessun oggetto del flusso) oppure collegato a uno o più nodi o nugget del modello del flusso. I commenti indipendenti sono utilizzati solitamente per descrivere le finalità generali del flusso; i commenti collegati descrivono il nodo o il nugget a cui sono associati. A nodi e nugget può essere associato più di un commento, mentre il flusso può avere un numero illimitato di commenti indipendenti.

Note: le annotazioni dei flussi possono essere visualizzate anche come commenti su schermo, ma questi non possono essere collegati né a nodi, né a nugget. Consultare l'argomento "Conversione di annotazioni in commenti" a pagina 59 per ulteriori informazioni.

L'aspetto della casella di testo cambia a indicare la modalità corrente del commento (o dell'annotazione visualizzata come commento), come illustrato dalla seguente tabella.

Tabella 3. Modalità caselle di testo commenti e annotazioni

Casella di testo commento	Casella di testo annotazione	Modalità	Indica	Ottenuto da...
		Modifica	Il commento può essere modificato.	Creare un nuovo commento o annotazione o fare doppio clic su un commento o un'annotazione esistente.
		Ultimo selezionato	Il commento può essere spostato, ridimensionato o eliminato.	Fare clic sullo sfondo del flusso dopo la modifica o fare clic su un commento o un'annotazione esistente.
		Visualizza	La modifica è terminata.	Dopo avere apportato le modifiche, fare clic su un altro nodo, commento o annotazione.

Quando si crea un nuovo commento indipendente, viene inizialmente visualizzato nell'angolo superiore sinistro dell'area del flusso.

Quando si associa un commento a un nodo o a un nugget, questo viene visualizzato inizialmente al di sopra dell'oggetto del flusso a cui è associato.

Quando la casella di testo è bianca significa che è possibile immettere il testo. Una volta digitato il testo, fare clic all'esterno della casella di testo. Lo sfondo del commento diventa giallo indicando che l'inserimento del testo è concluso. Il commento rimane selezionato consentendo all'utente di spostarlo, ridimensionarlo o eliminarlo.

Quando si fa clic nuovamente, il bordo diventa una linea continua per indicare che la modifica è terminata.

Se si fa doppio clic su un commento, la casella di testo entra in modalità di modifica, lo sfondo diventa bianco e il testo del commento può essere modificato.

È inoltre possibile associare commenti ai Supernodi.

Operazioni eseguibili sui commenti: Sui commenti è possibile eseguire una serie di operazioni. È possibile:

- Aggiungere un commento indipendente
- Associare un commento a un nodo o a un nugget
- Modificare un commento
- Ridimensionare un commento
- Spostare un commento
- Disconnettere un commento
- Eliminare un commento
- Visualizzare o nascondere tutti i commenti a un flusso

Per aggiungere un commento indipendente

1. Controllare che nel flusso non vi siano selezioni attive.
2. Eseguire una delle seguenti operazioni:
 - Nel menu principale, fare clic su:

Inserisci > Nuovo commento

- Fare clic con il pulsante destro del mouse sullo sfondo del flusso e scegliere **Nuovo commento** nel menu di scelta rapida.
 - Fare clic sul pulsante **Nuovo commento** nella barra degli strumenti.
3. Inserire il testo del commento (o incollarlo dagli Appunti).
 4. Fare clic su un nodo del flusso per salvare il commento.

Per associare un commento a un nodo o a un nugget

1. Selezionare uno o più nodi o nugget nell'area del flusso.
2. Eseguire una delle seguenti operazioni:
 - Nel menu principale, fare clic su:
Inserisci > Nuovo commento
 - Fare clic con il pulsante destro del mouse sullo sfondo del flusso e scegliere **Nuovo commento** nel menu di scelta rapida.
 - Fare clic sul pulsante **Nuovo commento** nella barra degli strumenti.
3. Immettere il testo del commento.
4. Fare clic su un altro nodo del flusso per salvare il commento.
In alternativa è possibile:
5. Inserire un commento indipendente (vedere sezione precedente).
6. Eseguire una delle seguenti operazioni:
 - Selezionare il commento, premere F2 e selezionare il nodo o il nugget.
 - Selezionare il nodo o il nugget, premere F2 e selezionare il commento.
 - (Solo per mouse a tre pulsanti) Portare il puntatore del mouse sul commento, tenere premuto il pulsante centrale e trascinare il puntatore sul nodo o sul nugget, quindi rilasciare il pulsante.

Per associare un commento a un nodo o a un nugget aggiuntivo

Se un commento è già allegato a un nodo o a un nugget o se si trova attualmente a livello del flusso e si desidera associarlo a un altro nodo o insieme, eseguire una delle seguenti operazioni:

- Selezionare il commento, premere F2 e selezionare il nodo o il nugget.
- Selezionare il nodo o il nugget, premere F2 e selezionare il commento.
- (Solo per mouse a tre pulsanti) Portare il puntatore del mouse sul commento, tenere premuto il pulsante centrale e trascinare il puntatore sul nodo o sul nugget, quindi rilasciare il pulsante.

Per modificare un commento esistente

1. Eseguire una delle seguenti operazioni:
 - Fare doppio clic sulla casella di testo del commento.
 - Selezionare la casella di testo e premere Invio.
 - Fare clic con il pulsante destro sulla casella di testo per visualizzare il menu, quindi scegliere Modifica.
2. Modificare il testo del commento. Durante la modifica, è possibile utilizzare i tasti di scelta rapida standard di Windows (per esempio Ctrl+C per copiare il testo). Le altre opzioni disponibili durante la modifica sono elencate nel menu di scelta rapida del commento.
3. Fare clic una volta all'esterno della casella di testo per visualizzare i comandi di ridimensionamento e poi un'altra volta per completare il commento.

Per ridimensionare la casella di testo di un commento

1. Selezionare il commento per visualizzare i comandi di ridimensionamento.
2. Fare clic e trascinare un comando per ridimensionare la casella.
3. Fare clic all'esterno della casella di testo per salvare la modifica.

Per spostare un commento esistente

Se si desidera spostare un commento, ma non gli eventuali oggetti allegati, eseguire le seguenti operazioni:

- Posizionare il puntatore del mouse sul commento, tenere premuto il pulsante sinistro e trascinare il commento nella nuova posizione.
- Selezionare il commento, tenere premuto il tasto Alt e spostare il commento con i tasti freccia.

Se si desidera spostare un commento insieme agli eventuali nodi o nugget a cui il commento è associato:

1. Selezionare tutti gli oggetti che si desidera spostare.
2. Eseguire una delle seguenti operazioni:
 - Posizionare il puntatore del mouse su uno degli oggetti, tenere premuto il pulsante sinistro e trascinare gli oggetti nella nuova posizione.
 - Selezionare uno degli oggetti, tenere premuto il tasto Alt e spostare gli oggetti utilizzando i tasti freccia.

Per disconnettere un commento da un nodo o da un nugget

1. Selezionare uno o più commenti da disconnettere.
2. Eseguire una delle seguenti operazioni:
 - Premere F3.
 - Fare clic con il pulsante destro del mouse su un commento selezionato e scegliere Disconnetti dal menu.

Per eliminare un commento

1. Selezionare uno o più commenti da eliminare.
2. Eseguire una delle seguenti operazioni:
 - Premere il tasto Canc.
 - Fare clic con il pulsante destro del mouse su un commento selezionato e scegliere Elimina dal menu.

Se il commento era associato a un nodo o a un nugget, viene eliminata anche la linea di connessione.

Se il commento era in origine un'annotazione di stream o di Supernodo che è stata convertita in un commento indipendente, il commento viene eliminato dall'area, a differenza del testo che viene mantenuto nella scheda Annotazioni del flusso o del Supernodo.

Per visualizzare o nascondere i commenti a un flusso

1. Eseguire una delle seguenti operazioni:
 - Nel menu principale, fare clic su:
Visualizza > Commenti
 - Fare clic sul pulsante **Mostra/nascondi commenti** nella barra degli strumenti.

Elenco di commenti su flussi: È possibile visualizzare un elenco di tutti i commenti che sono stati apportati a un flusso o Supernodo specifici.

Nell'elenco è possibile

- Modificare l'ordine dei commenti
- Modificare il testo del commento
- Cambiare il colore di primo piano o di sfondo di un commento

Elenco di commenti

Per elencare tutti i commenti relativi a un flusso, procedere in uno dei seguenti modi:

- Nel menu principale, fare clic su:
Strumenti > Proprietà flusso > Commenti
- Fare clic con il pulsante destro del mouse su un flusso nel riquadro dei manager e scegliere **Proprietà stream**, quindi **Commenti**.
- Fare clic con il pulsante destro del mouse sullo sfondo di un flusso nell'area e scegliere **Proprietà flusso**, quindi **Commenti**.

Testo. Testo del commento. Fare doppio clic sul testo per cambiare il campo in una casella di testo modificabile.

Collegamenti. Nome del nodo a cui è collegato il commento. Se questo campo è vuoto, il commento si applica al flusso.

Posizionamento dei pulsanti. Consentono di spostare un commento selezionato verso l'alto o verso il basso nell'elenco.

Colori commenti. Per cambiare il colore di primo piano o di sfondo di un commento, selezionare il commento, selezionare la casella di controllo **Colori personalizzati**, quindi selezionare un colore dall'elenco **Sfondo** o **Primo piano** (o da entrambi). Fare clic su **Applica**, quindi fare clic sullo sfondo del flusso per visualizzare l'effetto della modifica. Fare clic su **OK** per salvare la modifica.

Conversione di annotazioni in commenti: Le annotazioni apportate a flussi e Supernodi possono essere convertite in commenti.

Nel caso dei flussi, l'annotazione viene convertita in un commento "autonomo", vale a dire non collegato ad alcun nodo, nell'area del flusso.

Quando un'annotazione di un Supernodo viene convertita in commento, questo non è collegato al Supernodo nell'area del flusso, ma è visibile quando si ingrandisce il Supernodo.

Per convertire un'annotazione di flusso in commento

1. Fare clic su **Proprietà flusso** nel menu Strumenti. In alternativa, è possibile fare clic con il pulsante destro del mouse su un flusso nel riquadro dei manager e scegliere **Proprietà flusso**.
2. Fare clic sulla scheda **Annotazioni**.
3. Selezionare la casella di controllo **Mostra annotazione come commento**.
4. Fare clic su **OK**.

Per convertire un'annotazione di Supernodo in commento

1. Fare doppio clic sull'icona del Supernodo nell'area.
2. Fare clic sulla scheda **Annotazioni**.
3. Selezionare la casella di controllo **Mostra annotazione come commento**.
4. Fare clic su **OK**.

Annotazioni

È possibile annotare nodi, stream e modelli in diversi modi. È possibile aggiungere annotazioni descrittive e specificare un nome personalizzato. Queste opzioni sono particolarmente utili durante la creazione di report per i flussi aggiunti al riquadro dei progetti. Per i nodi e i nugget del modello, è inoltre possibile aggiungere un testo di suggerimento per distinguere i nodi simili nell'area del flusso.

Aggiunta di annotazioni

Se si modifica un nodo o un nugget del modello, verrà aperta una finestra di dialogo a schede contenente una scheda Annotazioni, nella quale è possibile impostare numerose opzioni per le annotazioni. È inoltre possibile aprire direttamente la scheda Annotazioni.

1. Per annotare un nodo o un nugget, fare clic con il pulsante destro del mouse sul nodo o sull'insieme nell'area del flusso e fare clic su **Rinomina e annota**. Verrà visualizzata la scheda Annotazioni della finestra di dialogo di modifica.
2. Per annotare un flusso, fare clic su **Proprietà flusso** nel menu Strumenti. In alternativa, è possibile fare clic con il pulsante destro del mouse su un flusso nel riquadro dei manager e scegliere **Proprietà flusso**. Fare clic sulla scheda Annotazioni.

Nome. Selezionare **Personalizzato** per modificare il nome creato automaticamente o per creare un nome univoco per il nodo da visualizzare nell'area del flusso.

Testo di suggerimento. (Nodi e nugget del modello soltanto) Inserire il testo di suggerimento visualizzato nell'area del flusso, che può risultare utile quando si utilizzano molti nodi simili.

Parole chiave. Specificare le parole chiave da utilizzare nei report dei progetti e per la ricerca di nodi in un flusso o per la registrazione di oggetti archiviati nel repository (vedere "Informazioni su IBM SPSS Collaboration and Deployment Services Repository" a pagina 117). Se sono presenti più parole chiave, è possibile separarle con un punto e virgola, per esempio reddito; tipo raccolto; valore richiesta. Gli spazi vuoti all'inizio o alla fine di una parola chiave vengono ritagliati, per esempio reddito ; tipo raccolto produrrà gli stessi risultati di reddito;tipo raccolto. Gli spazi vuoti all'interno delle parole chiave, invece, non vengono troncati. Per esempio, tipo di raccolto con uno spazio e tipo di raccolto con due spazi non si equivalgono.

Nell'area di testo principale è possibile inserire annotazioni più complesse relative alle operazioni o alle scelte eseguite nel nodo. Per esempio, se si condividono e riutilizzano i flussi, è utile inserire note sulle scelte, quali lo scarto di un campo con molti valori vuoti mediante un nodo Filtro. L'annotazione del nodo comporta la memorizzazione di queste informazioni con il nodo. È inoltre possibile includere le annotazioni in un report progetto creato nel riquadro dei progetti. Consultare l'argomento "Introduzione ai progetti" a pagina 141 per ulteriori informazioni.

Mostra annotazione come commento. (Annotazioni di flusso e Supernodi soltanto) Selezionare questa casella per convertire l'annotazione in un commento libero che sarà visibile nell'area del flusso. Consultare l'argomento "Aggiunta di commenti e annotazioni a nodi e flussi" a pagina 55 per ulteriori informazioni.

ID. Visualizza un ID univoco che può essere utilizzato per fare riferimento al nodo a scopo di script o automazione. Questo valore viene generato automaticamente alla creazione del nodo e non subisce variazioni. Si noti inoltre che per evitare confusione con la lettera "O", negli ID dei nodi non vengono utilizzati zero. Utilizzare il pulsante Copia a destra per copiare e incollare l'ID in script o altrove in base alle proprie esigenze.

Salvataggio dei flussi di dati

È possibile salvare i flussi creati in modo da poterli riutilizzare in seguito.

Per salvare un flusso

1. Fare clic su **Salva flusso** o **Salva flusso con nome** nel menu File.
2. Nella finestra di dialogo Salva passare alla cartella in cui si desidera salvare il file di stream.
3. Inserire un nome per il flusso nella casella di testo Nome file.
4. Selezionare **Aggiungi a progetto** se si desidera aggiungere il flusso salvato al progetto corrente.

Facendo clic su **Salva**, il flusso verrà memorizzato nella directory specificata con l'estensione **.str*.

File di backup automatici. Ogni volta che viene salvato un flusso, la versione precedente del file viene automaticamente conservata come backup, con un trattino aggiunto alla fine del nome del file, come in *mystream.str-*. Per ripristinare la versione di backup, basta eliminare il trattino e riaprire il file.

Salvataggio degli stati

Oltre ai flussi, è possibile salvare anche gli **stati**, che includono il diagramma del flusso correntemente visualizzato e tutti i nugget del modello creati (elencati nella scheda Modelli del riquadro dei manager).

Per salvare uno stato

1. Nel menu File, fare clic su:
Stato > Salva stato o Salva stato con nome
2. Nella finestra di dialogo Salva, passare alla cartella in cui si desidera salvare il file di stato.

Facendo clic su **Salva**, lo stato verrà memorizzato nella directory specificata con l'estensione **.cst*.

Salvataggio dei nodi

È inoltre possibile salvare un singolo nodo facendo clic con il pulsante destro del mouse sul nodo nell'area del flusso e fare clic su **Salva nodo** dal menu di scelta rapida. Utilizzare l'estensione di file **.nod*.

Salvataggio di più oggetti del flusso

Quando si esce da IBM SPSS Modeler con più oggetti non salvati, come flusso, progetti o nugget del modello, verrà richiesto di salvare tali oggetti prima della chiusura completa del software. Se si decide di salvare gli elementi, verrà visualizzata una finestra di dialogo, contenente opzioni per il salvataggio di ogni oggetto.

1. È sufficiente selezionare le caselle di controllo relative agli oggetti che si desidera salvare.
2. Per salvare ogni oggetto nel percorso richiesto, fare clic su **OK**.

Verrà quindi visualizzata una finestra di dialogo Salva standard per ogni oggetto. Al termine del salvataggio, l'applicazione verrà chiusa, come richiesto in origine.

Salvataggio dell'output

Le tabelle, i grafici e i report generati dai nodi Output di IBM SPSS Modeler possono essere salvati in formato oggetto di output (**.cou*).

1. Quando si visualizza l'output che si desidera salvare, nei menu della finestra di output fare clic su:
File > Salva
2. Specificare un nome e una posizione per il file di output.
3. Se si desidera, selezionare **Aggiungi file a progetto** nella finestra di dialogo Salva, per includere il file nel progetto corrente. Consultare l'argomento "Introduzione ai progetti" a pagina 141 per ulteriori informazioni.

In alternativa, è possibile fare clic con il pulsante destro del mouse su uno qualsiasi degli oggetti di output elencati nel riquadro dei manager e selezionare **Salva** dal menu di scelta rapida.

Crittografia e decrittografia di informazioni

Quando si salva un flusso, un nodo, un progetto, un file di output o un nugget del modello, è possibile crittografarlo per impedirne l'uso non autorizzato. A questo scopo, è necessario selezionare un'opzione aggiuntiva al momento del salvataggio e aggiungere una password all'elemento che si sta salvando. È possibile crittografare qualsiasi elemento salvato fornendogli una protezione aggiuntiva. Questa funzione non corrisponde alla crittografia SSL utilizzata quando si passano file tra IBM SPSS Modeler e IBM SPSS Modeler Server.

Quando si tenta di aprire un elemento crittografato, il sistema chiede di immettere la password. Dopo aver immesso la password corretta, l'elemento viene decrittografato automaticamente e si apre come di consueto.

Per crittografare un elemento

1. Nella finestra di dialogo Salva, fare clic su **Opzioni** per l'elemento da crittografare. Viene visualizzata la finestra di dialogo Opzioni di crittografia.
2. Selezionare **Crittografa questo file**.
3. Se lo si desidera, per aumentare la protezione selezionare **Nascondi password**. Tutti i caratteri immessi verranno visualizzati come una serie di punti.
4. Immettere la password. *Avviso:* se si dimentica la password, non è possibile aprire il file o il modello.
5. Se si seleziona **Nascondi password**, immettere nuovamente la password per confermarla.
6. Scegliere **OK** per ritornare alla finestra di dialogo Salva.

Nota: se si salva una copia di un elemento protetto mediante crittografia, il nuovo elemento viene automaticamente salvato in formato crittografato utilizzando la password originale, a meno che non si modifichino le impostazioni della finestra di dialogo Opzioni di crittografia.

Caricamento di file

È possibile ricaricare numerosi oggetti salvati in IBM SPSS Modeler:

- Flussi (.str)
- Stati (.cst)
- Modelli (.gm)
- Palette Modelli (.gen)
- Nodi (.nod)
- Output (.cou)
- Progetti (.cpj)

Apertura di nuovi file

È possibile caricare i flussi direttamente dal menu File.

1. Nel menu File, fare clic su **Apri flusso**.

Tutti gli altri tipi di file possono essere aperti mediante i comandi dei sottomenu disponibili nel menu File. Per esempio, per caricare un modello, nel menu File fare clic su:

Modelli > Apri modello o Carica tavolozza modelli

Apertura di file utilizzati di recente

Per caricare rapidamente i file utilizzati di recente, utilizzare le opzioni disponibili nella parte inferiore del menu File.

Selezionare **Flussi recenti**, **Progetti recenti** o **Stati recenti** per espandere un elenco di file utilizzati di recente.

Mappatura dei flussi di dati

Lo strumento per la mappatura dei dati consente di connettere una nuova sorgente dati a un flusso preesistente. Lo strumento per la mappatura non solo imposterà la connessione, ma consentirà anche di

specificare in quale modo i campi della nuova sorgente sostituiranno quelli del flusso esistente. Aniché ricreare un intero stream di dati per una nuova sorgente dati, sarà sufficiente connettersi a un flusso esistente.

Lo strumento per la mappatura consente di unire due frammenti di stream avendo la certezza che tutti i nomi dei campi essenziali corrispondano. La mappatura dei dati consiste semplicemente nella creazione di un nuovo nodo Filtro, che crea una corrispondenza dei campi appropriati rinominandoli.

Esistono due metodi equivalenti per mappare i dati:

Seleziona nodo sostitutivo. Questo metodo parte dal nodo che deve essere sostituito. Innanzitutto, fare clic con il pulsante destro del mouse sul nodo da sostituire; quindi, utilizzando l'opzione **Mappatura dati > Seleziona nodo sostitutivo** del menu di scelta rapida, selezionare il nodo con il quale sostituirlo.

Mappa a. Questo metodo parte dal nodo che deve essere inserito nel flusso. Innanzitutto, fare clic con il pulsante destro del mouse sul nodo da inserire; quindi, utilizzando l'opzione **Mappatura dati > Mappa a** del menu di scelta rapida, selezionare il nodo in cui inserirlo. Questo metodo è particolarmente utile per la mappatura su un nodo finale. *Nota:* non è possibile eseguire la mappatura su nodi Unione o Accodamento. In questi casi è sufficiente connettere il flusso al nodo Unione nel modo normale.

La mappatura dei dati è strettamente integrata con la creazione dei flussi. Se si tenta di connettersi a un nodo che ha già una connessione, sarà possibile sostituire la connessione o la mappatura al nodo specifico.

Mappatura dei dati su un modello

Per sostituire l'origine dati di un flusso di modelli con un nuovo nodo origine che consente di inserire i dati in IBM SPSS Modeler, è consigliabile utilizzare l'opzione **Seleziona nodo sostitutivo** dal menu di scelta rapida Mappatura dati. L'opzione è disponibile per tutti i nodi a eccezione dei nodi Unione, Aggregazione e tutti i nodi finali. Se si utilizza lo strumento per la mappatura dei dati per eseguire questa operazione, si avrà la certezza che esista una corrispondenza esatta tra i campi delle operazioni del flusso esistenti e della nuova sorgente dati. I passaggi seguenti forniscono una panoramica del processo di mappatura dei dati.

Fase 1: specificare i campi essenziali nel nodo origine. Affinché le operazioni del flusso vengano eseguite correttamente, è necessario specificare i campi essenziali. Consultare l'argomento "Definizione dei campi essenziali" a pagina 64 per ulteriori informazioni.

Fase 2: aggiungere la nuova origine dati all'area del flusso. Utilizzando uno dei nodi origine, inserire i nuovi dati sostitutivi.

fase 3: sostituire il nodo origine del modello. Utilizzando l'opzione Mappatura dati del menu di scelta rapida per il nodo origine del modello, fare clic su **Seleziona nodo sostitutivo**, quindi selezionare il nodo origine per i dati sostitutivi.

Fase 4: controllare i campi mappati. Nella finestra di dialogo visualizzata, controllare che venga eseguita correttamente la mappatura dei campi dalla sorgente dati sostitutiva al flusso. I campi essenziali che non sono stati mappati vengono visualizzati in rosso. Questi campi vengono utilizzati nelle operazioni del flusso e devono essere sostituiti con un campo simile nella nuova sorgente dati affinché le operazioni downstream possano essere eseguite correttamente. Consultare l'argomento "Analisi dei campi mappati" a pagina 64 per ulteriori informazioni.

Dopo che si è controllato che tutti i campi essenziali siano stati mappati correttamente, la sorgente dati precedente verrà disconnessa e la nuova sorgente dati verrà connessa al flusso mediante un nodo Filtro denominato *Mappa*. Questo nodo Filtro gestisce la mappatura effettiva dei campi nel flusso. Nell'area del flusso viene inoltre inserito un nodo Filtro denominato *Annulla mappatura*. Il nodo filtro *Unmap* consente

di invertire la mappatura del nome di campo aggiungendolo al flusso. I campi mappati verranno annullati, ma sarà necessario modificare i nodi downstream per riselectare i campi e le sovrapposizioni.

Mappatura tra flussi

Questo metodo di mappatura dei dati è simile alla connessione dei nodi, ma non richiede l'impostazione iniziale dei campi essenziali. Con questo metodo, si esegue semplicemente la connessione da un flusso a un altro utilizzando l'opzione **Mappa a** del menu di scelta rapida Mappatura dati. Questo tipo di mappatura dei dati è utile per la mappatura su nodi finali e per le operazioni di copia e incolla tra stream. *Nota:* se si utilizza l'opzione **Mappa a**, non è possibile eseguire la mappatura ai nodi Unione, Accodamento e a tutti i nodi origine.

Per mappare i dati tra flussi

1. Fare clic con il pulsante destro del mouse sul nodo da utilizzare per la connessione al nuovo stream.
2. Nel menu, fare clic su:
Mappatura di dati > Mappa a
3. Utilizzare il cursore per selezionare un nodo di destinazione nel flusso obiettivo.
4. Nella finestra di dialogo visualizzata, controllare che i campi corrispondano e fare clic su **OK**.

Definizione dei campi essenziali

Quando si esegue la mappatura su un flusso esistente, i campi essenziali vengono in genere specificati dall'autore del flusso. Questi campi essenziali indicano se un campo specifico viene utilizzato nelle operazioni downstream. Per esempio, è possibile che il flusso esistente crei un modello che utilizza un campo denominato *Tasso di abbandono*. In questo stream, *Tasso di abbandono* è un campo essenziale perché non è possibile creare il modello senza di esso. Allo stesso modo, i campi utilizzati nei nodi di manipolazione quali un nodo Ricava sono necessari per derivare il nuovo campo. L'impostazione esplicita di questi campi come campi essenziali garantisce che ad essi vengano mappati i campi corretti del nuovo nodo origine. Se non si mappano i campi obbligatori, verrà visualizzato un messaggio di errore. Se si stabilisce che alcune manipolazioni o alcuni nodi di output non sono necessari, è possibile eliminare i nodi dal flusso e rimuovere i campi appropriati dall'elenco dei campi essenziali.

Per impostare i campi essenziali

1. Fare clic con il pulsante destro del mouse sul nodo origine del flusso di modelli che verrà sostituito.
2. Nel menu, fare clic su:
Mappatura di dati > Specifica file essenziali
3. È possibile aggiungere o rimuovere campi dall'elenco mediante il selettore di campo. Per aprire il selettore di campo, fare clic sull'icona a destra dell'elenco di campi.

Analisi dei campi mappati

Dopo avere selezionato il punto nel quale un flusso di dati o una sorgente dati verranno mappati a un altro, verrà aperta una finestra di dialogo nella quale è possibile selezionare i campi da mappare o verificare che la mappatura di default di sistema sia corretta. Se i campi essenziali per il flusso o la sorgente dati sono stati impostati e non corrispondono, verranno visualizzati in rosso. I campi non mappati dalla sorgente dati passeranno attraverso il nodo Filtro senza essere modificati. Si noti tuttavia che è possibile mappare anche i campi non essenziali.

Originali. Elenca tutti i campi del modello o del flusso esistente, ovvero tutti i campi downstream disponibili. I campi della nuova sorgente dati verranno mappati a questi campi.

Mappati. Elenca i campi selezionati per la mappatura ai campi del modello. Si tratta dei campi il cui nome potrebbe essere modificato in modo da corrispondere a quello dei campi originali utilizzati nelle operazioni del flusso. Se si fa clic sulla cella di un campo, verrà attivato un elenco dei campi disponibili.

Se non si è certi dei campi da mappare, può essere utile analizzare in modo approfondito la sorgente dati prima di procedere alla mappatura. Per esempio, è possibile utilizzare la scheda Tipo nel nodo origine per esaminare un riepilogo dei dati di input.

Suggerimenti e tasti di scelta rapida

Per rendere più rapido e semplice l'utilizzo di Clementine, si consiglia di familiarizzare con i tasti di scelta rapida e i suggerimenti indicati di seguito.

- **Creare flusso rapidamente facendo doppio clic.** Fare semplicemente doppio clic su un nodo della palette per aggungerlo e connetterlo al flusso corrente.
- **Utilizzare le combinazioni dei tasti di scelta rapida per selezionare i nodi downstream.** Premere Ctrl+Q e Ctrl+W per attivare/disattivare la selezione di tutti i nodi downstream.
- **Utilizzare i tasti di scelta rapida per connettere e disconnettere i nodi.** Quando si seleziona un nodo nell'area, premere F2 per iniziare una connessione, premere Tab per passare al nodo richiesto e quindi Maiusc+Barra spaziatrice per completare la connessione. Utilizzare F3 per disconnettere tutti gli input e gli output al nodo selezionato.
- **Personalizza la scheda Tavolozza nodi con i nodi preferiti.** Nel menu Strumenti, fare clic su **Gestisci tavolozza** per aprire una finestra di dialogo e aggiungere, rimuovere o spostare i nodi presenti nella Tavolozza nodi.
- **Rinominare i nodi e aggiungere le descrizioni.** In ogni finestra di dialogo di nodo è disponibile una scheda Annotazioni in cui è possibile specificare un nome personalizzato per i nodi nell'area nonché descrizioni che consentano di organizzare il flusso. È inoltre possibile includere annotazioni più complesse per tenere traccia dell'avanzamento, salvare i dettagli del processo e indicare le decisioni di business necessarie o già raggiunte.
- **Inserire automaticamente i valori in un'espressione CLEM .** Tramite il generatore di espressioni, a cui è possibile accedere da diverse finestre di dialogo quali quelle dei nodi Ricava e Riempimento, è possibile inserire automaticamente i valori dei campi in un'espressione CLEM. Fare clic sul pulsante dei valori nel generatore di espressioni per scegliere tra i valori di campo esistenti.



Figura 13. Pulsante dei valori

- **Cercare file rapidamente.** Quando si cercano i file in una finestra di dialogo Apri, utilizzare l'elenco File (fare clic sul pulsante rombo giallo nella parte superiore della finestra di dialogo, accanto al campo Cerca in) per accedere alle directory utilizzate in precedenza nonché alle directory di default di IBM SPSS Modeler. Utilizzare i pulsanti avanti e indietro per scorrere le directory a cui si è effettuato l'accesso.
- **Ridurre lo spazio della finestra di output.** È possibile chiudere ed eliminare rapidamente l'output utilizzando il pulsante con la X rossa nell'angolo in alto a destra di tutte le finestre di output. In questo modo nella scheda Output del riquadro dei manager resteranno visualizzati solo i risultati interessanti.

Per il software è disponibile una gamma completa di tasti di scelta rapida. Consultare l'argomento "Accesso facilitato mediante tastiera" a pagina 166 per ulteriori informazioni.

Operazioni che è possibile effettuare

- Trascinare e selezionare un gruppo di nodi nell'area del flusso utilizzando il mouse.
- Copiare e incollare nodi da un flusso a un altro.
- Accedere alla Guida in linea da ogni finestra di dialogo e finestra di output.
- Visualizzare gli argomenti della Guida in linea relativi a CRISP-DM (Cross-Industry Standard Process for Data Mining). Fare clic su **Guida in linea di CRISP-DM** nel menu ?.

Capitolo 6. Gestione dei valori mancanti

Panoramica sui valori mancanti

Durante la fase Data Preparation del data mining è in genere opportuno sostituire i valori mancanti. I **valori mancanti** sono valori dell'insieme di dati sconosciuti, non raccolti o inseriti non correttamente e che in genere non sono validi per i relativi campi. Ad esempio, il campo *Sesso* dovrebbe contenere valori *M* e *F*. Se invece nel campo sono presenti i valori *Y* o *Z*, si può presumere che tali valori non siano validi e pertanto verranno interpretati come valori vuoti. Anche un valore negativo per il campo *Età* è privo di significato e va interpretato come un valore vuoto. Accade di frequente che in un questionario vengano inseriti valori chiaramente errati di questo tipo oppure che i campi vengano lasciati vuoti, a indicare che l'utente non ha voluto rispondere. A volte può essere utile analizzare i valori vuoti in modo più approfondito, per determinare se il rifiuto di fornire una risposta (per esempio relativa alla propria età) possa essere utile per prevedere un risultato specifico.

Alcune tecniche di modellazione consentono di gestire meglio i dati mancanti rispetto ad altre. Per esempio, C5.0 e Apriori sono particolarmente utili per i valori dichiarati esplicitamente come mancanti in un nodo Tipo. Altre tecniche non funzionano altrettanto bene con i valori mancanti, richiedono tempi di addestramento maggiori e i modelli risultanti sono meno precisi.

Esistono vari tipi di valori mancanti riconosciuti da IBM SPSS Modeler:

- **Valori null o mancanti di sistema.** Sono valori non stringa che vengono lasciati vuoti nel database o nel file di origine e che non sono specificati in modo esplicito come "mancanti" in un nodo Tipo o in un nodo origine. I valori mancanti di sistema vengono visualizzati come **\$null\$**. In IBM SPSS Modeler le stringhe vuote non vengono considerate valori null, ma possono essere considerate valori null da alcuni database.
- **Stringhe vuote e spazi vuoti.** I valori stringa vuote e spazi vuoti (stringhe prive di caratteri visibili) sono trattati in modo diverso rispetto ai valori null. Nella maggior parte dei casi, le stringhe vuote vengono considerate equivalenti agli spazi vuoti. Per esempio, se si seleziona l'opzione che consente di trattare gli spazi vuoti come valori vuoti in un nodo origine Tipo, questa impostazione risulta valida anche per le stringhe vuote.
- **Valori vuoti o mancanti definiti dall'utente.** Sono valori come, ad esempio, sconosciuto, 99 o -1 definiti in modo esplicito come mancanti in un nodo origine o un nodo Tipo. Se lo si desidera, si può anche decidere di considerare i valori null e gli spazi vuoti come valori vuoti: in questo modo è possibile evidenziarli in modo da poter essere elaborati diversamente ed esclusi dalla maggior parte dei calcoli. Per esempio, è possibile utilizzare la funzione @BLANK per trattare questi valori, insieme ad altri tipi di valori mancanti, come valori vuoti.

Letture di dati misti. Quando si leggono campi con archiviazione numerica (numeri interi, reali, ora, timestamp o data), gli eventuali valori non numerici vengono impostati come *valori null* o *valori mancanti di sistema* poiché, a differenza di altre applicazioni, Clementine non consente di utilizzare tipi di archiviazione misti all'interno dello stesso campo. Per evitare questo inconveniente, è necessario modificare opportunamente il tipo di archiviazione nel nodo origine o nell'applicazione esterna per fare in modo che tutti i campi contenenti dati misti siano letti come stringhe.

Letture di stringhe vuote da Oracle. Durante la lettura o la scrittura in un database Oracle, tenere presente che, a differenza di IBM SPSS Modeler e a differenza della maggior parte degli altri database, Oracle considera ed archivia i valori di stringa vuoti come valori null. Questo significa che gli stessi dati estratti da un database Oracle potrebbero comportarsi in modo diverso rispetto a quando vengono estratti da un file o da un altro database e i dati potrebbero restituire risultati diversi.

Gestione dei valori mancanti

Il tipo di trattamento scelto per i valori mancanti dipende da conoscenze specifiche di business o di dominio. La rimozione dei valori vuoti dall'insieme di dati in uso consente di ridurre i tempi di addestramento e di ottenere una maggiore precisione. La presenza dei valori vuoti può d'altra parte determinare nuove opportunità di business o fornire informazioni aggiuntive. Per stabilire la tecnica migliore da utilizzare, è necessario considerare i seguenti aspetti dei dati con cui si lavora:

- Dimensione dell'insieme di dati
- Numero dei campi contenenti valori vuoti
- Quantità di informazioni mancanti

In linea generale, è possibile adottare due approcci:

- Escludere i campi o i record con valori mancanti
- Assegnare, sostituire o forzare i valori mancanti mediante una serie di metodi diversi

Entrambi gli approcci possono essere in gran parte automatizzati mediante il nodo Esplora. Per esempio, è possibile generare un nodo Filtro che escluda i campi che presentano troppi valori mancanti per essere utili ai fini della modellazione e generare un Supernodo che assegni i valori mancanti ad alcuni o tutti i campi restanti. Proprio in questi casi viene dimostrata la vera efficacia dell'esplorazione, che consente non solo di valutare lo stato corrente dei dati, ma anche di agire sulla base di tale valutazione.

Gestione dei record con valori mancanti

Se la maggior parte dei valori mancanti è concentrata in un numero ridotto di record, è possibile semplicemente escludere quei record. Per esempio, le banche in genere dispongono di informazioni dettagliate sui clienti che hanno richiesto prestiti, ma se per l'approvazione dei prestiti al personale della banca vengono applicate condizioni meno restrittive, è probabile che nei relativi dati siano presenti molti campi vuoti. In questo caso, è possibile gestire i valori mancanti in due modi:

- È possibile utilizzare un nodo Seleziona per rimuovere i record relativi al personale.
- Se l'insieme di dati è grande, è possibile scartare tutti i record con valori vuoti.

Gestione dei campi con valori mancanti

Se la maggior parte dei valori mancanti si concentra in un numero ristretto di campi, è possibile gestire questi valori a livello di campo anziché a livello di record. Ciò consente inoltre di verificare l'importanza relativa di campi specifici prima di scegliere un approccio per la gestione dei valori mancanti. Se un campo è irrilevante per la creazione di modelli, probabilmente non vale la pena di conservarlo, indipendentemente dal suo numero di valori mancanti.

Per esempio, è possibile che una società di ricerche di mercato raccolga i dati mediante un questionario generale che include 50 domande. Due domande sono relative all'età e alle opinioni politiche, ovvero informazioni che la maggior parte delle persone preferisce non fornire. In questo caso, i campi *Età* e *Tendenze_politiche* contengono numerosi valori mancanti.

Livello di misurazione dei campi

Per stabilire il metodo da utilizzare, è opportuno considerare anche il livello di misurazione dei campi contenenti i valori mancanti.

Campi numerici. Per i tipi di campi numerici, per esempio *Continuo*, è sempre consigliabile eliminare tutti i valori non numerici prima di creare un modello, perché molti modelli non funzioneranno se i campi numerici contengono valori vuoti.

Campi relativi alla categoria. Per i campi relativi alla categoria, per esempio i campi *nominali* e *indicatore*, la modifica dei valori mancanti non è necessaria, ma consentirà comunque di aumentare la precisione del

modello. Per esempio, un modello che utilizza il campo *Sesso* funzionerà anche se sono presenti valori senza significato, come *Y* e *Z*, ma la rimozione di tutti i valori diversi da *M* e *F* aumenterà la precisione del modello.

Screening o rimozione di campi

I campi con troppi valori mancanti possono essere sottoposti a screening in diversi modi:

- I nodi *Esplora* possono essere utilizzati per filtrare i campi in base alla qualità.
- Si può utilizzare un nodo *Selezione funzioni* per escludere i campi i cui valori mancanti superano una percentuale specifica e per ordinare i campi in base all'importanza rispetto a un obiettivo specificato.
- Anziché rimuovere i campi, è possibile utilizzare un nodo *Tipo* per impostare il ruolo dei campi su **Nessuno**. In questo modo, i campi verranno mantenuti nell'insieme di dati, ma saranno esclusi dai processi di modellazione.

Assegnazione o riempimento dei valori mancanti

Se il numero dei valori mancanti è ridotto, può essere utile inserire valori sostitutivi al posto dei valori vuoti. Questa operazione può essere eseguita dal report del nodo *Esplora*, che consente di specificare secondo necessità le opzioni per i campi specifici e di generare un Supernodo che assegna i valori mediante diversi metodi. Questo è il metodo più flessibile, che consente inoltre di specificare la gestione di grandi quantità di campi in un unico nodo.

Per l'immissione di valori mancanti sono disponibili i seguenti metodi:

Fisso. Sostituisce con un valore fisso (la media di un campo, il valore intermedio dell'intervallo o una costante specificata dall'utente).

Casuale. Sostituisce con un valore casuale in base a una distribuzione normale o uniforme.

Espressione. Consente di specificare un'espressione personalizzata. Per esempio, è possibile sostituire valori con una variabile globale creata dal nodo *Calcola globali*.

Algoritmo. Sostituisce un valore previsto da un modello basato sull'algoritmo C&RT. Per ciascun campo immesso utilizzando questo metodo, sarà presente un modello C&RT separato, insieme ad un nodo *Riempimento* che sostituisce i valori vuoti e null con il valore previsto dal modello. Viene quindi utilizzato un nodo *Filtro* per rimuovere i campi di previsione generati dal modello.

In alternativa, per forzare i valori per campi specifici, è possibile utilizzare un nodo *Tipo* per fare in modo che i tipi di campi includano solo valori validi e quindi impostare su **Forza** la colonna *Controllo* dei campi per i quali è necessario sostituire i valori mancanti.

Funzioni CLEM per i valori mancanti

Sono disponibili numerose funzioni per la gestione dei valori mancanti. Per scartare o sostituire i valori mancanti, nei nodi *Selezione* e *Riempimento* vengono spesso utilizzate le seguenti funzioni:

- `count_nulls(LIST)`
- `@BLANK(FIELD)`
- `@NULL(FIELD)`
- `undef`

La combinazione delle funzioni @ e della funzione @FIELD consente di individuare i valori vuoti o null in uno o più campi. È possibile limitarsi a contrassegnare i campi che contengono valori vuoti o null, riempirli con valori sostitutivi o utilizzarli per molte altre operazioni.

All'interno di un elenco di campi, i valori null si possono contare nel modo seguente:

```
count_nulls(['possessocarta' 'possessocarta2' 'possessocarta3'])
```

Quando si utilizzano funzioni che accettano come input un elenco di campi, è possibile utilizzare le funzioni speciali @FIELDS_BETWEEN e @FIELDS_MATCHING come nell'esempio seguente:

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

La funzione undef consente di riempire i campi con il valore mancante di sistema, visualizzato come **\$null\$**. Per esempio, mediante un'istruzione condizionale simile alla seguente è possibile sostituire qualsiasi valore numerico:

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

Tutti i valori che non rientrano nell'intervallo verranno sostituiti da un valore mancante di sistema, visualizzato con **\$null\$**. Con la funzione not(), è possibile rilevare tutti gli altri valori numerici, inclusi quelli negativi. Consultare l'argomento "Funzioni per la gestione di valori vuoti e null" a pagina 114 per ulteriori informazioni.

Nota sullo scarto di record

Se si scartano i record utilizzando un nodo Seleziona, tenere presente che la sintassi utilizza la logica tripolare e include automaticamente i valori null nelle istruzioni di selezione. Per escludere i valori null (valori mancanti di sistema) in un'espressione di selezione è necessario specificarlo in modo esplicito utilizzando and not nell'espressione. Per esempio, per selezionare e includere tutti i record nei quali il farmaco prescritto è CuraC, verrà utilizzata l'istruzione di selezione seguente:

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

Nelle versioni precedenti in questo tipo di situazioni i valori null venivano esclusi.

Capitolo 7. Creazione di espressioni CLEM

Informazioni su CLEM

Control Language for Expression Manipulation (CLEM) è un linguaggio efficace per l'analisi e la manipolazione dei flussi di dati nei flussi di IBM SPSS Modeler. I data miner fanno ampio ricorso al linguaggio CLEM nelle operazioni stream per eseguire attività sia semplici, come la derivazione dei profitti dai dati relativi ai costi e alle entrate, sia complesse, come la trasformazione dei dati dei registri Web in un insieme di campi e record contenenti informazioni utilizzabili.

CLEM è utilizzato all'interno di IBM SPSS Modeler per:

- Confrontare e valutare le condizioni nei campi dei record.
- Derivare valori per nuovi campi.
- Derivare nuovi valori per campi esistenti.
- Analizzare le sequenze di record.
- Inserire dati di record in report.

Le espressioni CLEM sono indispensabili per la preparazione dei dati in IBM SPSS Modeler e possono essere utilizzate in una vasta gamma di nodi, da operazioni su record e campi (selezione, bilanciamento, riempimento) a grafici e output (Analisi, Report, Tabella). Per esempio, è possibile utilizzare il linguaggio CLEM in un nodo Ricava per creare un nuovo campo in base a una formula, quale un rapporto.

È inoltre possibile utilizzare le espressioni CLEM per operazioni di ricerca e sostituzione globali. Per esempio, è possibile utilizzare l'espressione @NULL(@FIELD) in un nodo Riempimento per sostituire **valori mancanti di sistema** con il valore intero 0. Per sostituire **valori mancanti definiti dall'utente**, denominati anche vuoti, si utilizza invece la funzione @BLANK.

È inoltre possibile creare espressioni CLEM più complesse. Ad esempio, è possibile derivare nuovi campi in base a un insieme condizionale di regole, come una categoria nuovo valore creato utilizzando la seguente espressione: `If: CardID = @OFFSET(CardID,1), Then: @OFFSET(ValueCategory,1), Else: 'exclude'`.

In questo esempio viene utilizzata la funzione @OFFSET per specificare che "se il valore del campo *id_carta* per un determinato record corrisponde al valore del record precedente, restituire il valore del campo denominato *CategoriaValore* per il record precedente. In caso contrario, assegnare la stringa "exclude". In altre parole, se i campi *id_carta* per i record adiacenti sono identici, dovranno essere assegnati alla stessa categoria di valore. (È possibile includere i record con la stringa exclude in un momento successivo tramite un nodo Selezione).

Esempi CLEM

Di seguito sono riportate alcune espressioni di esempio che illustrano la sintassi corretta nonché i tipi di espressione validi con CLEM.

Espressioni semplici

Le formule possono essere molto semplici, come quella riportata di seguito, che consente di creare un nuovo campo in base ai valori dei campi *Dopo* e *Prima*:

`(Dopo - Prima) / Prima * 100.0`

Si osservi che i nomi dei campi sono privi di virgolette quando viene fatto riferimento ai valori del campo.

Analogamente, la seguente espressione restituisce il logaritmo di ciascun valore per il campo *salario*.
`log(salario)`

Espressioni complesse

Le espressioni possono anche essere lunghe e complesse. La seguente espressione restituisce *vero* se il valore di due campi (*\$KX-Kohonen* e *\$KY-Kohonen*) rientra nell'intervallo specificato. Si noti che qui i nomi dei campi sono racchiusi tra virgolette singole, perché contengono caratteri speciali.

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

Per diverse funzioni, ad esempio le funzioni stringa, è necessario immettere vari parametri utilizzando la sintassi corretta. La funzione `subscrs` dell'esempio seguente restituisce il primo carattere di un campo *ID_prodotto*, per indicare se un elemento è biologico, geneticamente modificato o convenzionale. I risultati di un'espressione vengono descritti da `-> `result``.

```
subscrs(1, ID_prodotto) -> `c`
```

Analogamente, la seguente espressione è:

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

È importante osservare che i caratteri sono sempre racchiusi tra virgolette singole.

Combinazione di funzioni in un'espressione

Spesso le espressioni CLEM sono costituite da una combinazione di funzioni. Nella funzione seguente, la combinazione di `subscr` e `lowertoupper` restituisce il primo carattere di *ID_prodotto* e lo converte in una lettera maiuscola.

```
lowertoupper(subscr(1, ID_prodotto)) -> `C`
```

Questa stessa espressione può essere scritta in forma ridotta, vale a dire:

```
lowertoupper(ID_prodotto(1)) -> `C`
```

Di seguito è riportata un'altra combinazione comune di funzioni:

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

Questa espressione consente di individuare il carattere ``n`` nei valori del campo *pagina_web* partendo dall'ultimo carattere del valore del campo. Se si include anche la funzione `length`, l'espressione può calcolare dinamicamente la lunghezza del valore corrente, anziché utilizzare un numero statico come 7, che risulterebbe non valido per i valori con meno di sette caratteri.

Funzioni speciali

Sono disponibili numerose funzioni speciali, precedute dal simbolo `@`. Le funzioni di utilizzo più comune sono:

```
@BLANK('IDreferente') -> T
```

Spesso le funzioni speciali vengono utilizzate insieme. Si tratta di un metodo di utilizzo comune per evidenziare i valori vuoti in più campi alla volta.

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

Ulteriori esempi sono disponibili nella documentazione di CLEM. Consultare l'argomento "Panoramica di riferimento su CLEM" a pagina 85 per ulteriori informazioni.

Valori e tipi di dati

Le espressioni CLEM sono simili a formule costruite in base a valori, nomi di campi, operatori e funzioni. La più semplice delle espressioni CLEM valide è un valore o un nome di campo. Esempi di valori validi sono:

```
3
1.79
'banana'
```

Esempi di nomi di campi sono:

```
ID_Prodotto
'$P-CampoSuccessivo'
```

dove *Prodotto* è il nome di un campo derivante da un insieme di dati di market basket, '\$P-NextField' è il nome di un parametro e il valore dell'espressione è il valore del campo denominato. I nomi di campi iniziano in genere con una lettera, ma possono contenere anche cifre e caratteri di sottolineatura (_). Se si racchiude il nome tra virgolette, è possibile utilizzare nomi che non seguono queste regole. I valori CLEM possono essere dei seguenti tipi:

- Stringhe: ad esempio, "c1", "Tipo 2", "del testo libero"
- Interi: ad esempio, 12, 0, -189
- Numeri reali: ad esempio, 12,34, 0,0, -0,0045
- Campi di data/ora, per esempio 05/12/2002, 12/05/2002, 12/05/02

È possibile utilizzare anche i seguenti elementi:

- Codici carattere: per esempio `a` o 3
- Elenchi di elementi: per esempio [1 2 3], ['Tipo 1' 'Tipo 2']

In genere i codici carattere e gli elenchi non vengono utilizzati come valori di campo, ma come argomenti di funzioni CLEM.

Regole per l'inserimento tra virgolette

Sebbene il software sia flessibile nella determinazione dei campi, dei valori, dei parametri e delle stringhe utilizzati in un'espressione CLEM, di seguito sono suggerite alcune "procedure consigliate" da seguire durante la creazione di espressioni.

- **Stringhe:** utilizzare sempre le virgolette quando si scrivono stringhe ("Type 2" o "value"). È inoltre possibile utilizzare le virgolette singole, ma il rischio è di creare confusione con i campi racchiusi tra virgolette.
- **Caratteri:** utilizzare sempre virgolette singole, ovvero ` . Si noti, per esempio, il carattere d nella funzione stripchar(`d`,`curaA`). L'unica eccezione è rappresentata dall'utilizzo di un intero per fare riferimento a un carattere specifico in una stringa. Ad esempio il carattere 5 nella funzione lowertoupper("druga"(5)) -> "A". *Nota:* il codice Unicode della virgoletta singola è 0060.
- **Campi:** vengono in genere utilizzati senza virgolette nelle espressioni CLEM (subscr(2,arrayID)) -> CHAR). Utilizzare le virgolette singole solo se necessarie per racchiudere spazi o altri caratteri speciali ('Numero ordine'). I campi racchiusi tra virgolette ma non definiti nell'insieme di dati verranno interpretati come stringhe.
- **Parametri:** utilizzare sempre le virgolette singole ('\$P-threshold').

Espressioni e condizioni

Le espressioni CLEM possono restituire un risultato, utilizzato quando si derivano nuovi valori, per esempio:

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

oppure possono restituire *vero* o *falso*, utilizzato quando la selezione viene effettuata in base a una condizione, per esempio:

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

Nelle espressioni CLEM è possibile combinare in modo arbitrario operatori e funzioni, per esempio:
`sqrt(abs(Signal)) * max(T1, T2) + Baseline`

Le parentesi e la precedenza tra gli operatori determinano in quale ordine le espressioni verranno valutate. In questo esempio l'ordine di valutazione è il seguente:

- Viene valutato `abs(Signal)` e al risultato viene applicato `sqrt`.
- Viene valutato `max(T1, T2)`.
- I due risultati vengono moltiplicati: `x` ha la precedenza su `+`.
- Infine al risultato viene aggiunto `Baseline`.

Di seguito è specificato l'ordine decrescente di precedenza, ovvero dalle operazioni eseguite per prime a quelle eseguite per ultime:

- Argomenti della funzione
- Chiamate della funzione
- **xx**
- **x / mod div rem**
- **+ -**
- **> < >= <= /== == = /=**

Se si desidera ignorare la precedenza o se si hanno dei dubbi sull'ordine di valutazione, è possibile utilizzare le parentesi in modo esplicito, per esempio

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

Parametri flusso, sessione e Supernodo

È possibile definire parametri da utilizzare in espressioni CLEM e negli script. Si tratta a tutti gli effetti di variabili definite dall'utente che vengono salvate e mantenute con il flusso, la sessione o il Supernodo corrente e a cui è possibile accedere dall'interfaccia utente nonché attraverso script. Per esempio, durante il salvataggio di un flusso, qualsiasi parametro impostato per tale stream viene salvato contemporaneamente. In questo modo i parametri vengono distinti dalle variabili di script locali, utilizzabili solo nello script in cui sono dichiarate. I parametri vengono spesso utilizzati nello script per controllare il comportamento dello script, fornendo informazioni sui campi e i valori che non necessitano di essere attentamente codificati nello script.

L'ambito di un parametro dipende dalla posizione in cui è impostato:

- i parametri impostati per un flusso possono essere impostati in uno script del flusso o nella finestra di dialogo delle proprietà del flusso e sono disponibili a tutti i nodi del flusso. Vengono visualizzati nell'elenco Parametri nel generatore di espressioni.

- I parametri di sessione possono essere impostati in uno script autonomo oppure nella finestra di dialogo dei parametri di sessione e sono disponibili per tutti i flussi utilizzati nella sessione corrente, ovvero per tutti i flussi elencati nella scheda Flusso del riquadro dei manager.

È inoltre possibile impostare parametri per Supernodi. In questo caso, sono visibili solo ai nodi incapsulati all'interno del Supernodo.

Utilizzo dei parametri nelle espressioni CLEM

I parametri vengono rappresentati nelle espressioni CLEM con `$P-nomep`, dove `nomep` rappresenta il nome del parametro. Quando vengono utilizzati nelle espressioni CLEM, i parametri devono essere racchiusi tra virgolette singole, per esempio `'$P-scale'`.

I parametri disponibili possono essere facilmente visualizzati mediante il Generatore di espressioni. Per visualizzare i parametri correnti:

1. In qualsiasi finestra di dialogo che accetta espressioni CLEM, fare clic sul pulsante Generatore espressioni.
2. Selezionare **Parametri** dall'elenco Campi.

È quindi possibile selezionare dall'elenco i parametri da inserire nell'espressione CLEM. Consultare l'argomento "Selezione di campi, parametri e variabili globali" a pagina 80 per ulteriori informazioni.

Utilizzo di stringhe

Esistono svariate operazioni disponibili per le stringhe, tra cui:

- Conversione di una stringa in caratteri maiuscoli o minuscoli: `upperlower(CHAR)`.
- Rimozione dei caratteri specificati, ad esempio ``ID_`` o ``$``, da una variabile stringa: `stripchar(CHAR,STRING)`.
- Determinazione della lunghezza (numero di caratteri) di una variabile stringa, `length(STRING)`.
- Controllo dell'ordine alfabetico dei valori stringa, `alphabefore(STRING1, STRING2)`.
- Rimozione dello spazio vuoto iniziale o finale dai valori, `trim(STRING)`, `trim_start(STRING)` o `trimend(STRING)`.
- Estrazione dei primi o ultimi *n* caratteri da una stringa: `startstring(LENGTH, STRING)` o `endstring(LENGTH, STRING)`. Si supponga, per esempio, di disporre di un campo denominato *articolo* che combina il nome di un prodotto con un codice ID di 4 cifre (CAMERA ACME-D109). Per creare un nuovo campo contenente solo il codice di 4 cifre, specificare la formula seguente in un nodo Ricava: `endstring(4, articolo)`
- Corrispondenza di un modello specifico; `STRING matches PATTERN`. Per esempio, per selezionare le persone il cui titolo professionale include "market", è possibile specificare quanto segue in un nodo Selezione:
posizione matches `"*market*"`
- Sostituzione di tutte le istanze di una sottostringa all'interno di una stringa: `replace(SUBSTRING, NEWSUBSTRING, STRING)`. Per esempio, per sostituire tutte le istanze di un carattere non supportato, come una barra verticale (|), con un punto e virgola prima del text mining, utilizzare la funzione `replace` in un nodo Riempimento. In **Riempi campi**, selezionare tutti i campi in cui il carattere può essere presente. Impostare la condizione **Sostituisci** su **Sempre** e specificare la condizione seguente per **Sostituisci con**:
`replace('|',';',@FIELD)`
- Derivazione di un campo flag in base alla presenza di una specifica sottostringa. Per esempio, sarebbe possibile utilizzare una funzione di stringa in un nodo Ricava per generare un campo flag separato per ciascuna risposta con un'espressione quale:
`hassubstring(museums,"museum_of_design")`

Consultare l'argomento "Funzioni stringa" a pagina 99 per ulteriori informazioni.

Gestione di vuoti e valori mancanti

La sostituzione dei vuoti o dei valori mancanti è una comune attività di preparazione dei dati per i data miner. In CLEM sono disponibili diversi strumenti per automatizzare la gestione dei vuoti. Il nodo Riempimento è l'ambito più comune per la gestione dei vuoti. È tuttavia possibile utilizzare le seguenti funzioni in qualsiasi nodo che accetti le espressioni CLEM:

- @BLANK(FIELD) può essere utilizzata per determinare i record i cui valori sono vuoti per un determinato campo, quale *Age*.
- @NULL(FIELD) può essere utilizzata per determinare i record i cui valori sono valori mancanti di sistema per i campi specificati. In IBM SPSS Modeler, i valori mancanti di sistema sono visualizzati come valori \$null\$.

Consultare l'argomento "Funzioni per la gestione di valori vuoti e null" a pagina 114 per ulteriori informazioni.

Utilizzo di numeri

In IBM SPSS Modeler sono disponibili numerose operazioni standard sui valori numerici, quali:

- Calcolo del seno dell'angolo specificato - $\sin(\text{NUM})$.
- Calcolo del logaritmo naturale dei campi numerici - $\log(\text{NUM})$
- Calcolo della somma di due numeri - $\text{NUM1} + \text{NUM2}$

Consultare l'argomento "Funzioni numeriche" a pagina 96 per ulteriori informazioni.

Utilizzo di date e ore

I formati di data e ora possono variare a seconda della sorgente dati e delle impostazioni internazionali. I formati di data e ora sono specifici di ogni stream e sono impostati nella finestra di dialogo delle proprietà del flusso. I seguenti sono esempi di funzioni comunemente utilizzate per i campi di data/ora.

Calcolo del tempo trascorso

È possibile calcolare facilmente il tempo trascorso da una data di riferimento tramite una famiglia di funzioni simili a quella seguente, che restituisce il tempo, in mesi, dalla data di riferimento alla data rappresentata dalla stringa della data DATE, come numero reale. Tale numero corrisponde a una cifra approssimativa basata su un mese formato da 30 giorni.

`date_in_months(Date)`

Confronto di valori Data/Ora

È possibile confrontare i valori dei campi di data/ora nei record tramite funzioni simili a quella seguente, che restituisce il valore *vero*, se la stringa della data DATE1 rappresenta una data antecedente a quella rappresentata dalla stringa della data DATE2. In caso contrario, questa funzione restituirà un valore 0.

`date_before(Date1, Date2)`

Calcolo di differenze

È anche possibile calcolare la differenza tra due ore e due date tramite funzioni quali:

`date_weeks_difference(Date1, Date2)`

Questa funzione restituisce il tempo, in settimane, dalla data rappresentata dalla stringa della data DATE1 alla data rappresentata dalla stringa della data DATE2, come numero reale. Tale numero è basato su una settimana formata da sette giorni. Se DATE2 è antecedente a DATE1, questa funzione restituisce un numero negativo.

Data odierna

È possibile aggiungere la data corrente all'insieme di dati tramite la funzione @TODAY. La data odierna viene aggiunta come stringa al campo specificato o al nuovo campo con il formato di data selezionato nella finestra di dialogo delle proprietà del flusso. Consultare l'argomento "Funzioni di data e ora" a pagina 105 per ulteriori informazioni.

Riepilogo di più campi

Il linguaggio CLEM include una serie di funzioni che restituiscono statistiche riepilogative di più campi. Tali funzioni possono essere particolarmente utili per l'analisi di dati di indagine in cui è possibile che risposte multiple a una singola domanda siano state archiviate in più campi. Consultare l'argomento "Utilizzo di dati a risposta multipla" a pagina 78 per ulteriori informazioni.

Funzioni di confronto

È possibile confrontare valori di più campi utilizzando le funzioni min_n e max_n, per esempio:

```
max_n(['importocarta1' 'importocarta2' 'importocarta3' 'importocarta4'])
```

È inoltre possibile utilizzare una serie di funzioni di conteggio per ottenere conteggi di valori in base a determinati criteri, anche quando tali valori sono archiviati in più campi. Per esempio, per contare il numero di carte che sono state possedute per oltre cinque anni:

```
count_greater_than(5, ['possessocarta' 'possessocarta2' 'possessocarta3'])
```

Per contare valori null nello stesso insieme di campi:

```
count_nulls(['possessocarta' 'possessocarta2' 'possessocarta3'])
```

Si noti che nell'esempio viene conteggiato il numero di carte possedute e non il numero di possessori. Consultare l'argomento "Funzioni di confronto" a pagina 93 per ulteriori informazioni.

Per conteggiare il numero di occorrenze di un valore specifico in più campi, è possibile utilizzare la funzione count_equal. Nell'esempio seguente viene conteggiato il numero di campi dell'elenco che contengono il valore Y.

```
count_equal("Y", [Answer1, Answer2, Answer3])
```

Utilizzando i seguenti valori per i campi dell'elenco, la funzione restituisce i risultati per il valore Y come indicato di seguito.

Tabella 4. Valori funzioni.

Answer1	Answer2	Answer3	Count
Y	N	Y	2
Y	N	N	1

Funzioni numeriche

È possibile ottenere statistiche di più campi utilizzando le funzioni sum_n, mean_n e sdev_n, per esempio:

```
sum_n(['saldocarta1' 'saldocarta2' 'saldocarta3'])
```

```
mean_n(['saldocarta1' 'saldocarta2' 'saldocarta3'])
```

Consultare l'argomento "Funzioni numeriche" a pagina 96 per ulteriori informazioni.

Creazione di elenchi di campi

Quando si utilizza una qualsiasi funzione che accetta un elenco di campi come input, è possibile utilizzare come input le funzioni speciali @FIELDS_BETWEEN(iniziale, finale) e @FIELDS_MATCHING(schema). Per esempio, supponendo che l'ordine dei campi sia quello mostrato nell'esempio sum_n precedente, la seguente espressione sarebbe equivalente:

```
sum_n(@FIELDS_BETWEEN(saldocarta1, saldocarta3))
```

In alternativa, per contare il numero di valori null in tutti i campi che iniziano con "carta":

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

Consultare l'argomento "Campi speciali" a pagina 115 per ulteriori informazioni.

Utilizzo di dati a risposta multipla

Per analizzare dati a risposta multipla è possibile utilizzare una serie di funzioni di confronto, quali:

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

Si supponga, per esempio, una domanda a risposta multipla sulla prima, seconda e terza più importante ragione per decidere su un determinato acquisto (come prezzo, consiglio personale, commenti su pubblicazioni specializzate, fornitore locale, altro). In tal caso è possibile determinare l'importanza del prezzo derivando l'indice del campo in cui è stato incluso per primo:

```
first_index("prezzo", [Motivo1 Motivo2 Motivo3])
```

Analogamente, si supponga di avere chiesto ai clienti di classificare tre auto in ordine di probabilità di acquisto e di avere codificato le risposte in tre campi separati, come segue:

Tabella 5. Esempio classifica automobili.

ID cliente	auto1	auto2	auto3
101	1	3	2
102	3	2	1
103	2	3	1

In tal caso è possibile determinare l'indice del campo dell'auto preferita (classificata al numero 1 o con il rango inferiore) utilizzando la funzione min_index:

```
min_index(['auto1' 'auto2' 'auto3'])
```

Consultare l'argomento "Funzioni di confronto" a pagina 93 per ulteriori informazioni.

Riferimento agli insiemi di risposte multiple

La funzione speciale @MULTI_RESPONSE_SET può essere utilizzata per fare riferimento a tutti i campi in un insieme di risposte multiple. Per esempio, se i tre campi auto dell'esempio precedente venissero inclusi in un insieme di risposte multiple denominato classifica_auto, l'espressione seguente restituirebbe lo stesso risultato:

```
max_index(@MULTI_RESPONSE_SET("classifica_auto"))
```

Generatore di espressioni

È possibile immettere le espressioni CLEM manualmente oppure utilizzare il Builder di espressioni, che visualizza un elenco completo di funzioni e operatori CLEM, nonché i campi di dati del flusso corrente, consentendo di creare espressioni in modo rapido senza la necessità di memorizzare il nome esatto di campi o funzioni. Inoltre il Generatore esegue controlli automatici e aggiunge le virgolette appropriate a campi e valori, facilitando la creazione di espressioni sintatticamente corrette.

Nota: il Generatore di espressioni non è supportato nelle impostazioni di script o di parametri.

Accesso al Generatore di espressioni

Il Generatore di espressioni è disponibile in tutti i nodi in cui vengono utilizzate espressioni CLEM, inclusi i nodi seleziona, bilanciamento, nuovo campo, riempimento, analisi, report e tabella. Per aprire il Generatore di espressioni, fare clic sul pulsante del calcolatore immediatamente a destra del campo della formula.

Creazione di espressioni

Il Generatore di espressioni non solo fornisce elenchi completi di campi, funzioni e operatori, ma consente anche di accedere ai valori dei dati, se i dati sono istanziati.

Per creare un'espressione tramite il Generatore di espressioni

1. Digitare nel campo dell'espressione, utilizzando gli elenchi di funzioni e campi come riferimenti.
o
2. Selezionare i campi e le funzioni richiesti dagli elenchi scorrevoli.
3. Fare doppio clic o fare clic sul pulsante con la freccia gialla per aggiungere il campo o la funzione al campo dell'espressione.
4. Utilizzare i pulsanti degli operandi al centro della finestra di dialogo per inserire le operazioni nell'espressione.

Selezione di funzioni

Nell'elenco di funzioni sono visualizzate tutte le funzioni e gli operatori CLEM disponibili. Scorrere l'elenco per selezionare una funzione o, per effettuare la ricerca in modo più semplice, utilizzare l'elenco a discesa per visualizzare un sottoinsieme di funzioni o di operatori. Le funzioni disponibili sono state raggruppate in categorie, per semplificare la ricerca.

La maggior parte di queste categorie sono descritte nella sezione Riferimento della descrizione del linguaggio CLEM. Consultare l'argomento "Riferimenti alle funzioni" a pagina 90 per ulteriori informazioni.

Le altre categorie sono le seguenti:

- **Funzioni generali** contiene una selezione di alcune delle funzioni utilizzate più comunemente.
- **Elementi usati di recente** contiene un elenco delle funzioni CLEM utilizzate nella sessione corrente.
- **@ Functions** contiene un elenco di tutte le funzioni speciali, il cui nome è preceduto dal simbolo della chiocciolina "@".
- **Funzioni database.** Se il flusso include una connessione di database (mediante un nodo origine del database), questa selezione elenca le funzioni disponibili all'interno del database, incluse le funzioni definite dall'utente (UDF).
- **Aggregati database.** Se il flusso include una connessione al database (mediante un nodo origine database), questa selezione elenca le opzioni di aggregazione disponibili all'interno del database.
- **Aggregati finestra database.** Se il flusso include una connessione al database (mediante un nodo origine database), questa selezione elenca le opzioni di aggregazione della finestra che è possibile utilizzare all'interno del database.

Sebbene chiamati "aggregati" queste opzioni non sono progettate per l'utilizzo nel nodo di aggregazione; sono meglio applicabili ai nodi come i nodi Nuovo campo o nodi selezione. Questo perché il loro output è scalare invece di essere un vero aggregato; così che non riducono la quantità di dati mostrati nell'output nello stesso modo in cui lo fanno i nodi di aggregazione. Ad esempio, si potrebbe utilizzare questo ordinamento di aggregazione per fornire una media mobile in basso attraverso le righe dei dati come ad esempio "media della riga corrente e delle righe precedenti".

- **Aggregati incorporati.** Contiene un elenco dei possibili nodi di aggregazione che è possibile utilizzare.
- **Operatori** elenca tutti gli operatori utilizzabili quando si creano le espressioni. Gli operatori sono disponibili anche dai pulsanti al centro della finestra di dialogo.
- **Tutte le funzioni** contiene un elenco completo delle funzioni CLEM disponibili.

Dopo avere selezionato un gruppo di funzioni, fare doppio clic per inserire le funzioni nel campo dell'espressione nel punto indicato dalla posizione del cursore.

Selezione di campi, parametri e variabili globali

Nell'elenco dei campi sono visualizzati tutti i campi disponibili in questo punto del flusso di dati. Scorrere l'elenco per selezionare un campo. Fare doppio clic o utilizzare il pulsante con la freccia gialla per aggiungere un campo all'espressione.

Consultare l'argomento "Parametri flusso, sessione e Supernodo" a pagina 74 per ulteriori informazioni.

Oltre ai campi, è possibile scegliere tra i seguenti elementi:

Insiemi di risposte multiple. Per maggiori informazioni, consultare il manuale *Nodi origine, di elaborazione e di output in IBM SPSS Modeler*.

Elementi usati di recente contiene un elenco di campi, insieme a risposta multipla, parametri e valori globali utilizzati all'interno della sessione corrente.

Parametri. Consultare l'argomento "Parametri flusso, sessione e Supernodo" a pagina 74 per ulteriori informazioni.

valori globali. Per maggiori informazioni, consultare il manuale *Nodi origine, di elaborazione e di output in IBM SPSS Modeler*.

Visualizzazione o selezione di valori

I valori dei campi possono essere visualizzati da svariati punti del sistema, inclusi il Generatore di espressioni e i report del nodo Esplora, e nel corso della modifica dei valori futuri in un nodo Intervalli di tempo. Si noti che per utilizzare questa funzionalità è necessario che i dati vengano completamente istanziati in un nodo origine o Tipo, in modo che l'archiviazione, i tipi e i valori siano noti.

Per visualizzare i valori dal Generatore di espressioni o da un nodo Intervalli di tempo, selezionare il campo richiesto e fare clic sul pulsante di selezione dei valori per aprire una finestra di dialogo contenente i valori del campo selezionato. È quindi possibile selezionare un valore e fare clic su **Inserisci** per incollarlo nell'espressione o elenco corrente.



Figura 14. Pulsante di selezione dei valori

Per i campi flag e nominali, vengono elencati tutti i valori definiti. Per i campi continui (intervallo numerico), vengono visualizzati i valori minimi e massimi.

Controllo di espressioni CLEM

Per convalidare l'espressione, fare clic su **Controlla** nel Generatore di espressioni (angolo inferiore destro). Le espressioni non controllate sono visualizzate in rosso. Se vengono rilevati degli errori, viene visualizzato un messaggio che ne comunica la causa.

Vengono controllati i seguenti elementi:

- Inserimento corretto tra virgolette di valori e nomi di campi
- Utilizzo corretto di parametri e variabili globali
- Utilizzo valido di operatori
- Esistenza di campi di riferimento
- Esistenza e definizione di valori globali di riferimento

Se vengono rilevati errori di sintassi, provare a ricreare l'espressione utilizzando gli elenchi e i pulsanti degli operatori anziché la digitazione manuale. In questo modo le virgolette appropriate per campi e valori verranno aggiunte automaticamente.

Trova e sostituisci

La finestra di dialogo Trova/Sostituisci è disponibile ogni volta che è possibile modificare il testo di script o di espressioni, compreso l'editor di script, il generatore di espressioni CLEM e quando si definisce un modello nel nodo Report. Quando si modifica un testo in una di queste aree, premere Ctrl+F per accedere alla finestra di dialogo, assicurandosi che il cursore sia posizionato in un'area di testo. In un nodo Riempimento, per esempio, è possibile accedere alla finestra di dialogo da qualsiasi area di testo della scheda Impostazioni oppure dal campo testo nel Generatore di espressioni.

1. Con il cursore posizionato in un'area di testo, premere Ctrl+F per accedere alla finestra di dialogo Trova/Sostituisci.
2. Immettere il testo da cercare oppure sceglierne uno dall'elenco a discesa degli elementi cercati di recente.
3. Se necessario, immettere il testo sostitutivo.
4. Fare clic su **Trova successivo** per avviare la ricerca.
5. Fare clic su **Sostituisci** per sostituire la selezione corrente oppure scegliere **Sostituisci tutto** per aggiornare tutte le istanze o quelle selezionate.
6. Al termine di ogni operazione, la finestra di dialogo si chiude. Premere F3 da qualsiasi area di testo per ripetere l'ultima operazione di ricerca oppure premere Ctrl+F per accedere nuovamente alla finestra di dialogo.

Opzioni di ricerca

Caratteri maiuscoli/minuscoli. Specifica se l'operazione di ricerca fa distinzione tra caratteri maiuscoli/minuscoli, per esempio se *miavar* corrisponde a *miaVar*. Il testo sostitutivo viene sempre inserito esattamente come viene digitato, indipendentemente da questa impostazione.

Solo parole intere. Specifica se l'operazione di ricerca cerca le occorrenze che sono parole intere. Se questa opzione è selezionata, la ricerca di *palla* non consentirà di trovare per esempio *pallavolo* o *Palladio*.

Espressioni regolari. Specifica se è utilizzata la sintassi delle espressioni regolari (vedere la sezione seguente). Quando questa opzione è selezionata, l'opzione **Solo parole intere** è disattivata e il relativo valore viene ignorato.

Solo testo selezionato. Controlla l'ambito della ricerca quando si utilizza l'opzione **Sostituisci tutto**.

Sintassi delle espressioni regolari

Le espressioni regolari consentono di cercare caratteri speciali, quali tabulazioni o caratteri di nuova riga, classi o intervalli di caratteri quali *a - d*, cifre e caratteri diversi da cifre, nonché limiti, per esempio l'inizio o la fine di una riga. Sono supportati i seguenti tipi di espressioni.

Tabella 6. Corrispondenze di caratteri.

Caratteri	Corrispondenze
x	Il carattere x
\\	Il carattere barra rovesciata
\0n	Il carattere con valore ottale 0n (0 <= n <= 7)
\0nn	Il carattere con valore ottale 0nn (0 <= n <= 7)
\0mnn	Il carattere con valore ottale 0mnn (0 <= m <= 3, 0 <= n <= 7)
\xhh	Il carattere con valore esadecimale 0xhh
\uhhhh	Il carattere con valore esadecimale 0xhhhh
\t	Il carattere di tabulazione ('\u0009')
\n	Il carattere di nuova riga (avanzamento riga) ('\u000A')
\r	Il carattere di ritorno a capo ('\u000D')
\f	Il carattere di avanzamento carta ('\u000C')
\a	Il carattere di avviso (campanello) ('\u0007')
\e	Il carattere di escape ('\u001B')
\cx	Il carattere di controllo corrispondente a x

Tabella 7. Corrispondenze di classi di caratteri.

Classi di caratteri	Corrispondenze
[abc]	a, b o c (classe semplice)
[^abc]	Qualsiasi carattere, eccetto a, b o c (sottrazione)
[a-zA-Z]	a-z oppure A-Z, incluse (intervallo)
[a-d[m-p]]	a-d oppure m-p (unione). In alternativa è possibile specificare [a-dm-p].
[a-z&&[def]]	a-z + d, e oppure f (intersezione)
[a-z&&[^bc]]	a-z, eccetto b e c (sottrazione). In alternativa è possibile specificare [ad-z].
[a-z&&[^m-p]]	a-z, eccetto m-p (sottrazione). In alternativa è possibile specificare [a-lq-z].

Tabella 8. Classi di caratteri predefinite.

Classi di caratteri predefinite	Corrispondenze
.	Qualsiasi carattere (può corrispondere o meno a terminazioni di riga)
\d	Qualsiasi cifra: [0-9]
\D	Un carattere diverso da una cifra: [^0-9]
\s	Un spazio vuoto: [\n\x0B \t \f\r]
\S	Uno spazio non vuoto: [^\s]
\w	Un carattere alfanumerico: [a-zA-Z_0-9]
\W	Un carattere diverso da alfanumerico: [^\w]

Tabella 9. Corrispondenze di limiti.

Corrispondenze di limiti	Corrispondenze
^	L'inizio di una riga

Tabella 9. Corrispondenze di limiti (Continua).

Corrispondenze di limiti	Corrispondenze
\$	La fine di una riga
\b	Un limite di parola
\B	Un limite diverso da un limite di parola
\A	L'inizio dell'input
\Z	La fine dell'input ma per la terminazione finale, se disponibile
\z	La fine dell'input

Capitolo 8. Riferimenti al linguaggio CLEM

Panoramica di riferimento su CLEM

Questa sezione illustra il linguaggio CLEM (Control Language for Expression Manipulation), un potente strumento per l'analisi e la manipolazione dei dati utilizzati nei flussi di IBM SPSS Modeler. È possibile utilizzare CLEM all'interno di nodi per eseguire svariate attività, dalla valutazione di condizioni o derivazione di valori fino all'inserimento di dati in report.

Le espressioni CLEM sono costituite da valori, nomi di campi, operatori e funzioni. Tramite la sintassi corretta è possibile creare una vasta gamma di potenti operazioni sui dati.

Tipi di dati CLEM

Nel linguaggio CLEM è possibile utilizzare i seguenti tipi di dati:

- Interi
- Numeri reali
- Caratteri
- Stringhe
- Elenchi
- Campi
- Data/ora

Regole per l'inserimento tra virgolette

Sebbene IBM SPSS Modeler sia flessibile quando l'utente determina campi, valori, parametri e stringhe utilizzati in un'espressione CLEM, di seguito sono suggerite alcune "procedure consigliate" da seguire durante la creazione di espressioni.

- Stringhe: utilizzare sempre virgolette per la scrittura di stringhe, per esempio "Tipo 2". È inoltre possibile utilizzare le virgolette singole, ma il rischio è di creare confusione con i campi racchiusi tra virgolette.
- Campi: utilizzare virgolette singole solo se necessarie per racchiudere spazi o altri caratteri speciali, per esempio 'Numero ordine'. I campi racchiusi tra virgolette ma non definiti nell'insieme di dati verranno interpretati come stringhe.
- Parametri: utilizzare sempre virgolette singole quando si specificano parametri, per esempio '\$P-soglia'.
- Per i caratteri devono essere utilizzate virgolette singole (`), per esempio stripchar(`d`, "curaA").

Queste regole verranno illustrate in modo più dettagliato negli argomenti riportati di seguito.

Interi

I numeri interi vengono rappresentati come sequenze di cifre decimali. Facoltativamente, è possibile inserire un segno meno (-) prima del numero reale per indicare un numero negativo, ad esempio, 1234, 999, -77.

Il linguaggio CLEM è in grado di gestire numeri interi a precisione arbitraria. La dimensione massima dei numeri interi dipende dalla piattaforma in uso. Se i valori hanno dimensioni superiori a quelle visualizzabili in un campo di numeri interi, l'impostazione del tipo di campo su Reale consente in genere di ripristinare il valore.

Numeri reali

Con *numero reale* si intende un numero a virgola mobile. I numeri reali sono rappresentati da una o più cifre seguite da un punto decimale, seguito a sua volta da una o più cifre. Nel linguaggio CLEM, vengono utilizzati numeri reali con precisione doppia.

Facoltativamente, è possibile inserire un segno meno (-) prima del numero reale per indicare un numero negativo, ad esempio, 1,234, 0,999, -77,001. Utilizzare il formato `<numero>` e `<esponente>` per esprimere un numero reale con la notazione esponenziale, ad esempio, 1234,0e5, 1,7e-2. Quando le stringhe di numeri vengono lette dai file e convertite automaticamente in numeri dall'applicazione IBM SPSS Modeler i numeri senza cifre iniziali prima del separatore decimale e senza cifre dopo il separatore vengono accettati, ad esempio 999. o .11. Questi formati non sono tuttavia validi nelle espressioni CLEM.

Nota: quando si fa riferimento a numeri reali in espressioni CLEM, è necessario utilizzare il punto come separatore decimale indipendentemente da eventuali impostazioni del flusso corrente o dalle impostazioni internazionali. Per esempio, specificare

```
Na > 0.6
```

anziché

```
Na > 0,6
```

Questa regola è valida anche se nella finestra di dialogo delle proprietà del flusso è stata selezionata la virgola come simbolo decimale ed è conforme alla linea guida generale secondo cui la sintassi del codice deve essere indipendente da specifiche impostazioni internazionali o convenzioni.

Caratteri

I caratteri sono in genere visualizzati come CHAR e nelle espressioni CLEM consentono di eseguire test sulle stringhe. Per esempio, la funzione `isuppercode` consente di stabilire se il primo carattere di una stringa è maiuscolo. L'espressione CLEM seguente utilizza un carattere per indicare che il test deve essere eseguito sul primo carattere della stringa:

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

Per esprimere il codice (in la posizione) di un carattere specifico in un'espressione CLEM, utilizzare virgolette singole nel formato ``<carattere>``, ad esempio, ``A``, ``Z``.

Nota: non esiste un tipo di archiviazione CHAR per un campo, pertanto se un campo deriva o viene riempito con un'espressione che produce un tipo CHAR, il risultato verrà convertito in una stringa.

Stringhe

È in genere opportuno racchiudere le stringhe tra virgolette. Tra gli esempi di stringhe sono inclusi "c35prodotto2" e "IDreferente". Per indicare caratteri speciali in una stringa, utilizzare una barra rovesciata, per esempio "\\$65443". (Per indicare una barra rovesciata, utilizzare una doppia barra rovesciata, \\). È possibile racchiudere una stringa tra virgolette singole, ma in questo caso il risultato (per esempio 'IDreferente') non è distinguibile da un campo tra virgolette. Consultare l'argomento "Funzioni stringa" a pagina 99 per ulteriori informazioni.

Elenchi

Un elenco rappresenta una sequenza ordinata di elementi, che possono essere di tipo eterogeneo. Gli elenchi sono racchiusi tra parentesi quadre ([]). Tra gli esempi di elenchi sono inclusi [1 2 4 16] e ["abc" "def"]. Gli elenchi non vengono utilizzati come valori dei campi di IBM SPSS Modeler, bensì per fornire argomenti alle funzioni, quali `member` e `oneof`.

Campi

I nomi inclusi nelle espressioni CLEM non corrispondenti a nomi di funzioni vengono interpretati come nomi di campi. Questi possono essere scritti semplicemente come `Power`, `val27`, `state_flag` e così via. Tuttavia, se i nomi iniziano con una cifra o comprendono caratteri non alfabetici come gli spazi (ad eccezione del carattere di sottolineatura), è necessario inserire il nome tra virgolette singole, per esempio `'Power Increase'`, `'2nd answer'`, `'#101'` e `'$P-NextField'`.

Nota: i campi racchiusi tra virgolette ma non definiti nel dataset verranno interpretati come stringhe.

Date

I calcoli di data sono basati su una data di riferimento, specificata nella finestra di dialogo delle proprietà del flusso. La data di riferimento di default è 1 gennaio 1900.

Il linguaggio CLEM supporta i formati di data seguenti.

Tabella 10. Formati data lingua CLEM

Formato	Esempi
GGMAAA	150163
MMGGAA	011563
AAMMGG	630115
AAAAMGG	19630115
AAAAGGG	Anno a quattro cifre seguito da numero a tre cifre rappresentante il giorno dell'anno, per esempio 2000032 rappresenta il 32esimo giorno del 2000 oppure l'1 febbraio 2000.
DAY	Giorno della settimana nelle opzioni internazionali correnti, per esempio, Lunedì, Martedì, ..., ecc., in inglese
MONTH	Mese nella locale corrente, ad esempio, Gennaio, Febbraio,
GG/MM/AA	15/01/63
GG/MM/AAAA	15/01/1963
MM/GG/AA	01/15/63
MM/GG/AAAA	01/15/1963
GG-MM-AA	15-01-63
GG-MM-AAAA	15-01-1963
MM-GG-AA	01-15-63
MM-GG-AAAA	01-15-1963
GG.MM.AA	15.01.63
GG.MM.AAAA	15.01.1963
MM.GG.AA	01.15.63
MM.GG.AAAA	01.15.1963
GG-MES-AA	15-GEN-63, 15-gen-63, 15-Gen-63
GG/MES/AA	15/GEN/63, 15/gen/63, 15/Gen/63
GG.MES.AA	15.GEN.63, 15.gen.63, 15.Gen.63
GG-MES-AAAA	15-GEN-1963, 15-gen-1963, 15-Gen-1963
GG/MES/AAAA	15/GEN/1963, 15/gen/1963, 15/Gen/1963
GG.MES.AAAA	15.GEN.1963, 15.gen.1963, 15.Gen.1963
MON YYYY	Gen 2004

Tabella 10. Formati data lingua CLEM (Continua)

Formato	Esempi
q Q YYYY	Data rappresentata come cifra (1–4) rappresentante il trimestre seguito dalla lettera <i>T</i> e un anno a quattro cifre, per esempio, 25 dic 2004 viene rappresentato nel seguente modo: 4 T 2004.
ww WK YYYY	Numero a due cifre rappresentante la settimana dell'anno, seguito dalle lettere <i>ST</i> e un anno a quattro cifre. La settimana dell'anno viene calcolata presupponendo che il primo giorno della settimana è Lunedì e che esiste almeno un giorno nella prima settimana.

Ora

Il linguaggio CLEM supporta i formati di ora seguenti.

Tabella 11. Formati ora lingua CLEM

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

Operatori CLEM

Sono disponibili i seguenti operatori:

Tabella 12. Operatori lingua CLEM.

Operazione	Commenti	Precedenza (vedere la sezione seguente)
or	Utilizzata tra due espressioni CLEM. Restituisce il valore vero se una delle due espressioni è vera oppure se entrambe sono vere.	10
and	Utilizzata tra due espressioni CLEM. Restituisce il valore vero se entrambe le espressioni sono vere.	9

Tabella 12. Operatori lingua CLEM (Continua).

Operazione	Commenti	Precedenza (vedere la sezione seguente)
=	Utilizzata tra qualsiasi coppia di elementi confrontabili. Restituisce vero se ITEM1 è uguale a ITEM2.	7
==	Identico a =.	7
/=	Utilizzata tra qualsiasi coppia di elementi confrontabili. Restituisce vero se ITEM1 <i>non</i> è uguale a ITEM2.	7
/==	Identico a /=.	7
>	Utilizzata tra qualsiasi coppia di elementi confrontabili. Restituisce vero se ITEM1 è maggiore di ITEM2.	6
>=	Utilizzata tra qualsiasi coppia di elementi confrontabili. Restituisce vero se ITEM1 è maggiore o uguale a ITEM2.	6
<	Utilizzata tra qualsiasi coppia di elementi confrontabili. Restituisce vero se ITEM1 è minore di ITEM2.	6
<=	Utilizzata tra qualsiasi coppia di elementi confrontabili. Restituisce vero se ITEM1 è minore o uguale a ITEM2.	6
&&= _0	Utilizzata tra due numeri interi. Equivalente dell'espressione Booleano $INT1 \&\& INT2 = 0$.	6
&&/= _0	Utilizzata tra due numeri interi. Equivalente dell'espressione Booleano $INT1 \&\& INT2 \neq 0$.	6
+	Somma due numeri: $NUM1 + NUM2$.	5
>>	Concatena due stringhe, per esempio $STRING1 \>> STRING2$.	5
-	Sottrae un numero dall'altro: $NUM1 - NUM2$. Può inoltre precedere un numero: $-NUM$.	5
*	Utilizzata per moltiplicare due numeri: $NUM1 * NUM2$.	4
&&	Utilizzata tra due numeri interi. Il risultato è l'AND bitwise di INT1 e INT2.	4
&&~	Utilizzata tra due numeri interi. Il risultato è l'AND bitwise di INT1 e il complemento bitwise di INT2.	4
	Utilizzata tra due numeri interi. Il risultato è l'OR inclusivo bitwise di INT1 e INT2.	4
~	Utilizzata prima di un numero intero. Genera il complemento bitwise di INT.	4
/&	Utilizzata tra due numeri interi. Il risultato è l'OR esclusivo bitwise di INT1 e INT2.	4
INT1 << N	Utilizzata tra due numeri interi. Genera lo schema di bit di INT con spostamento a sinistra di N posizioni.	4

Tabella 12. Operatori lingua CLEM (Continua).

Operazione	Commenti	Precedenza (vedere la sezione seguente)
INT1 >> N	Utilizzata tra due numeri interi. Genera lo schema di bit di INT con spostamento a destra di N posizioni.	4
/	Utilizzata per dividere un numero per un altro: NUM1 / NUM2.	4
**	Utilizzata tra due numeri: BASE ** POTENZA. Restituisce BASE elevata alla potenza POTENZA.	3
rem	Utilizzata tra due numeri interi: INT1 rem INT2. Restituisce il resto, INT1 - (INT1 div INT2) * INT2.	2
div	Utilizzata tra due numeri interi: INT1 div INT2. Esegue una divisione di interi.	2

Precedenza degli operatori

Le precedenze influiscono sull'analisi delle espressioni complesse, in particolare delle espressioni non racchiuse tra parentesi e contenenti più operatori infissi. Per esempio,

$3 + 4 * 5$

viene analizzato come $3 + (4 * 5)$ anziché come $(3 + 4) * 5$, poiché in base alle precedenze relative * deve essere analizzato prima di +. Nel linguaggio CLEM, a ogni operatore è associato un valore di precedenza. Minore è il valore, maggiore è l'importanza che gli viene attribuita nell'elenco di analisi. Ciò significa che verrà elaborato prima degli altri operatori con valori di precedenza maggiori.

Riferimenti alle funzioni

Per le operazioni con i dati in IBM SPSS Modeler sono disponibili le seguenti funzioni CLEM. È possibile inserire queste funzioni sotto forma di codice in numerose finestre di dialogo, tra cui i nodi Ricava e Crea flag, oppure utilizzare il Generatore espressioni per creare espressioni CLEM valide senza che sia necessario memorizzare elenchi di funzioni o nomi di campi.

Tabella 13. Funzioni CLEM da utilizzare con dati IBM SPSS Modeler

Tipo di funzione	Descrizione
Informazioni	Utilizzate per ottenere informazioni sui valori dei campi. Per esempio, la funzione <code>is_string</code> restituisce un true per tutti i record il cui tipo è una stringa.
Conversione	Utilizzate per creare nuovi campi o convertire tipi di archiviazione. Per esempio, la funzione <code>to_timestamp</code> converte il campo selezionato in un timestamp.
Confronto	Utilizzate per confrontare i valori dei campi tra di loro oppure con una stringa determinata. Per esempio, <code><=</code> viene utilizzata per confrontare se i valori di due campi sono uno uguale o inferiore all'altro.
Logica	Utilizzate per eseguire operazioni logiche, quali le operazioni <code>if</code> , <code>then</code> , <code>else</code> .
Numerica	Utilizzate per eseguire calcoli numerici, quale il logaritmo naturale dei valori dei campi.
Trigonometriche	Utilizzate per eseguire calcoli trigonometrici, quale l'arcocoseno di un angolo specificato.

Tabella 13. Funzioni CLEM da utilizzare con dati IBM SPSS Modeler (Continua)

Tipo di funzione	Descrizione
Probabilità	Restituiscono probabilità in base a varie distribuzioni, quali la probabilità che un valore dalla distribuzione <i>t</i> per studenti sia inferiore a un determinato valore.
Bitwise	Utilizzate per manipolare i numeri interi come schemi di bit.
Casuale	Utilizzate per la selezione casuale di elementi o per la creazione casuale di numeri.
Stringa	Utilizzate per eseguire numerose operazioni sulle stringhe, per esempio <code>stripchar</code> , che consente di rimuovere un carattere specificato.
SoundEx	Utilizzate per trovare stringhe quando non se ne conosce l'esatta ortografia, in base a ipotesi fonetiche sulla pronuncia di determinate lettere.
Data e orario	Utilizzate per eseguire svariate operazioni sui campi di data, ora e timestamp.
Sequenza	Utilizzate per avere informazioni sulla sequenza di record di un insieme di dati o per eseguire operazioni in base a tale sequenza.
Globali	Utilizzate per accedere a valori globali creati da un nodo Calcola globali. Per esempio, <code>@MEAN</code> viene utilizzata per fare riferimento alla media di tutti i valori di un campo nell'intero insieme di dati.
Vuoti e null	Utilizzate per accedere, contrassegnare e spesso riempire i vuoti specificati dall'utente o i valori mancanti di sistema. Per esempio, la funzione <code>@BLANK(FIELD)</code> viene utilizzata per assegnare un flag true a record contenenti valori vuoti.
Campi speciali	Utilizzate per denotare i campi specifici da esaminare. Per esempio, la funzione <code>@FIELD</code> viene utilizzata quando si derivano più campi.

Convenzioni utilizzate nelle descrizioni delle funzioni

Le convenzioni seguenti vengono utilizzate in questa guida quando fanno riferimento agli elementi in una funzione.

Tabella 14. Convenzioni nelle descrizioni delle funzioni.

Convenzione	Descrizione
<i>BOOL</i>	Valore booleano o flag, per esempio vero o falso.
<i>NUM, NUM1, NUM2</i>	Qualsiasi numero.
<i>REAL, REAL1, REAL2</i>	Qualsiasi numero reale, come 1,234 o -77,01.
<i>INT, INT1, INT2</i>	Qualsiasi numero intero, come 1 o -77.
<i>CHAR</i>	Un codice di carattere come <code>`A`</code> .
<i>STRING</i>	Una stringa come <code>"IDreferente"</code> .
<i>LIST</i>	Un elenco di elementi come <code>["abc" "def"]</code> .
<i>ITEM</i>	Un campo come <code>Cliente</code> oppure <code>estrazione_concetto</code> .
<i>DATE</i>	Un campo di data come <code>data_inizio</code> , i cui valori sono in un formato quale GG-MES-AAAA.
<i>TIME</i>	Un campo temporale come <code>flusso_potenza</code> , i cui valori sono in un formato quale HHMMSS.

Le funzioni in questa guida vengono elencate con la funzione in una colonna, il tipo di risultato (intero, stringa, ecc.) in un'altra e una descrizione (se disponibile) in una terza colonna. Per esempio, di seguito è riportata la descrizione della funzione `rem`.

Tabella 15. descrizione della funzione rem.

Funzione	Risultato	Descrizione
INT1 rem INT2	Numero	Restituisce il resto della divisione di INT1 diviso per INT2. Ad esempio, INT1 – (INT1 div INT2) * INT2.

I dettagli sulle convenzioni di utilizzo, quali la modalità di elencare gli elementi oppure di specificare i caratteri in una funzione, vengono descritti altrove. Consultare l'argomento "Tipi di dati CLEM" a pagina 85 per ulteriori informazioni.

Funzioni di informazione

Le funzioni di informazione vengono utilizzate per ottenere informazioni sui valori di un determinato campo e in genere per derivare campi flag. Per esempio, è possibile utilizzare la funzione @BLANK per creare un campo flag che indica i record contenenti valori vuoti per il campo selezionato. Allo stesso modo, è possibile controllare il tipo di archiviazione per un campo utilizzando qualsiasi funzione relativa al tipo di archiviazione, per esempio is_string.

Tabella 16. Funzioni di informazione CLEM.

Funzione	Risultato	Descrizione
@BLANK(FIELD)	Booleano	Restituisce un valore vero per tutti i record contenenti valori vuoti in base alle regole di gestione dei vuoti impostate in un nodo origine o Tipo a monte mediante la scheda Tipi.
@NULL(ITEM)	Booleano	Restituisce vero per tutti i record contenenti valori non definiti. I valori non definiti sono rappresentati da valori di sistema null, visualizzati in IBM SPSS Modeler come \$null\$.
is_date(ITEM)	Booleano	Restituisce vero per tutti i record il cui tipo è rappresentato da una data.
is_datetime(ITEM)	Booleano	Restituisce vero per tutti i record il cui tipo è rappresentato da una data, un'ora o un timestamp.
is_integer(ITEM)	Booleano	Restituisce vero per tutti i record il cui tipo è rappresentato da un numero intero.
is_number(ITEM)	Booleano	Restituisce vero per tutti i record il cui tipo è rappresentato da un numero.
is_real(ITEM)	Booleano	Restituisce vero per tutti i record il cui tipo è rappresentato da un numero reale.
is_string(ITEM)	Booleano	Restituisce vero per tutti i record il cui tipo è rappresentato da una stringa.
is_time(ITEM)	Booleano	Restituisce vero per tutti i record il cui tipo è rappresentato da un'ora.
is_timestamp(ITEM)	Booleano	Restituisce vero per tutti i record il cui tipo è rappresentato da un timestamp.

Funzioni di conversione

Le funzioni di conversione consentono di creare nuovi campi e convertire il tipo di archiviazione dei file esistenti. Per esempio, per formare nuove stringhe è possibile unire o separare stringhe. Per unire due stringhe, utilizzare l'operatore ><. Per esempio, se al campo Site è assegnato il valore "BRAMLEY", "xx" >< Site restituirà "xxBRAMLEY". Il risultato di >< è sempre rappresentato da una stringa, anche se gli argomenti non sono stringhe. Di conseguenza, se il campo V1 è 3 e il campo V2 è 5, V1 >< V2 restituirà "35", come stringa e non come numero.

Le funzioni di conversione, e qualsiasi altra funzione che richieda un tipo di input specifico, quale un valore di data o ora, dipendono dai formati correnti specificati nella finestra di dialogo Opzioni flusso. Per esempio, se si desidera convertire un campo stringa con valori *Gen 2003, Feb 2003* e così via, selezionare il formato di data corrispondente **MES AAAA** come formato di data di default per il flusso.

Tabella 17. Funzioni di conversione CLEM

Funzione	Risultato	Descrizione
ITEM1 >< ITEM2	Stringa	Esegue la concatenazione dei valori di due campi e restituisce la stringa risultante come <i>ITEM1ITEM2</i> .
to_integer(ITEM)	Numero intero	Converte il tipo di archiviazione del campo specificato in un numero intero.
to_real(ITEM)	Numero reale	Converte il tipo di archiviazione del campo specificato in un numero reale.
to_number(ITEM)	Numero	Converte il tipo di archiviazione del campo specificato in un numero.
to_string(ITEM)	Stringa	Converte il tipo di archiviazione del campo specificato in una stringa.
to_time(ITEM)	Ora	Converte il tipo di archiviazione del campo specificato in un'ora.
to_date(ITEM)	Data	Converte il tipo di archiviazione del campo specificato in una data.
to_timestamp(ITEM)	Timestamp	Converte il tipo di archiviazione del campo specificato in un timestamp.
to_datetime(ITEM)	Data e ora	Converte il tipo di archiviazione del campo specificato in un valore di data, ora o timestamp.
datetime_date(ITEM)	Data	Restituisce il valore di data per un <i>numero</i> , una <i>stringa</i> o un <i>timestamp</i> . Si noti che questa è la sola funzione che consente di convertire un numero (in secondi) nuovamente in una data. Se ITEM è una stringa, crea una data analizzando una stringa nel formato di data corrente. La funzione verrà eseguita correttamente solo se il formato di data specificato nella finestra di dialogo Proprietà flusso è corretto. Se ITEM è un numero, viene interpretato come numero di secondi dalla data (o epoca) di base. Le frazioni di giorno vengono troncate. Se ITEM è un timestamp, viene restituita la parte relativa alla data del timestamp. Se ITEM è una data, la data viene restituita invariata.

Funzioni di confronto

Le funzioni di confronto vengono utilizzate per confrontare i valori dei campi tra di loro oppure con una stringa specificata. Per esempio, è possibile controllare l'uguaglianza delle stringhe utilizzando =. Un esempio di verifica dell'uguaglianza delle stringhe è costituito da: `Class = "classe 1"`.

Ai fini del confronto numerico, *maggiore* indica una maggiore vicinanza all'infinito positivo, mentre *minore* corrisponde a una maggiore prossimità all'infinito negativo. Tutti i numeri negativi sono quindi minori di qualsiasi numero positivo.

Tabella 18. Funzioni di confronto CLEM

Funzione	Risultato	Descrizione
count_equal(ITEM1, LIST)	Numero intero	Restituisce il numero di valori da un elenco di campi che sono uguali a <i>ELEMENTO1</i> o null se <i>ELEMENTO1</i> è null.
count_greater_than(ITEM1, LIST)	Numero intero	Restituisce il numero di valori da un elenco di campi che sono maggiori di <i>ELEMENTO1</i> o null se <i>ELEMENTO1</i> è null.

Tabella 18. Funzioni di confronto CLEM (Continua)

Funzione	Risultato	Descrizione
count_less_than(ITEM1, LIST)	Numero intero	Restituisce il numero di valori da un elenco di campi che sono minori di <i>ELEMENTO1</i> o null se <i>ELEMENTO1</i> è null.
count_not_equal(ITEM1, LIST)	Numero intero	Restituisce il numero di valori da un elenco di campi che non sono uguali a <i>ELEMENTO1</i> o null se <i>ELEMENTO1</i> è null.
count_nulls(LIST)	Numero intero	Restituisce il numero di valori null da un elenco di campi.
count_non_nulls(LIST)	Numero intero	Restituisce il numero di valori non null da un elenco di campi.
date_before(DATA1, DATA2)	Booleano	Viene utilizzata per controllare l'ordine dei valori di data. Restituisce un valore vero se <i>DATE1</i> precede <i>DATE2</i> .
first_index(ITEM, LIST)	Numero intero	Restituisce l'indice del primo campo contenente <i>ITEM</i> da un <i>LIST</i> di campi oppure 0 se il valore non viene trovato. Supportata solo per tipi stringa, numero intero e numero reale.
first_non_null(LIST)	Qualsiasi	Restituisce il primo valore non null nell'elenco di campi fornito. Sono supportati tutti i tipi di archiviazione.
first_non_null_index(LIST)	Numero intero	Restituisce l'indice del primo campo dell' <i>LIST</i> specificato contenente un valore non null oppure 0 se tutti i valori sono null. Sono supportati tutti i tipi di archiviazione.
ITEM1 = ITEM2	Booleano	Restituisce vero per i record in cui <i>ITEM1</i> è uguale a <i>ITEM2</i> .
ITEM1 /= ITEM2	Booleano	Restituisce vero se le due stringhe non sono identiche e 0 se sono identiche.
ELEMENTO1 < ELEMENTO2	Booleano	Restituisce vero per i record in cui <i>ITEM1</i> è minore di <i>ITEM2</i> .
ELEMENTO1 <= ELEMENTO2	Booleano	Restituisce vero per i record in cui <i>ITEM1</i> è minore o uguale a <i>ITEM2</i> .
ITEM1 > ITEM2	Booleano	Restituisce vero per i record in cui <i>ITEM1</i> è maggiore di <i>ITEM2</i> .
ITEM1 >= ITEM2	Booleano	Restituisce vero per i record in cui <i>ITEM1</i> è maggiore o uguale a <i>ITEM2</i> .
last_index(ITEM, LIST)	Numero intero	Restituisce l'indice dell'ultimo campo contenente <i>ITEM</i> da un <i>LIST</i> di campi oppure 0 se il valore non viene trovato. Supportata solo per tipi stringa, numero intero e numero reale.
last_non_null(LIST)	Qualsiasi	Restituisce l'ultimo valore non null nell'elenco di campi fornito. Sono supportati tutti i tipi di archiviazione.
last_non_null_index(LIST)	Numero intero	Restituisce l'indice dell'ultimo campo dell' <i>LIST</i> specificato contenente un valore non null oppure 0 se tutti i valori sono null. Sono supportati tutti i tipi di archiviazione.
max(ELEMENTO1, ELEMENTO2)	Qualsiasi	Restituisce il maggiore dei due elementi, <i>ITEM1</i> o <i>ITEM2</i> .
max_index(LIST)	Numero intero	Restituisce l'indice del campo contenente il valore massimo da un elenco di campi numerici oppure 0 se tutti i valori sono null. Per esempio, se il terzo campo elencato contiene il valore massimo, viene restituito il valore di indice 3. Se più campi contengono il valore massimo, viene restituito quello elencato per primo (più a sinistra).
max_n(LIST)	Numero	Restituisce il valore massimo da un elenco di campi numerici oppure null se tutti i valori di campo sono null.
member(ELEMENTO, ELENCO)	Booleano	Restituisce vero se <i>ITEM</i> è membro dell'elenco <i>LIST</i> specificato. In caso contrario, verrà restituito un valore falso. È inoltre possibile specificare un elenco di nomi di campo.
min(ELEMENTO1, ELEMENTO2)	Qualsiasi	Restituisce il minore dei due elementi, <i>ITEM1</i> o <i>ITEM2</i> .

Tabella 18. Funzioni di confronto CLEM (Continua)

Funzione	Risultato	Descrizione
min_index(LIST)	Numero intero	Restituisce l'indice del campo contenente il valore minimo da un elenco di campi numerici oppure 0 se tutti i valori sono null. Per esempio, se il terzo campo elencato contiene il valore minimo, viene restituito il valore di indice 3. Se più campi contengono il valore minimo, viene restituito quello elencato per primo (più a sinistra).
min_n(LIST)	Numero	Restituisce il valore minimo da un elenco di campi numerici oppure null se tutti i valori di campo sono null.
time_before(ORA1, ORA2)	Booleano	Viene utilizzata per controllare l'ordine dei valori di ora. Restituisce un valore vero se <i>TIME1</i> precede <i>TIME2</i> .
value_at(INT, LIST)		Restituisce il valore di ogni campo elencato all'offset INT oppure NULL se l'offset non è compreso nell'intervallo di valori validi, cioè minore di 1 o maggiore del numero dei campi elencati. Sono supportati tutti i tipi di archiviazione.

Funzioni logiche

Le espressioni CLEM possono essere utilizzate per eseguire operazioni logiche.

Tabella 19. Funzioni logiche CLEM

Funzione	Risultato	Descrizione
COND1 and COND2	Booleana	Questa operazione rappresenta una congiunzione logica e restituisce un valore vero se <i>COND1</i> and <i>COND2</i> sono entrambe vere. Se <i>COND1</i> è falsa, la valutazione di <i>COND2</i> non verrà eseguita. Si può quindi disporre di congiunzioni in cui <i>COND1</i> verifica innanzitutto la validità di un'operazione in <i>COND2</i> . Ad esempio, <code>length(Label) >=6 and Label(6) = 'x'</code> .
COND1 or COND2	Booleana	Questa operazione rappresenta una disgiunzione logica inclusiva e restituisce un valore vero se <i>COND1</i> o <i>COND2</i> risulta vera oppure se entrambe vengono rilevate come vere. Se <i>COND1</i> è vera, la valutazione di <i>COND2</i> non verrà eseguita.
not(COND)	Booleana	Questa operazione rappresenta una negazione logica e restituisce un valore vero se <i>COND</i> è falsa. In caso contrario, questa operazione restituirà un valore 0.
if COND then EXPR1 else EXPR2 endif	Qualsiasi	Questa operazione rappresenta una valutazione condizionale. Se <i>COND</i> è vera, questa operazione restituirà il risultato di <i>EXPR1</i> . Altrimenti, verrà restituito il risultato della valutazione di <i>EXPR_N</i> .
if COND1 then EXPR1 elseif COND2 then EXPR2 else EXPR_N endif	Qualsiasi	Questa operazione rappresenta una valutazione condizionale con più diramazioni. Se <i>COND1</i> è vera, questa operazione restituisce il risultato di <i>EXPR1</i> . In caso contrario, se <i>COND2</i> è vera, restituirà il risultato della valutazione di <i>EXPR2</i> . Altrimenti, verrà restituito il risultato della valutazione di <i>EXPR_N</i> .

Funzioni numeriche

CLEM include alcune funzioni numeriche di uso comune.

Tabella 20. Funzioni numeriche CLEM.

Funzione	Risultato	Descrizione
-NUM	Numero	Viene utilizzata per negare NUM. Restituisce il numero corrispondente con segno opposto.
NUM1 + NUM2	Numero	Restituisce la somma di NUM1 e NUM2.
NUM1 - NUM2	Numero	Restituisce il risultato della sottrazione del valore di NUM2 da NUM1.
NUM1 * NUM2	Numero	Restituisce il valore di NUM1 moltiplicato per NUM2.
NUM1 / NUM2	Numero	Restituisce il resto della divisione di NUM1 diviso per NUM2.
INT1 div INT2	Numero	Viene utilizzata per eseguire la divisione di numeri interi. Restituisce il valore di INT1 diviso per INT2.
INT1 rem INT2	Numero	Restituisce il resto della divisione di INT1 diviso per INT2. Ad esempio, INT1 - (INT1 div INT2) * INT2.
INT1 mod INT2	Numero	Questa funzione è obsoleta. In alternativa utilizzare la funzione rem.
BASE ** POWER	Numero	Restituisce il valore BASE elevato alla potenza POTENZA. Entrambi i valori possono essere rappresentati da qualsiasi numero, ma BASE non deve essere uguale a zero se POTENZA corrisponde a un valore zero di tipo diverso dal numero intero 0. Se POWER rappresenta un numero intero, il calcolo viene eseguito mediante moltiplicazioni successive delle potenze di BASE. Di conseguenza, se BASE è un numero intero, il risultato sarà rappresentato da un numero intero. Se POTENZA corrisponde al numero intero 0, il risultato sarà sempre rappresentato da un 1 dello stesso tipo di BASE. In caso contrario, ovvero se POWER non corrisponde a un numero intero, il risultato verrà calcolato come $\exp(\text{POWER} * \log(\text{BASE}))$.
abs(NUM)	Numero	Restituisce il valore assoluto di NUM, che corrisponde sempre a un numero dello stesso tipo.
exp(NUM)	Numero reale	Restituisce e elevato alla potenza NUM, dove e rappresenta la base dei logaritmi naturali.
fracof(NUM)	Numero reale	Restituisce la parte frazionaria di NUM, definita come NUM-intof(NUM).
intof(NUM)	Numero intero	Tronca l'argomento in modo da ottenere un numero intero. Restituisce il numero intero dello stesso segno di NUM e con l'ampiezza maggiore per cui abs(INT) <= abs(NUM).
log(NUM)	Numero reale	Restituisce il logaritmo naturale (base e) di NUM, che non deve essere uno zero di alcun tipo.
log10(NUM)	Numero reale	Restituisce il logaritmo in base 10 di NUM, che non deve essere uno zero di alcun tipo. Questa funzione è definita come $\log(\text{NUM}) / \log(10)$.
negate(NUM)	Numero	Viene utilizzata per negare NUM. Restituisce il numero corrispondente con segno opposto.
round(NUM)	Numero intero	Utilizzata per arrotondare NUM a un intero applicando intof(NUM+0.5) se NUM è positivo o intof(NUM-0.5) se NUM è negativo.

Tabella 20. Funzioni numeriche CLEM (Continua).

Funzione	Risultato	Descrizione
sign(NUM)	Numero	Viene utilizzata per determinare il segno di NUM. Se NUM è un numero intero, questa operazione restituirà -1, 0 o 1. Se NUM è un numero reale, restituirà -1,0, 0,0, o 1,0, a seconda che NUM sia negativo, uguale a zero o positivo.
sqrt(NUM)	Numero reale	Restituisce la radice quadrata di NUM. NUM deve essere positivo.
sum_n(LIST)	Numero	Restituisce la somma di valori da un elenco di campi numerici oppure null se tutti i valori di campo sono null.
mean_n(LIST)	Numero	Restituisce il valore medio da un elenco di campi numerici oppure null se tutti i valori di campo sono null.
sdev_n(LIST)	Numero	Restituisce la deviazione standard da un elenco di campi numerici oppure null se tutti i valori di campo sono null.

Funzioni trigonometriche

Tutte le funzioni illustrate in questa sezione accettano un angolo come argomento oppure ne restituiscono uno come risultato. In entrambi i casi, le unità dell'angolo, rappresentate da radianti o gradi, vengono controllate mediante l'impostazione dell'opzione corrispondente del flusso.

Tabella 21. Funzioni trigonometriche CLEM

Funzione	Risultato	Descrizione
arccos(NUM)	Numero reale	Calcola l'arcocoseno dell'angolo specificato.
arccosh(NUM)	Numero reale	Calcola l'arcocoseno iperbolico dell'angolo specificato.
arcsin(NUM)	Numero reale	Calcola l'arcoseno dell'angolo specificato.
arcsinh(NUM)	Numero reale	Calcola l'arcoseno iperbolico dell'angolo specificato.
arctan(NUM)	Numero reale	Calcola l'arcotangente dell'angolo specificato.
arctan2(NUM_Y, NUM_X)	Numero reale	Calcola l'arcotangente di NUM_Y / NUM_X e utilizza i segni dei due numeri per derivare le informazioni sui quadranti. Il risultato è rappresentato da un numero reale compreso nell'intervallo $-\pi < \text{ANGLE} \leq \pi$ (radians) - $180 < \text{ANGLE} \leq 180$ (degrees)
arctanh(NUM)	Numero reale	Calcola l'arcotangente iperbolico dell'angolo specificato.
cos(NUM)	Numero reale	Calcola il coseno dell'angolo specificato.
cosh(NUM)	Numero reale	Calcola il coseno iperbolico dell'angolo specificato.
pi	Numero reale	Questa costante è la migliore approssimazione reale di pi.
sin(NUM)	Numero reale	Calcola il seno dell'angolo specificato.
sinh(NUM)	Numero reale	Calcola il seno iperbolico dell'angolo specificato.
tan(NUM)	Numero reale	Calcola la tangente dell'angolo specificato.
tanh(NUM)	Numero reale	Calcola la tangente iperbolica dell'angolo specificato.

Funzioni di probabilità

Le funzioni di probabilità restituiscono probabilità in base a varie distribuzioni, quali la probabilità che un valore dalla distribuzione *t* di Student sia minore di un determinato valore.

Tabella 22. Funzioni di probabilità CLEM

Funzione	Risultato	Descrizione
<code>cdf_chisq(NUM, DF)</code>	Numero reale	Restituisce la probabilità che un valore dalla distribuzione chi-quadrato con i gradi di libertà specificati sia inferiore al numero specificato.
<code>cdf_f(NUM, DF1, DF2)</code>	Numero reale	Restituisce la probabilità che un valore dalla distribuzione <i>F</i> , con gradi di libertà <i>DF1</i> e <i>DF2</i> , sia inferiore al numero specificato.
<code>cdf_normal(NUM, MEAN, STDDEV)</code>	Numero reale	Restituisce la probabilità che un valore dalla distribuzione normale con la deviazione standard e media specificate sia inferiore al numero specificato.
<code>cdf_t(NUM, DF)</code>	Numero reale	Restituisce la probabilità che un valore dalla distribuzione <i>t</i> per studenti con i gradi di libertà specificati sia inferiore al numero specificato.

Operazioni bitwise sui numeri interi

Queste funzioni consentono la manipolazione dei numeri interi come schemi di bit rappresentanti valori in complemento a due, in cui alla posizione di bit *N* corrisponde il peso 2^{*N} . I bit sono numerati da 0 a salire. Queste operazioni vengono eseguite come se il bit del segno di un numero intero venisse esteso a sinistra in modo indefinito. Di conseguenza, in qualsiasi posizione precedente al bit più significativo, un numero intero positivo dispone di 0 bit e un numero intero negativo di 1 bit.

Tabella 23. Operazioni sui numeri interi bitwise CLEM.

Funzione	Risultato	Descrizione
<code>~~ INT1</code>	Numero intero	Genera il complemento bitwise del numero intero <i>INT1</i> . Nel risultato è quindi presente un 1 per ciascuna posizione di bit in cui si riscontra 0 in <i>INT1</i> . È sempre valida la regola $~~INT = -(INT + 1)$.
<code>INT1 INT2</code>	Numero intero	Il risultato di questa operazione è costituito dall'OR inclusivo bitwise di <i>INT1</i> e <i>INT2</i> . Nel risultato è quindi presente un 1 per ogni posizione di bit in cui si riscontra un 1 in <i>INT1</i> o <i>INT2</i> oppure in entrambi.
<code>INT1 /& INT2</code>	Numero intero	Il risultato di questa operazione è costituito dall'OR esclusivo bitwise di <i>INT1</i> e <i>INT2</i> . Nel risultato è quindi presente un 1 per ogni posizione di bit in cui si riscontra un 1 in <i>INT1</i> o in <i>INT2</i> , ma non in entrambi.
<code>INT1 && INT2</code>	Numero intero	Genera l'AND bitwise dei numeri interi <i>INT1</i> e <i>INT2</i> . Nel risultato è quindi presente un 1 per ciascuna posizione di bit in cui si riscontra un 1 sia in <i>INT1</i> che in <i>INT2</i> .
<code>INT1 &&~~ INT2</code>	Numero intero	Genera l'AND bitwise di <i>INT1</i> e il complemento bitwise di <i>INT2</i> . Nel risultato è quindi presente un 1 per ogni posizione di bit in cui si riscontra un 1 in <i>INT1</i> e uno 0 in <i>INT2</i> . Equivale a <code>INT1 && (~~INT2)</code> e si rivela utile per cancellare i bit di <i>INT1</i> impostati in <i>INT2</i> .
<code>INT << N</code>	Numero intero	Genera lo schema di bit di <i>INT1</i> con spostamento a sinistra di <i>N</i> posizioni. L'utilizzo di un valore negativo per <i>N</i> determina uno spostamento a destra.
<code>INT >> N</code>	Numero intero	Genera lo schema di bit di <i>INT1</i> con spostamento a destra di <i>N</i> posizioni. L'utilizzo di un valore negativo per <i>N</i> determina uno spostamento a sinistra.
<code>INT1 &&=_0 INT2</code>	Booleano	Equivale all'espressione booleana <code>INT1 && INT2 /= 0</code> , ma offre una maggiore efficienza.

Tabella 23. Operazioni sui numeri interi bitwise CLEM (Continua).

Funzione	Risultato	Descrizione
<code>INT1 &&/=_0 INT2</code>	Booleano	Equivale all'espressione booleana <code>INT1 && INT2 == 0</code> , ma offre una maggiore efficienza.
<code>integer_bitcount(INT)</code>	Numero intero	Conta il numero di bit 1 o 0 nella rappresentazione in complemento a due di <i>INT</i> . Se <i>INT</i> non è negativo, <i>N</i> corrisponderà al numero di bit 1. Se <i>INT</i> è negativo, rappresenterà il numero di bit 0. A causa dell'estensione del segno, in un numero intero non negativo è presente un numero infinito di bit 0 e un numero intero negativo contiene un numero infinito di bit 1. È sempre valida la regola <code>integer_bitcount(INT) = integer_bitcount(-(INT+1))</code> .
<code>integer_leastbit(INT)</code>	Numero intero	Restituisce la posizione di bit <i>N</i> dell'insieme di bit meno significativo nel numero intero <i>INT</i> . <i>N</i> rappresenta la più elevata potenza di 2 per cui <i>INT</i> può essere diviso esattamente.
<code>integer_length(INT)</code>	Numero intero	Restituisce la lunghezza in bit di <i>INT</i> come intero in complemento a due. <i>N</i> rappresenta quindi il numero intero più basso per cui <code>INT < (1 << N) if INT >= 0 INT >= (-1 << N) if INT < 0</code> . If <i>INT</i> non è negativo, quindi la rappresentazione di <i>INT</i> come numero intero senza segno richiede un campo di almeno <i>N</i> bit. In alternativa, per rappresentare <i>INT</i> come numero intero con segno, sarà richiesto un minimo di <i>N</i> +1 bit, indipendentemente dal segno.
<code>testbit(INT, N)</code>	Booleano	Verifica il bit in posizione <i>N</i> nel numero intero <i>INT</i> e restituisce lo stato del bit <i>N</i> sotto forma di valore booleano, ovvero vero per 1 e falso per 0.

Funzioni casuali

Le funzioni illustrate di seguito vengono utilizzate per la selezione casuale di elementi o la generazione casuale di numeri.

Tabella 24. Funzioni casuali CLEM.

Funzione	Risultato	Descrizione
<code>oneof(LIST)</code>	Qualsiasi	Restituisce un elemento di <i>LIST</i> scelto in modo casuale. Gli elementi dell'elenco devono essere specificati nel formato <code>[ITEM1, ITEM2, ..., ITEM_N]</code> . Si noti che è inoltre possibile specificare un elenco di nomi di campo.
<code>random(NUM)</code>	Numero	Restituisce un numero random distribuito in modo uniforme dello stesso tipo (<i>INT</i> o <i>REAL</i>), da 1 a <i>NUM</i> . Se si utilizza un numero intero, verranno restituiti solo numeri interi. Se si utilizza un numero reale (decimale), verranno restituiti numeri reali la cui precisione decimale è determinata dalle opzioni del flusso. Il numero random più grande restituito dalla funzione può essere uguale a <i>NUM</i> .
<code>random0(NUM)</code>	Numero	Ha le stesse proprietà di <code>random(NUM)</code> , ma a partire da 0. Il numero random più grande restituito dalla funzione non sarà mai uguale a <i>NUM</i> .

Funzioni stringa

Nel linguaggio CLEM sono supportate le seguenti operazioni con le stringhe:

- Confronto di stringhe

- Creazione di stringhe
- Accesso ai caratteri

In un'espressione CLEM, una stringa è rappresentata da qualsiasi sequenza di caratteri racchiusa tra virgolette corrispondenti ("virgolette stringa"). I caratteri (CHAR) possono essere rappresentati da qualsiasi carattere alfanumerico singolo. Nelle espressioni CLEM vengono dichiarati utilizzando virgolette singole nel formato ``<carattere>``, per esempio ``z``, ``A`` o ``2``. I caratteri che rappresentano indici fuori intervallo o negativi in una stringa determineranno un comportamento non definito.

Nota: i confronti tra stringhe che utilizzano e non utilizzano il push back SQL potrebbero generare risultati diversi in presenza di spazi finali.

Tabella 25. Funzioni stringa di CLEM.

Funzione	Risultato	Descrizione
<code>allbutfirst(N, STRING)</code>	Stringa	Restituisce una stringa risultante dalla rimozione dei primi <i>N</i> caratteri da <i>STRING</i> .
<code>allbutlast(N, STRING)</code>	Stringa	Restituisce una stringa risultante dalla rimozione degli ultimi caratteri da <i>STRING</i> .
<code>alphabefore(STRING1, STRING2)</code>	Booleano	Viene utilizzata per controllare l'ordine alfabetico delle stringhe. Restituisce un valore vero se <i>STRING1</i> precede <i>STRING2</i> .
<code>endstring(LENGTH, STRING)</code>	Stringa	Estrae gli ultimi <i>N</i> caratteri dalla stringa specificata. Se la lunghezza della stringa è inferiore o uguale alla lunghezza specificata, rimane invariata.
<code>hasendstring(STRING, SUBSTRING)</code>	Numero intero	Questa funzione è equivalente a <code>isendstring(SUBSTRING, STRING)</code> .
<code>hasmidstring(STRING, SUBSTRING)</code>	Numero intero	Questa funzione è equivalente a <code>ismidstring(SUBSTRING, STRING)</code> , in relazione alla sottostringa incorporata.
<code>hasstartstring(STRING, SUBSTRING)</code>	Numero intero	Questa funzione è equivalente a <code>isstartstring(SUBSTRING, STRING)</code> .
<code>hassubstring(STRING, N, SUBSTRING)</code>	Numero intero	Questa funzione è equivalente a <code>issubstring(SUBSTRING, N, STRING)</code> , con <i>N</i> impostato per default su 1.
<code>count_substring(STRING, SUBSTRING)</code>	Numero intero	Restituisce il numero di volte in cui la sottostringa specificata è presente all'interno della stringa. Per esempio, <code>count_substring("foooo.txt", "oo")</code> restituisce 3.
<code>hassubstring(STRING, SUBSTRING)</code>	Numero intero	Questa funzione è equivalente a <code>issubstring(SUBSTRING, 1, STRING)</code> , con <i>N</i> impostato per default su 1.
<code>isalphacode(CHAR)</code>	Booleano	Restituisce un valore vero se <i>CHAR</i> rappresenta un carattere della stringa specificata (spesso un nome di campo) con codice alfabetico. In caso contrario, questa funzione restituirà un valore 0. Per esempio, <code>isalphacode(num_prodotto(1))</code> .
<code>isendstring(SUBSTRING, STRING)</code>	Numero intero	Se la stringa <i>STRING</i> termina con la sottostringa <i>SUBSTRING</i> , questa funzione restituirà il numero di indice di <i>SUBSTRING</i> in <i>STRING</i> . In caso contrario, questa funzione restituirà un valore 0.

Tabella 25. Funzioni stringa di CLEM (Continua).

Funzione	Risultato	Descrizione
islowercode(CHAR)	Booleano	Restituisce il valore vero se <i>CHAR</i> rappresenta una lettera minuscola per la stringa specificata (spesso un nome di campo). In caso contrario, restituirà il valore 0. Per esempio, sia <code>islowercode(``)</code> che <code>islowercode(nome_paese((2)))</code> sono espressioni valide.
ismidstring(SUBSTRING, STRING)	Numero intero	Se <i>SUBSTRING</i> è una sottostringa di <i>STRING</i> ma non ha inizio con il primo carattere di <i>STRING</i> né termina con l'ultimo, questa funzione restituirà l'indice in corrispondenza del quale inizia la sottostringa. In caso contrario, questa funzione restituirà un valore 0.
isnumbercode(CHAR)	Booleano	Restituisce il valore vero se <i>CHAR</i> per la stringa specificata (spesso un nome di campo) rappresenta un carattere con codice numerico. In caso contrario, restituirà il valore 0. Per esempio, <code>isnumbercode(id_prodotto(2))</code> .
isstartstring(SUBSTRING, STRING)	Numero intero	Se la stringa <i>STRING</i> ha inizio con la sottostringa <i>SUBSTRING</i> , questa funzione restituirà l'indice 1. In caso contrario, restituirà il valore 0.
issubstring(SUBSTRING, N, STRING)	Numero intero	Cerca una sottostringa uguale a <i>SUBSTRING</i> all'interno della stringa <i>STRING</i> , a partire dal carattere <i>N</i> . Se la ricerca ha esito positivo, questa funzione restituisce il valore di indice numero intero in corrispondenza del quale ha inizio la sottostringa corrispondente. In caso contrario, restituirà un valore 0. Se non viene specificato, <i>N</i> verrà impostato per default su 1.
issubstring(SUBSTRING, STRING)	Numero intero	Cerca una sottostringa uguale a <i>SUBSTRING</i> all'interno della stringa <i>STRING</i> , a partire dal carattere <i>N</i> . Se la ricerca ha esito positivo, questa funzione restituisce il valore di indice numero intero in corrispondenza del quale ha inizio la sottostringa corrispondente. In caso contrario, restituirà un valore 0. Se non viene specificato, <i>N</i> verrà impostato per default su 1.
issubstring_count(SUBSTRING, N, STRING):	Numero intero	Restituisce l'indice dell' <i>N</i> esima occorrenza della <i>SUBSTRING</i> all'interno della <i>STRING</i> specificata. Se le occorrenze della <i>SUBSTRING</i> sono inferiori a <i>N</i> , viene restituito 0.
issubstring_lim(SUBSTRING, N, STARTLIM, ENDLIM, STRING)	Numero intero	Questa funzione è equivalente a <code>issubstring</code> , ma la corrispondenza deve iniziare prima dell'indice <i>STARTLIM</i> o a partire da esso e terminare con l'indice <i>ENDLIM</i> o prima di esso. I vincoli <i>STARTLIM</i> o <i>ENDLIM</i> possono essere disattivati fornendo un valore falso per questi argomenti, per esempio <code>issubstring_lim(SSUBSTRING, N, false, false, STRING)</code> equivale a <code>issubstring</code> .
isuppercode(CHAR)	Booleano	Restituisce un valore vero se <i>CHAR</i> rappresenta una lettera maiuscola. In caso contrario, restituirà il valore 0. Per esempio, sia <code>isuppercode(``)</code> che <code>isuppercode(nome_paese((2)))</code> sono espressioni valide.

Tabella 25. Funzioni stringa di CLEM (Continua).

Funzione	Risultato	Descrizione
last(CHAR)	Stringa	Restituisce l'ultimo carattere CHAR di STRING (che deve contenere almeno un carattere).
length(STRING)	Numero intero	Restituisce la lunghezza della stringa STRING, ovvero il numero di caratteri in essa contenuti.
locchar(CHAR, N, STRING)	Numero intero	Utilizzata per identificare la posizione dei caratteri nei campi simbolici. La funzione cerca il carattere CHAR all'interno della stringa STRING, iniziando la ricerca dal carattere Nesimo di STRING. Questa funzione restituisce un valore che indica la posizione (a partire da N) in cui si trova il carattere. Se il carattere non viene trovato, questa funzione restituirà un valore 0. Se la funzione include un offset (N) non valido (per esempio un offset superiore alla lunghezza della stringa), la funzione restituisce \$null\$. Per esempio, locchar(`n`, 2, pagina_web) cerca il carattere `n` all'interno del campo denominato pagina_web, a partire dal secondo carattere nel valore del campo. <i>Nota:</i> racchiudere il carattere specificato tra virgolette singole.
locchar_back(CHAR, N, STRING)	Numero intero	Funzione analoga a locchar, tranne per il fatto che la ricerca viene eseguita all'indietro, a partire dall'Nesimo carattere. Per esempio, locchar_back(`n`, 9, pagina_web) esegue la ricerca all'interno del campo pagina_web a partire dal nono carattere e procedendo all'indietro verso l'inizio della stringa. Se contiene un offset non valido, per esempio superiore alla lunghezza della stringa, questa funzione restituirà \$null\$. In teoria, per utilizzare in modo dinamico la lunghezza del valore corrente del campo sarebbe opportuno utilizzare locchar_back insieme alla funzione length(<campo>). Per esempio, locchar_back(`n`, (length(pagina_web)), pagina_web).
lowertoupper(CHAR) lowertoupper (STRING)	Carattere o Stringa	L'input può essere rappresentato da una stringa o da un carattere, utilizzato in questa funzione per restituire un nuovo elemento dello stesso tipo, con i caratteri minuscoli convertiti nei relativi maiuscoli. Ad esempio, lowertoupper(`a`), lowertoupper("My string") e lowertoupper(nome_campo(2)) sono espressioni valide.

Tabella 25. Funzioni stringa di CLEM (Continua).

Funzione	Risultato	Descrizione
matches	Booleano	Restituisce vero se una stringa corrisponde allo schema specificato. Lo schema deve essere una stringa letterale e non può essere un nome di campo contenente uno schema. È possibile includere nello schema un punto interrogativo (?) corrispondente esattamente a un carattere; un asterisco (*) corrisponde a zero o più caratteri. Per indicare un punto interrogativo o un asterisco letterale, ovvero non utilizzato come carattere jolly, è possibile utilizzare la barra rovesciata (\).
replace(SUBSTRING, NEWSUBSTRING, STRING)	Stringa	All'interno della <i>STRING</i> specificata, sostituisce tutte le istanze della <i>SUBSTRING</i> con la <i>NEWSUBSTRING</i> .
replicate(COUNT, STRING)	Stringa	Restituisce una stringa costituita dalla stringa originale copiata il numero di volte specificato.
stripchar(Char, STRING)	Stringa	Consente di rimuovere i caratteri specificati da una stringa o campo. È possibile utilizzare questa funzione, per esempio, per rimuovere dai dati i simboli aggiuntivi, come i simboli di valuta, in modo da ottenere un nome o un numero semplice. Per esempio, la sintassi stripchar('\$', 'Costo') restituisce un nuovo campo in cui il simbolo del dollaro è stato rimosso da tutti i valori. <i>Nota:</i> racchiudere il carattere specificato tra virgolette singole.
skipchar(Char, N, STRING)	Numero intero	Cerca nella stringa <i>STRING</i> qualsiasi carattere diverso da <i>CHAR</i> , a partire dall' <i>Nesimo</i> carattere. Questa funzione restituisce una sottostringa di numeri interi che indica il punto in cui è stato trovato un carattere oppure 0, se ogni carattere successivo a <i>Nesimo</i> corrisponde a <i>CHAR</i> . Se contiene un offset non valido, per esempio superiore alla lunghezza della stringa, questa funzione restituirà \$null\$. locchar è spesso utilizzato insieme alle funzioni skipchar per determinare il valore di <i>N</i> (punto dal quale iniziare la ricerca nella stringa). Per esempio, skipchar(`s`, (locchar(`s`, 1, "MiaStringa))), "MiaStringa").
skipchar_back(Char, N, STRING)	Numero intero	Questa funzione è analoga a skipchar, tranne che per il fatto che la ricerca viene eseguita a ritroso , a partire dall' <i>Nesimo</i> carattere.
startstring(LENGTH, STRING)	Stringa	Estrae i primi <i>N</i> caratteri dalla stringa specificata. Se la lunghezza della stringa è inferiore o uguale alla lunghezza specificata, rimane invariata.
strmember(Char, STRING)	Numero intero	Equivale a locchar(Char, 1, STRING). Restituisce una sottostringa di numeri interi che indica il punto in cui viene rilevata la prima occorrenza di <i>CHAR</i> oppure 0. Se contiene un offset non valido, per esempio superiore alla lunghezza della stringa, questa funzione restituirà \$null\$.

Tabella 25. Funzioni stringa di CLEM (Continua).

Funzione	Risultato	Descrizione
subscrs(N, STRING)	Carattere	Restituisce l'Nesimo carattere CHAR della stringa di input STRING. Per questa funzione è inoltre disponibile la forma abbreviata STRING(N). Ad esempio, lowertoupper("name" (1)) è un'espressione valida.
substring(N, LEN, STRING)	Stringa	Restituisce una stringa SUBSTRING, rappresentata dal numero di caratteri di STRING indicati da LUNGH, a partire dal carattere in corrispondenza dell'indice N.
substring_between(N1, N2, STRING)	Stringa	Restituisce la sottostringa di STRING che ha inizio in corrispondenza dell'indice N1 e termina con l'indice N2.
trim(STRING)	Stringa	Rimuove i caratteri spazio vuoto iniziale e finale dalla stringa specificata.
trim_start(STRING)	Stringa	Rimuove i caratteri spazio vuoto iniziale dalla stringa specificata.
trimend(STRING)	Stringa	Rimuove i caratteri spazio vuoto finale dalla stringa specificata.
unicode_char(NUM)	Carattere	Restituisce il carattere con valore Unicode NUM
unicode_value(CHAR)	NUM	Restituisce il valore Unicode di CAR
uppertolower(CHAR) uppertolower (STRING)	Carattere o Stringa	L'input può essere rappresentato da una stringa o da un carattere, utilizzato in questa funzione per restituire un nuovo elemento dello stesso tipo, con i caratteri maiuscoli convertiti nei relativi minuscoli. <i>Nota:</i> ricordarsi di specificare le stringhe con le virgolette e i caratteri con le virgolette singole. I nomi semplici di campo non devono essere racchiusi tra virgolette.

Funzioni SoundEx

SoundEx è un metodo utilizzato per trovare stringhe quando si conosce un suono ma non l'esatta ortografia. Sviluppato nel 1918, cerca parole con suoni simili in base a ipotesi fonetiche sulla pronuncia di determinate lettere. Può essere per esempio utilizzato per cercare nomi in un database, in cui ortografia e pronuncia di nomi simili possono variare. L'algoritmo SoundEx di base è documentato in numerose fonti e nonostante i limiti conosciuti (per esempio, la combinazione iniziale ph e la lettera f non corrispondono nonostante il loro suono sia simile) è supportato in qualche forma dalla maggioranza dei database.

Tabella 26. Funzioni soundex CLEM

Funzione	Risultato	Descrizione
soundex(STRING)	Numero intero	Restituisce il codice SoundEx di quattro caratteri per la STRING specificata.
soundex_difference(STRING1, STRING2)	Numero intero	Restituisce un numero intero compreso tra 0 e 4 che indica il numero di caratteri che sono identici nella codifica SoundEx per le due stringhe, dove 0 indica nessuna somiglianza e 4 indica una forte somiglianza o stringhe identiche.

Funzioni di data e ora

Nel linguaggio CLEM è disponibile un gruppo di funzioni per la gestione dei campi con archiviazione di tipo datetime delle variabili che rappresentano le date e le ore. I formati di data e ora utilizzati sono specifici di ogni flusso e vengono impostati nella finestra di dialogo delle proprietà del flusso. Le funzioni di data e ora analizzano le stringhe di data e ora in base al formato correntemente selezionato.

Quando si specifica un anno in una data che utilizza solo due cifre, ovvero in cui non è indicato il secolo, IBM SPSS Modeler utilizza il secolo di default specificato nella finestra di dialogo delle proprietà del flusso.

Tabella 27. Funzioni data e ora CLEM.

Funzione	Risultato	Descrizione
@TODAY	Stringa	Se si seleziona Cambia giorni/min nella finestra di dialogo delle proprietà del flusso, questa funzione restituirà la data corrente sotto forma di stringa nel formato di data corrente. Se si utilizza un formato di data a due cifre e non si seleziona Cambia giorni/min , questa funzione restituirà \$null\$ sul server corrente.
to_time(ELEMENTO)	Ora	Converte il tipo di archiviazione del campo specificato in un'ora.
to_date(ITEM)	Data	Converte il tipo di archiviazione del campo specificato in una data.
to_timestamp(ITEM)	Timestamp	Converte il tipo di archiviazione del campo specificato in un timestamp.
to_datetime(ITEM)	Data/ora	Converte il tipo di archiviazione del campo specificato in un valore di data, ora o timestamp.
datetime_date(ITEM)	Data	Restituisce il valore di data per un <i>numero</i> , una <i>stringa</i> o un <i>timestamp</i> . Si noti che questa è la sola funzione che consente di convertire un numero (in secondi) nuovamente in una data. Se ITEM è una stringa, crea una data analizzando una stringa nel formato di data corrente. La funzione verrà eseguita correttamente solo se il formato di data specificato nella finestra di dialogo Proprietà flusso è corretto. Se ITEM è un numero, viene interpretato come numero di secondi dalla data (o epoca) di base. Le frazioni di giorno vengono troncate. Se ITEM è un timestamp, viene restituita la parte relativa alla data del timestamp. Se ITEM è una data, la data viene restituita invariata.
date_before(DATE1, DATE2)	Booleano	Restituisce un valore vero se DATE1 rappresenta una data o un timestamp precedente a quella rappresentata da DATE2. In caso contrario, questa funzione restituirà un valore 0.
date_days_difference(DATE1, DATE2)	Numero intero	Restituisce il tempo in giorni dalla data o dal timestamp rappresentato da DATE1 a quello rappresentato da DATE2, sotto forma di numero intero. Se DATE2 è antecedente a DATE1, questa funzione restituisce un numero negativo.
date_in_days(DATE)	Numero intero	Restituisce il tempo in giorni trascorso dalla data di riferimento alla data o al timestamp rappresentato da DATE, sotto forma di numero intero. Se il valore di DATE è precedente alla data di riferimento, questa funzione restituisce un numero negativo. Per ottenere un calcolo corretto è necessario includere una data valida. Per esempio, 29 febbraio 2001 non è una data valida. Infatti, dal momento che il 2001 non è un anno bisestile, questa data non esiste.

Tabella 27. Funzioni data e ora CLEM (Continua).

Funzione	Risultato	Descrizione
date_in_months(<i>DATE</i>)	Numero reale	Restituisce il tempo in mesi trascorso dalla data di riferimento alla data o al timestamp rappresentato da <i>DATE</i> , sotto forma di numero reale. Tale numero corrisponde a una cifra approssimativa basata su un mese formato da 30.4375 giorni. Se il valore di <i>DATE</i> è precedente alla data di riferimento, questa funzione restituisce un numero negativo. Per ottenere un calcolo corretto è necessario includere una data valida. Per esempio, 29 febbraio 2001 non è una data valida. Infatti, dal momento che il 2001 non è un anno bisestile, questa data non esiste.
date_in_weeks(<i>DATE</i>)	Numero reale	Restituisce il tempo in settimane trascorso dalla data di riferimento alla data o al timestamp rappresentato da <i>DATE</i> , sotto forma di numero reale. Tale numero è basato su una settimana formata da sette giorni. Se il valore di <i>DATE</i> è precedente alla data di riferimento, questa funzione restituisce un numero negativo. Per ottenere un calcolo corretto è necessario includere una data valida. Per esempio, 29 febbraio 2001 non è una data valida. Infatti, dal momento che il 2001 non è un anno bisestile, questa data non esiste.
date_in_years(<i>DATE</i>)	Numero reale	Restituisce il tempo in anni trascorso dalla data di riferimento alla data o al timestamp rappresentato da <i>DATE</i> , sotto forma di numero reale. Tale numero corrisponde a una cifra approssimativa basata su un anno formato da 365.25 giorni. Se il valore di <i>DATE</i> è precedente alla data di riferimento, questa funzione restituisce un numero negativo. Per ottenere un calcolo corretto è necessario includere una data valida. Per esempio, 29 febbraio 2001 non è una data valida. Infatti, dal momento che il 2001 non è un anno bisestile, questa data non esiste.
date_months_difference (<i>DATE1</i> , <i>DATE2</i>)	Numero reale	Restituisce il tempo in mesi dalla data o dal timestamp rappresentato da <i>DATE1</i> a quello rappresentato da <i>DATE2</i> , sotto forma di numero reale. Tale numero corrisponde a una cifra approssimativa basata su un mese formato da 30.4375 giorni. Se <i>DATE2</i> è antecedente a <i>DATE1</i> , questa funzione restituisce un numero negativo.
datetime_date(<i>YEAR</i> , <i>MONTH</i> , <i>DAY</i>)	Data	Crea il valore di data per gli argomenti <i>YEAR</i> , <i>MONTH</i> e <i>DAY</i> specificati. Gli argomenti devono essere numeri interi.
datetime_day(<i>DATE</i>)	Numero intero	Restituisce il giorno del mese di una <i>DATE</i> o un timestamp. Il risultato è un numero intero nell'intervallo da 1 a 31.
datetime_day_name(<i>DAY</i>)	Stringa	Restituisce il nome completo dell'argomento <i>DAY</i> specificato. L'argomento deve essere un numero intero nell'intervallo compreso tra 1 (domenica) e 7 (sabato).
datetime_hour(<i>TIME</i>)	Numero intero	Restituisce l'ora dal valore <i>TIME</i> o da un timestamp. Il risultato è un numero intero nell'intervallo da 0 a 23.
datetime_in_seconds(<i>TIME</i>)	Numero reale	Restituisce la parte in secondi memorizzata in <i>TIME</i> .
datetime_in_seconds(<i>DATE</i>), datetime_in_seconds(<i>DATETIME</i>)	Numero reale	Restituisce il numero accumulato, convertito in secondi, dalla differenza tra il valore <i>DATE</i> o <i>DATETIME</i> corrente e la data di riferimento (1900-01-01).
datetime_minute(<i>TIME</i>)	Numero intero	Restituisce i minuti dal valore <i>TIME</i> o da un timestamp. Il risultato è un numero intero nell'intervallo compreso tra 0 e 59.

Tabella 27. Funzioni data e ora CLEM (Continua).

Funzione	Risultato	Descrizione
datetime_month(DATE)	Numero intero	Restituisce il mese da una DATE o da un timestamp. Il risultato è un numero intero nell'intervallo da 1 a 12.
datetime_month_name (MONTH)	Stringa	Restituisce il nome completo dell'argomento MONTH specificato. L'argomento deve essere un numero intero nell'intervallo da 1 a 12.
datetime_now	Timestamp	Restituisce l'ora corrente come timestamp.
datetime_second(TIME)	Numero intero	Restituisce i secondi dal valore TIME o da un timestamp. Il risultato è un numero intero nell'intervallo da 0 a 59.
datetime_day_short_name (DAY)	Stringa	Restituisce il nome abbreviato dell'argomento DAY specificato. L'argomento deve essere un numero intero nell'intervallo compreso tra 1 (domenica) e 7 (sabato).
datetime_month_short_name (MONTH)	Stringa	Restituisce il nome abbreviato dell'argomento MONTH specificato. L'argomento deve essere un numero intero nell'intervallo da 1 a 12.
datetime_time(HOUR, MINUTE, SECOND)	Ora	Restituisce il valore di ora per gli argomenti TIME, MINUTE, e SECOND specificati. Gli argomenti devono essere numeri interi.
datetime_time(ITEM)	Ora	Restituisce il valore di ora dell'ITEM specificato.
datetime_timestamp(YEAR, MONTH, DAY, HOUR, MINUTE, SECOND)	Timestamp	Restituisce il valore di timestamp per l'YEAR, il MONTH, il DAY, l'HOUR, i MINUTE e i SECONDI specificati.
datetime_timestamp(DATE, TIME)	Timestamp	Restituisce il valore di timestamp per la DATE e l'TIME specificate.
datetime_timestamp (NUMBER)	Timestamp	Restituisce il valore di timestamp del numero di secondi specificato.
datetime_weekday(DATE)	Numero intero	Restituisce il giorno della settimana dalla DATE o dal timestamp specificato.
datetime_year(DATE)	Numero intero	Restituisce l'anno da una DATE o da un timestamp. Il risultato è un numero intero come 2002.
date_weeks_difference (DATE1, DATE2)	Numero reale	Restituisce il tempo in settimane dalla data o dal timestamp rappresentato da DATE1 a quello rappresentato da DATE2, sotto forma di numero reale. Tale numero è basato su una settimana formata da sette giorni. Se DATE2 è antecedente a DATE1, questa funzione restituisce un numero negativo.
date_years_difference (DATE1, DATE2)	Numero reale	Restituisce il tempo in anni dalla data o dal timestamp rappresentato da DATE1 a quello rappresentato da DATE2, sotto forma di numero reale. Tale numero corrisponde a una cifra approssimativa basata su un anno formato da 365.25 giorni. Se DATE2 è antecedente a DATE1, questa funzione restituisce un numero negativo.
time_before(TIME1, TIME2)	Booleano	Restituisce un valore vero se TIME1 rappresenta un'ora o un timestamp precedente a quello rappresentata da TIME2. In caso contrario, questa funzione restituirà un valore 0.
time_hours_difference (TIME1, TIME2)	Numero reale	Restituisce il tempo trascorso in ore tra le ore o timestamp rappresentati da TIME1 e TIME2, sotto forma di numero reale. Se si seleziona Cambia giorni/min nella finestra di dialogo delle proprietà del flusso, un valore superiore di TIME1 verrà interpretato come riferito al giorno precedente. Se non si seleziona l'opzione di trasformazione, un valore superiore di TIME1 determina la restituzione di un valore negativo.

Tabella 27. Funzioni data e ora CLEM (Continua).

Funzione	Risultato	Descrizione
<code>time_in_hours(TIME)</code>	Numero reale	Restituisce il tempo in ore rappresentato dall'ora <i>TIME</i> , sotto forma di numero reale. Con il formato di ora HHMM, per esempio, l'espressione <code>time_in_hours('0130')</code> restituisce 1.5. <i>TIME</i> può rappresentare un'ora o un timestamp.
<code>time_in_mins(TIME)</code>	Numero reale	Restituisce il tempo in minuti rappresentato da <i>TIME</i> , sotto forma di numero reale. <i>TIME</i> può rappresentare un'ora o un timestamp.
<code>time_in_secs(TIME)</code>	Numero intero	Restituisce il tempo in secondi rappresentato da <i>TIME</i> , sotto forma di numero intero. <i>TIME</i> può rappresentare un'ora o un timestamp.
<code>time_mins_difference(TIME1, TIME2)</code>	Numero reale	Restituisce il tempo trascorso in minuti tra le ore o timestamp rappresentati da <i>TIME1</i> e <i>TIME2</i> , sotto forma di numero reale. Se si seleziona Cambia giorni/min nella finestra di dialogo delle proprietà del flusso, un valore superiore a <i>TIME1</i> verrà interpretato come riferito al giorno precedente (o all'ora precedente, se per il formato corrente sono specificati solo ore e minuti). Se non si seleziona l'opzione di trasformazione, un valore superiore a <i>TIME1</i> determina la restituzione di un valore negativo.
<code>time_secs_difference(TIME1, TIME2)</code>	Numero intero	Restituisce la differenza in secondi tra le ore o timestamp rappresentati da <i>TIME1</i> e <i>TIME2</i> , sotto forma di numero intero. Se si seleziona Cambia giorni/min nella finestra di dialogo delle proprietà del flusso, un valore superiore a <i>TIME1</i> verrà interpretato come riferito al giorno precedente (o all'ora precedente, se per il formato corrente sono specificati solo ore e minuti). Se non si seleziona l'opzione di trasformazione, un valore superiore di <i>TIME1</i> determina la restituzione di un valore negativo.

Conversione dei valori di data e ora

Si noti che le funzioni di conversione, e qualsiasi altra funzione che richieda un tipo di input specifico quale un valore di data o ora, dipendono dai formati correnti specificati nella finestra di dialogo Opzioni flusso. Per esempio, un campo denominato *DATE*, archiviato come stringa con i valori *Gen 2003*, *Feb 2003* e così via, potrebbe essere convertito in campo di archiviazione di date nel modo seguente:

```
to_date(DATE)
```

Perché la conversione funzioni è necessario selezionare il formato di data **MES AAAA** corrispondente come default per il flusso.

Per un esempio di conversione di valori stringa in date mediante un nodo Riempimento, vedere il flusso *broadband_create_models.str*, installato nella cartella *\Demos* all'interno della sottocartella *streams*.

Date archiviate come numeri. Tenere presente che *DATE* nell'esempio precedente è il nome di un campo, mentre `to_date` è una funzione CLEM. È possibile convertire le date archiviate come numeri utilizzando la funzione `datetime_date`, in cui il numero viene interpretato come numero di secondi trascorsi dalla data (o epoca) di base.

```
datetime_date(DATE)
```

Mediante la conversione di una data in un numero di secondi e viceversa, è possibile effettuare calcoli, quali la data corrente più o meno un numero determinato di giorni. Per esempio:

```
datetime_date((date_in_days(DATE)-7)*60*60*24)
```

Funzioni sequenza

Per alcune operazioni, la sequenza degli eventi si rivela importante. L'applicazione supporta le seguenti sequenze di record:

- Sequenze e serie temporali
- Funzioni sequenza
- Indicizzazione di record
- Calcolo della media, somma e confronto di valori
- Monitoraggio delle modifiche e differenziazione
- @SINCE
- Valori di offset
- Funzionalità aggiuntive relative alle sequenze

In numerose applicazioni, ogni record in un flusso può essere considerato come un caso singolo, indipendente da tutti gli altri. In queste circostanze, l'ordine dei record è in genere irrilevante.

Per alcune classi di problemi, tuttavia, la sequenza di record è estremamente importante. Si tratta in genere di situazioni relative a serie temporali, in cui la sequenza di record rappresenta una sequenza ordinata di occorrenze ed eventi. Ogni record rappresenta una snapshot di un momento specifico nel tempo. È tuttavia possibile che la maggior parte delle informazioni utili non sia contenuta nei valori del momento, bensì risieda nella variazione e nel comportamento di tali valori nel corso del tempo.

Il parametro rilevante può anche non essere rappresentato dal tempo. Gli stessi principi possono essere applicati anche a record che rappresentano, per esempio, analisi eseguite a distanze diverse lungo una linea.

Le funzioni speciali e di sequenza sono immediatamente riconoscibili grazie alle seguenti caratteristiche:

- Sono tutte precedute dal prefisso @.
- I nomi sono riportati in caratteri maiuscoli.

Le funzioni sequenza possono fare riferimento al record correntemente elaborato da un nodo, ai record che hanno già attraversato un nodo e anche, in un caso, ai record non ancora passati attraverso un nodo. Le funzioni sequenza possono essere combinate liberamente con altri componenti delle espressioni CLEM, benché alcune presentino restrizioni relative agli argomenti utilizzabili.

Esempi

Può rivelarsi utile conoscere da quanto tempo si è verificato un evento specifico o una condizione si è rivelata vera. A tale scopo, è possibile utilizzare la funzione @SINCE, per esempio:

```
@SINCE(Reddito > Risultati)
```

Questa funzione restituisce l'offset dell'ultimo record in cui questa condizione è stata rilevata come vera, ovvero il numero di record precedenti a quello considerato in cui la condizione era vera. Se la condizione non è mai stata rilevata come vera, @SINCE restituirà @INDEX + 1.

Talvolta può rivelarsi utile fare riferimento a un valore del record corrente nell'espressione utilizzata con @SINCE. A questo scopo, è possibile utilizzare la funzione @THIS, con cui si specifica che il nome di un campo deve essere sempre applicato al record corrente. Per trovare l'offset dell'ultimo record contenente un campo Concentrazione con valore due volte superiore rispetto a quello del record corrente, si può utilizzare quanto segue:

```
@SINCE(Concentration > 2 * @THIS(Concentration))
```

In alcuni casi, la condizione specificata per @SINCE è vera per il record corrente per definizione, per esempio:

```
@SINCE(ID == @THIS(ID))
```

Per questo motivo, @SINCE non valuta la condizione per il record corrente. Per valutare la condizione per il record corrente così come per i precedenti, utilizzare l'analoga funzione @SINCE0. Se la condizione è vera nel record corrente, @SINCE0 restituirà 0.

Tabella 28. Finzioni di sequenza CLEM.

Funzione	Risultato	Descrizione
MEAN(FIELD)	Numero reale	Restituisce la media dei valori per l'argomento <i>FIELD</i> o <i>FIELDS</i> specificato.
@MEAN(FIELD, EXPR)	Numero reale	Restituisce la media dei valori per <i>FIELD</i> negli ultimi record <i>EXPR</i> ricevuti dal nodo corrente, incluso il record corrente. <i>FIELD</i> deve essere il nome di un campo numerico. <i>EXPR</i> può essere qualsiasi valutazione di espressione per un numero intero maggiore di 0. Se <i>EXPR</i> non viene specificato o se il valore specificato è superiore al numero di record ricevuti, viene restituita la media dei valori inclusi in tutti i record ricevuti.
@MEAN(FIELD, EXPR, INT)	Numero reale	Restituisce la media dei valori per <i>FIELD</i> negli ultimi record <i>EXPR</i> ricevuti dal nodo corrente, incluso il record corrente. <i>FIELD</i> deve essere il nome di un campo numerico. <i>EXPR</i> può essere qualsiasi valutazione di espressione per un numero intero maggiore di 0. Se <i>EXPR</i> non viene specificato o se il valore specificato è superiore al numero di record ricevuti, viene restituita la media dei valori inclusi in tutti i record ricevuti. <i>INT</i> indica il numero massimo di valori da cercare all'indietro. Se si specifica questo terzo argomento, si otterrà un'operazione più precisa.
@DIFF1(FIELD)	Numero reale	Restituisce il primo differenziale di <i>FIELD1</i> . La forma con un singolo argomento restituisce semplicemente la differenza tra il valore corrente e il valore precedente del campo. Restituisce 0 se per il campo specificato non esiste alcun valore precedente.
@DIFF1(FIELD1, FIELD2)	Numero reale	La forma con due argomenti fornisce il primo differenziale di <i>FIELD1</i> rispetto a <i>FIELD2</i> . Restituisce 0 se per il campo specificato non esiste alcun valore precedente.
@DIFF2(FIELD)	Numero reale	Restituisce il secondo differenziale di <i>FIELD1</i> . La forma con un singolo argomento restituisce semplicemente la differenza tra il valore corrente e il valore precedente del campo. Restituisce 0 se per il campo specificato non esiste alcun valore precedente.
@DIFF2(FIELD1, FIELD2)	Numero reale	La forma con due argomenti fornisce il primo differenziale di <i>FIELD1</i> rispetto a <i>FIELD2</i> . Restituisce 0 se per il campo specificato non esiste alcun valore precedente.
@INDEX	Numero intero	Restituisce l'indice del record corrente. Gli indici vengono assegnati ai record mano a mano che arrivano al nodo corrente. Al primo record viene assegnato l'indice 1, che viene incrementato di 1 per ciascun record successivo.

Tabella 28. Finzioni di sequenza CLEM (Continua).

Funzione	Risultato	Descrizione
@LAST_NON_BLANK(FIELD)	Qualsiasi	Restituisce l'ultimo valore per <i>FIELD</i> che non era vuoto, come definito in un nodo origine o Tipo a monte. Se i record letti non contengono valori non vuoti per <i>FIELD</i> , viene restituito \$null\$. Si noti che è possibile definire separatamente per ogni campo valori vuoti, denominati anche valori mancanti definiti dall'utente.
@MAX(FIELD)	Numero	Restituisce il valore massimo per il campo specificato con l'argomento <i>FIELD</i> .
@MAX(FIELD, EXPR)	Numero	Restituisce il valore massimo per <i>FIELD</i> negli ultimi record <i>EXPR</i> ricevuti, incluso il record corrente. <i>FIELD</i> deve essere il nome di un campo numerico. <i>EXPR</i> può essere qualsiasi valutazione di espressione per un numero intero maggiore di 0.
@MAX(FIELD, EXPR, INT)	Numero	Restituisce il valore massimo per <i>FIELD</i> negli ultimi record <i>EXPR</i> ricevuti, incluso il record corrente. <i>FIELD</i> deve essere il nome di un campo numerico. <i>EXPR</i> può essere qualsiasi valutazione di espressione per un numero intero maggiore di 0. Se <i>EXPR</i> non viene specificato o se il valore specificato è superiore al numero di record ricevuti, viene restituito il valore massimo dei valori inclusi in tutti i record ricevuti. <i>INT</i> indica il numero massimo di valori da cercare all'indietro. Se si specifica questo terzo argomento, si otterrà un'operazione più precisa.
@MIN(FIELD)	Numero	Restituisce il valore minimo per il campo specificato con l'argomento <i>FIELD</i> .
@MIN(FIELD, EXPR)	Numero	Restituisce il valore minimo per <i>FIELD</i> negli ultimi record <i>EXPR</i> ricevuti, incluso il record corrente. <i>FIELD</i> deve essere il nome di un campo numerico. <i>EXPR</i> può essere qualsiasi valutazione di espressione per un numero intero maggiore di 0.
@MIN(FIELD, EXPR, INT)	Numero	Restituisce il valore minimo per <i>FIELD</i> negli ultimi record <i>EXPR</i> ricevuti, incluso il record corrente. <i>FIELD</i> deve essere il nome di un campo numerico. <i>EXPR</i> può essere qualsiasi valutazione di espressione per un numero intero maggiore di 0. Se <i>EXPR</i> non viene specificato o se il valore specificato è superiore al numero di record ricevuti, viene restituito il valore minimo dei valori inclusi in tutti i record ricevuti. <i>INT</i> indica il numero massimo di valori da cercare all'indietro. Se si specifica questo terzo argomento, si otterrà un'operazione più precisa.

Tabella 28. Finzioni di sequenza CLEM (Continua).

Funzione	Risultato	Descrizione
@OFFSET(FIELD, EXPR)	Qualsiasi	Restituisce il valore di <i>FIELD</i> nel record con un offset dal record corrente pari al valore di <i>EXPR</i> . Un offset positivo fa riferimento a un record già passato, mentre un offset negativo specifica un "lookahead" di un record non ancora pervenuto. Per esempio, @OFFSET(Status, 1) restituisce il valore del campo Status nel record precedente mentre @OFFSET(Status, 4) viene eseguita la ricerca in avanti su quattro record nella sequenza, ovvero su record che non sono ancora passati attraverso questo nodo, per ottenere il valore. <i>Si noti che un offset negativo (lookahead) deve essere specificato come costante.</i> Per i soli offset positivi, <i>EXPR</i> può inoltre rappresentare un'espressione CLEM arbitraria, che può essere valutata per fornire l'offset in relazione al record corrente. In tal caso, l'utilizzo di una versione a tre argomenti di questa funzione dovrebbe migliorare la performance (vedere la funzione successiva). Se questa espressione restituisce un valore diverso da un numero intero non negativo, verrà generato un errore. Non è quindi consentito il calcolo di offset per il lookahead. <i>Nota:</i> Una funzione @OFFSET autoreferenziale non può utilizzare in "lookahead" letterale. Ad esempio, in un nodo Riempimento, non è possibile sostituire il valore di field1 utilizzando un'espressione come @OFFSET(field1,-2).
@OFFSET(FIELD, EXPR, INT)	Qualsiasi	Esegue la stessa operazione della funzione @OFFSET con l'aggiunta di un terzo argomento, <i>INT</i> , che specifica il numero massimo di valori da cercare all'indietro. In casi in cui l'offset viene calcolato da un'espressione, il terzo argomento dovrebbe migliorare le prestazioni. Per esempio in un'espressione come @OFFSET(Foo, Mese, 12) il sistema conserva solo gli ultimi dodici valori di Foo, altrimenti deve archiviare ogni valore. Per i casi in cui il valore di offset sia una costante, inclusi gli offset di "lookahead" negativi, che devono essere costanti, il terzo argomento è inutile ed è necessario usare la versione a due argomenti di questa funzione. Vedere anche la nota relativa alle funzioni autoreferenziali nella versione a due argomenti descritte sopra.
@SDEV(FIELD)	Numero reale	Restituisce la deviazione standard dei valori per l'argomento <i>FIELD</i> o <i>FIELD</i> specificato.
@SDEV(FIELD, EXPR)	Numero reale	Restituisce la deviazione standard dei valori per <i>FIELD</i> negli ultimi record <i>EXPR</i> ricevuti dal nodo corrente, incluso il record corrente. <i>FIELD</i> deve essere il nome di un campo numerico. <i>EXPR</i> può essere qualsiasi valutazione di espressione per un numero intero maggiore di 0. Se <i>EXPR</i> non viene specificato o se il valore specificato è superiore al numero di record ricevuti, viene restituita la deviazione standard in tutti i record ricevuti.

Tabella 28. Finzioni di sequenza CLEM (Continua).

Funzione	Risultato	Descrizione
@SDEV(FIELD, EXPR, INT)	Numero reale	Restituisce la deviazione standard dei valori per <i>FIELD</i> negli ultimi record <i>EXPR</i> ricevuti dal nodo corrente, incluso il record corrente. <i>FIELD</i> deve essere il nome di un campo numerico. <i>EXPR</i> può essere qualsiasi valutazione di espressione per un numero intero maggiore di 0. Se <i>EXPR</i> non viene specificato o se il valore specificato è superiore al numero di record ricevuti, viene restituita la deviazione standard in tutti i record ricevuti. <i>INT</i> indica il numero massimo di valori da cercare all'indietro. Se si specifica questo terzo argomento, si otterrà un'operazione più precisa.
@SINCE(EXPR)	Qualsiasi	Restituisce il numero di record passati da quando l'espressione CLEM arbitraria rappresentata da <i>EXPR</i> è stata rilevata come vera.
@SINCE(EXPR, INT)	Qualsiasi	L'aggiunta del secondo argomento, <i>INT</i> , consente di specificare il numero massimo di record da cercare all'indietro. Se <i>EXPR</i> non è mai stata vera <i>INT</i> corrisponderà a @INDEX+1.
@SINCE@(EXPR)	Qualsiasi	A differenza di @SINCE, considera il record corrente. @SINCE@ restituisce 0 se <i>EXPR</i> risulta vera per il record corrente.
@SINCE@(EXPR, INT)	Qualsiasi	L'aggiunta del secondo argomento, <i>INT</i> , consente di specificare il numero massimo di record da cercare all'indietro.
@SUM(FIELD)	Numero	Restituisce la somma dei valori per l'argomento <i>FIELD</i> o <i>FIELDS</i> specificato.
@SUM(FIELD, EXPR)	Numero	Restituisce la somma dei valori per <i>FIELD</i> negli ultimi record <i>EXPR</i> ricevuti dal nodo corrente, incluso il record corrente. <i>FIELD</i> deve essere il nome di un campo numerico. <i>EXPR</i> può essere qualsiasi valutazione di espressione per un numero intero maggiore di 0. Se <i>EXPR</i> non viene specificato o se il valore specificato è superiore al numero di record ricevuti, viene restituita la somma dei valori inclusi in tutti i record ricevuti.
@SUM(FIELD, EXPR, INT)	Numero	Restituisce la somma dei valori per <i>FIELD</i> negli ultimi record <i>EXPR</i> ricevuti dal nodo corrente, incluso il record corrente. <i>FIELD</i> deve essere il nome di un campo numerico. <i>EXPR</i> può essere qualsiasi valutazione di espressione per un numero intero maggiore di 0. Se <i>EXPR</i> non viene specificato o se il valore specificato è superiore al numero di record ricevuti, viene restituita la somma dei valori inclusi in tutti i record ricevuti. <i>INT</i> indica il numero massimo di valori da cercare all'indietro. Se si specifica questo terzo argomento, si otterrà un'operazione più precisa.
@THIS(FIELD)	Qualsiasi	Restituisce il valore del campo denominato <i>FIELD</i> nel record corrente. Viene utilizzata solo nelle espressioni @SINCE.

Funzioni globali

Le funzioni @MEAN,@SUM, @MIN, @MAX e @SDEV sono applicabili, al massimo, ai record precedentemente esaminati e al record corrente. In alcuni casi, tuttavia, è utile poter confrontare i valori nel record corrente

con i valori rilevati nell'intero insieme di dati. Se si utilizza un nodo Calcola globali per generare valori nell'intero insieme di dati, è possibile accedere a tali valori in un'espressione CLEM mediante le funzioni globali.

Per esempio,

@GLOBAL_MAX(Age)

restituisce il valore più elevato di Age nell'insieme di dati, mentre l'espressione

(Value - @GLOBAL_MEAN(Value)) / @GLOBAL_SDEV(Value)

esprime la differenza tra il valore Value di questo record e la media globale come numero di deviazioni standard. È possibile utilizzare i valori globali solo dopo che sono stati calcolati in un nodo Calcola globali. Tutti i valori globali correnti possono essere annullati facendo clic sul pulsante **Cancella valori globali** nella scheda Valori globali della finestra di dialogo delle proprietà del flusso.

Tabella 29. Funzioni globali CLEM.

Funzione	Risultato	Descrizione
@GLOBAL_MAX(CAMPO)	Numero	Restituisce il valore massimo per <i>FIELD</i> nell'intero insieme di dati, generato in precedenza da un nodo Calcola globali. <i>FIELD</i> deve essere il nome di un campo numerico. Se il valore globale corrispondente non è stato impostato, si verifica un errore.
@GLOBAL_MIN(CAMPO)	Numero	Restituisce il valore minimo per <i>FIELD</i> nell'intero insieme di dati, generato in precedenza da un nodo Calcola globali. <i>FIELD</i> deve essere il nome di un campo numerico. Se il valore globale corrispondente non è stato impostato, si verifica un errore.
@GLOBAL_SDEV(CAMPO)	Numero	Restituisce la deviazione standard dei valori per <i>FIELD</i> nell'intero insieme di dati, generato in precedenza da un nodo Calcola globali. <i>FIELD</i> deve essere il nome di un campo numerico. Se il valore globale corrispondente non è stato impostato, si verifica un errore.
@GLOBAL_MEAN(CAMPO)	Numero	Restituisce la media dei valori per <i>FIELD</i> nell'intero insieme di dati, generato in precedenza da un nodo Calcola globali. <i>FIELD</i> deve essere il nome di un campo numerico. Se il valore globale corrispondente non è stato impostato, si verifica un errore.
@GLOBAL_SUM(CAMPO)	Numero	Restituisce la somma dei valori per <i>FIELD</i> nell'intero insieme di dati, generato in precedenza da un nodo Calcola globali. <i>FIELD</i> deve essere il nome di un campo numerico. Se il valore globale corrispondente non è stato impostato, si verifica un errore.

Funzioni per la gestione di valori vuoti e null

Mediante il linguaggio CLEM è possibile stabilire che valori specifici in un campo devono essere considerati come valori vuoti o mancanti. Le funzioni illustrate di seguito gestiscono i valori vuoti.

Tabella 30. Funzioni con valore null e vuoto di CLEM.

Funzione	Risultato	Descrizione
@BLANK(FIELD)	Booleano	Restituisce un valore vero per tutti i record contenenti valori vuoti in base alle regole di gestione dei vuoti impostate in un nodo origine o Tipo a monte mediante la scheda Tipi.

Tabella 30. Funzioni con valore null e vuoto di CLEM (Continua).

Funzione	Risultato	Descrizione
@LAST_NON_BLANK(CAMPO)	Qualsiasi	Restituisce l'ultimo valore per <i>FIELD</i> che non era vuoto, come definito in un nodo origine o Tipo a monte. Se i record letti non contengono valori non vuoti per <i>FIELD</i> , viene restituito \$null\$. Si noti che è possibile definire separatamente per ogni campo valori vuoti, denominati anche valori mancanti definiti dall'utente.
@NULL(FIELD)	Booleano	Restituisce un valore vero se il valore di <i>FIELD</i> corrisponde al valore mancante di sistema \$null\$. Restituisce falso per tutti gli altri valori, inclusi i vuoti definiti dall'utente. Per verificare entrambi, utilizzare @BLANK(FIELD) e @NULL(FIELD).
undef	Qualsiasi	Viene utilizzata in genere in CLEM per immettere un valore \$null\$, per esempio per sostituire i valori vuoti con valori null nel nodo Riempimento.

I campi vuoti possono essere "riempiti" con il nodo Riempimento. In entrambi i nodi Riempimento e Ricava, solo in modalità multipla, la funzione speciale CLEM @FIELD fa riferimento ai campi correnti esaminati.

Campi speciali

Le funzioni speciali vengono utilizzate per indicare i campi specifici esaminati oppure per generare un elenco di campi come input. In caso di derivazione di più campi, per esempio, è opportuno utilizzare @FIELD per indicare che l'operazione di derivazione deve essere eseguita sui campi selezionati. Se si utilizza l'espressione $\log(@FIELD)$, viene derivato un nuovo campo di logaritmo per ogni campo selezionato.

Tabella 31. Campi speciali CLEM.

Funzione	Risultato	Descrizione
@FIELD	Qualsiasi	Esegue un'operazione su tutti i campi specificati nel contesto dell'espressione.
@TARGET	Qualsiasi	In caso di utilizzo di un'espressione CLEM in una funzione di analisi definita dall'utente @TARGET rappresenta il campo obiettivo o il "valore corretto" per la coppia obiettivo/previsto analizzata. Questa funzione viene in genere utilizzata in un nodo Analisi.
@PREDICTED	Qualsiasi	In caso di utilizzo di un'espressione CLEM in una funzione di analisi definita dall'utente, @PREDICTED rappresenta il campo previsto per la coppia obiettivo/previsto analizzata. Questa funzione viene in genere utilizzata in un nodo Analisi.
@PARTITION_FIELD	Qualsiasi	Sostituisce il nome del campo partizione corrente.

Tabella 31. Campi speciali CLEM (Continua).

Funzione	Risultato	Descrizione
@TRAINING_PARTITION	Qualsiasi	Restituisce il valore della partizione dell'addestramento corrente. Per esempio, per selezionare i record di addestramento tramite un nodo Selezione, usare l'espressione CLEM: @PARTITION_FIELD = @TRAINING_PARTITION. In tal modo, il nodo Selezione funzionerà indipendentemente dai valori utilizzati per rappresentare ogni partizione nei dati.
@TESTING_PARTITION	Qualsiasi	Restituisce il valore della partizione di test corrente.
@VALIDATION_PARTITION	Qualsiasi	Restituisce il valore della partizione di convalida corrente.
@FIELDS_BETWEEN(inizio, fine)	Qualsiasi	Restituisce l'elenco di nomi di campo compresi tra i campi iniziale e finale (inclusi) specificati in base all'ordine naturale (cioè di inserimento) dei campi nei dati.
@FIELDS_MATCHING(schema)	Qualsiasi	Restituisce un elenco di nomi di campo corrispondenti a uno schema specificato. È possibile includere nello schema un punto interrogativo (?) corrispondente esattamente a un carattere; un asterisco (*) corrisponde a zero o più caratteri. Per indicare un punto interrogativo o un asterisco letterale, ovvero non utilizzato come carattere jolly, è possibile utilizzare la barra rovesciata (\).
@MULTI_RESPONSE_SET	Qualsiasi	Restituisce l'elenco dei campi nell'insieme di risposte multiple denominato.

Capitolo 9. Utilizzo di IBM SPSS Modeler con un repository

Informazioni su IBM SPSS Collaboration and Deployment Services Repository

IBM SPSS Modeler può essere utilizzato insieme a un repository IBM SPSS Collaboration and Deployment Services, il che consente di gestire il ciclo di vita dei modelli di data mining e degli oggetti predittivi correlati, oltre a consentire l'utilizzo di questi oggetti dalle applicazioni, dagli strumenti e dalle soluzioni aziendali. Gli oggetti IBM SPSS Modeler che possono essere condivisi in questo modo includono i flussi, i nodi, gli output dei flussi, gli scenari, i progetti e i modelli. Gli oggetti vengono memorizzati nel repository centralizzato, da dove è possibile condividerli con altre applicazioni e registrarli utilizzando funzionalità avanzate per il controllo delle versioni, dei metadati e della ricerca.

Nota: per l'accesso ad un repository IBM SPSS Collaboration and Deployment Services è richiesta una licenza separata. Per ulteriori informazioni, consultare <http://www.ibm.com/software/analytics/spss/products/deployment/cds/>

Prima di poter utilizzare IBM SPSS Modeler con il repository, è necessario installare un adattatore sull'host del repository. Senza l'adattatore, quando si tenta di accedere agli oggetti del repository da alcuni nodi o modelli di IBM SPSS Modeler è possibile che venga visualizzato il seguente messaggio:

Potrebbe essere necessario aggiornare il repository in modo che supporti nuovi tipi di nodi, modelli e output.

Per istruzioni sull'installazione dell'adattatore, vedere la guida *IBM SPSS Modeler Deployment Installation*, disponibile in IBM SPSS Modeler Deployment DVD. Per informazioni su come accedere agli oggetti del repository IBM SPSS Modeler da IBM SPSS Collaboration and Deployment Services Deployment Manager, consultare il documento *IBM SPSS Modeler Deployment Guide*.

Le sezioni che seguono contengono informazioni su come accedere al repository da IBM SPSS Modeler.

Supporto completo del controllo delle versioni e della ricerca

Il repository offre funzionalità complete di controllo delle versioni e di ricerca. Per esempio, si supponga di creare un flusso e di memorizzarlo nel repository, in modo da poterlo condividere con i ricercatori di altre divisioni. Se in seguito si aggiorna il flusso in IBM SPSS Modeler, è possibile aggiungere la versione aggiornata al repository senza sovrascrivere la versione precedente. Tutte le versioni rimangono accessibili ed è possibile eseguire ricerche per nome, etichetta, campi utilizzati o altri attributi. Per esempio, è possibile cercare tutte le versioni del modello che utilizzano le entrate nette come input o tutti i modelli creati da un autore specifico. Per eseguire queste operazioni in un file system tradizionale, sarebbe necessario salvare ogni versione con un nome file diverso e il programma non sarebbe a conoscenza delle relazioni tra le diverse versioni.

Single Sign-On (SSO)

La funzione Single Sign-On (SSO) consente agli utenti di collegarsi al repository senza dover immettere ogni volta nome utente e password. I dettagli di accesso alla rete locale dell'utente forniscono l'autenticazione necessaria per accedere a IBM SPSS Collaboration and Deployment Services. Questa funzione dipende dalle seguenti condizioni:

- IBM SPSS Collaboration and Deployment Services deve essere configurato in modo da utilizzare un provider di Single Sign-On (SSO)
- L'utente deve avere eseguito l'accesso a un host compatibile con il provider

Consultare l'argomento "Connessione al repository" a pagina 119 per ulteriori informazioni.

Archiviazione e deployment di oggetti di repository

I flussi creati in IBM SPSS Modeler possono essere **archiviati** nel repository così come sono, come file con estensione *.str*. In questo modo, più utenti in tutta l'azienda possono accedere a un unico stream. Consultare l'argomento "Archiviazione di oggetti in repository" a pagina 120 per ulteriori informazioni.

È possibile anche eseguire il **deployment** di un flusso nel repository. Uno stream sottoposto a deployment è memorizzato come file con metadati aggiuntivi. Uno stream sottoposto a deployment può beneficiare al massimo delle funzioni applicabili a livello aziendale di IBM SPSS Collaboration and Deployment Services, come il calcolo automatico del punteggio e l'aggiornamento del modello. Per esempio, un modello può essere aggiornato automaticamente a intervalli regolari man mano che sono disponibili nuovi dati. In alternativa, è possibile eseguire la distribuzione di un insieme di stream per l'analisi Campione/Sfidante, in cui si confrontano i flussi per determinare quale contiene il modello predittivo più efficace.

È possibile distribuire un flusso in uno dei due modi: come flusso (con estensione *.str*) o come scenario (con estensione *.scn*). La distribuzione come flusso consente al flusso di essere utilizzato dall'applicazione thin client IBM SPSS Modeler Advantage. Consultare l'argomento "Apertura di un flusso in IBM SPSS Modeler Advantage" a pagina 137 per ulteriori informazioni. La distribuzione come scenario consente al flusso di essere utilizzato da Applicazioni predittive versione 5, il predecessore di IBM SPSS Modeler Advantage.

Per ulteriori informazioni, consultare "Opzioni di distribuzione dei flussi" a pagina 131.

Requisiti per i flussi implementati come scenari

- Per garantire un accesso coerente ai dati aziendali, è necessario accedere ai flussi di cui è stata eseguita la distribuzione tramite il componente Enterprise View di IBM SPSS Collaboration and Deployment Services. Ciò significa che in IBM SPSS Modeler deve esserci almeno un nodo origine Enterprise View all'interno di ogni ramo designato del flusso, sia di calcolo del punteggio che di creazione del modello.
- Per utilizzare il nodo Enterprise View, IBM SPSS Collaboration and Deployment Services deve essere installato, configurato e accessibile dal proprio sito, con una Enterprise View, le Visualizzazioni applicazione e le DPD già definite. Per ulteriori informazioni, rivolgersi all'amministratore di sistema oppure visitare il sito Web aziendale all'indirizzo <http://www.ibm.com/software/analytics/spss/products/deployment/cds/>.
- Una Definizione dei provider di dati (DPD) viene definita rispetto a una determinata sorgente dati ODBC. Per utilizzare una DPD da IBM SPSS Modeler, è necessario che sull'host server di IBM SPSS Modeler sia stata definita una sorgente dati ODBC con lo stesso nome e che si connette allo stesso archivio dati di quella a cui fa riferimento la Definizione dei provider di dati.
- È inoltre necessario che IBM SPSS Collaboration and Deployment Services Enterprise View Driver sia installato su ogni computer utilizzato per modificare o eseguire il flusso. Per Windows, installare semplicemente il driver sul computer su cui è installato SPSS Modeler o SPSS Modeler Server. Non sono necessarie ulteriori configurazioni del driver. In UNIX, è necessario aggiungere allo script di avvio un riferimento allo script *pev.sh*. Per dettagli sull'installazione di IBM SPSS Collaboration and Deployment Services Enterprise View Driver, rivolgersi all'amministratore di sistema.

Altre opzioni di distribuzione

Mentre IBM SPSS Collaboration and Deployment Services offre le più estese e avanzate funzionalità di gestione del contenuto aziendale, sono disponibili altri meccanismi per la distribuzione o l'esportazione di stream, tra cui:

- Esportare il flusso e il modello da utilizzare successivamente con IBM SPSS Modeler Solution Publisher Runtime.

- Esportare uno o più modelli in PMML, un formato basato su XML per la codifica delle informazioni sui modelli. Consultare l'argomento "Importazione ed esportazione di modelli come PMML" a pagina 138 per ulteriori informazioni.

Connessione al repository

1. Per connettersi al repository, nel menu principale di IBM SPSS Modeler fare clic su:
Strumenti > Repository > Opzioni...
2. Nel campo **RepositoryURL**, immettere o selezionare il percorso della directory o l'URL dell'installazione del repository a cui si desidera accedere. È possibile connettersi a un solo repository alla volta.
Le impostazioni sono specifiche di ogni sito o installazione. Per i dettagli di un accesso specifico contattare l'amministratore del sistema locale.

Nota: per l'accesso ad un repository IBM SPSS Collaboration and Deployment Services è richiesta una licenza separata. Per ulteriori informazioni, consultare <http://www.ibm.com/software/analytics/spss/products/deployment/cds/>

Imposta credenziali. Lasciare questa casella deselezionata per abilitare la funzione di **Single Sign-On (SSO)**, che cercherà di eseguire l'accesso dell'utente utilizzando il nome utente e la password del computer locale. Se la funzione di Single Sign-On (SSO) non è disponibile oppure se si seleziona la casella per disabilitare Single Sign-On (SSO) (per esempio, per accedere a un account amministratore), viene visualizzata un'ulteriore schermata per l'inserimento delle credenziali.

Immissione di credenziali per il repository

A seconda delle impostazioni, i seguenti campi possono essere necessari nella finestra di dialogo Repository: Credenziali:

ID utente e password. Specificare un nome utente e una password validi per il login. Se necessario, rivolgersi all'amministratore locale per ulteriori informazioni.

Provider. Scegliere un provider di protezione per l'autenticazione. Il repository può essere configurato in modo da utilizzare diversi provider di protezione. Se sono necessarie ulteriori informazioni, rivolgersi all'amministratore di sistema.

Memorizza repository e ID utente. Salva le impostazioni correnti come di default, in modo tale che non sia necessario reinserirle ogni volta che si desidera eseguire la connessione.

Esplorazione dei contenuti di repository

Il repository consente di esplorare i contenuti archiviati in modo simile a Esplora risorse di Windows; è possibile anche cercare le *versioni* di ogni oggetto archiviato.

1. Per aprire la finestra del IBM SPSS Collaboration and Deployment Services Repository, nei menu di SPSS Modeler fare clic su:

Strumenti > Repository > Esplora...

1. Specificare le impostazioni relative alla connessione al repository, se necessario. Consultare l'argomento "Connessione al repository" per ulteriori informazioni. Per informazioni specifiche sulla porta, la password e altri dettagli per la connessione, rivolgersi all'amministratore del sistema locale.

La finestra di esplorazione visualizza inizialmente una struttura ad albero della gerarchia delle cartelle. Fare clic sul nome di una cartella per visualizzarne il contenuto.

Nel riquadro destro vengono visualizzati gli oggetti corrispondenti al criterio di selezione o di ricerca corrente, nonché informazioni dettagliate sulla versione selezionata visualizzata nel riquadro in basso a destra. Gli attributi visualizzati sono validi per la versione più recente.

Archiviazione di oggetti in repository

Flussi, nodi, modelli, palette di modelli, progetti e oggetti di output possono essere archiviati nel repository, da dove altri utenti e applicazioni potranno accedervi.

Nota: per l'accesso ad un repository IBM SPSS Collaboration and Deployment Services è richiesta una licenza separata. Per ulteriori informazioni, consultare <http://www.ibm.com/software/analytics/spss/products/deployment/cds/>

Nel repository è anche possibile pubblicare output dei flussi in un formato che ne consenta la visualizzazione ad altri utenti via Internet mediante IBM SPSS Collaboration and Deployment Services Deployment Portal.

Impostazione delle proprietà degli oggetti

Quando si archivia un oggetto, viene visualizzata la finestra di dialogo Repository: Archivia, che consente di impostare i valori di diverse proprietà dell'oggetto. È possibile:

- Scegliere il nome e la cartella del repository in cui archiviare l'oggetto
- Aggiungere informazioni sull'oggetto, per esempio l'etichetta versione e altre proprietà consultabili tramite ricerche
- Assegnare all'oggetto uno o più argomenti di classificazione
- Impostare le opzioni di sicurezza per l'oggetto

Le sezioni seguenti descrivono le proprietà che è possibile impostare.

Scelta della posizione di archiviazione degli oggetti

Nella finestra di dialogo Repository: Archivia, immettere quanto riportato di seguito.

Salva in. Mostra la cartella corrente, ovvero la posizione in cui verrà archiviato l'oggetto. Fare doppio clic sul nome di una cartella nell'elenco per impostare quella cartella come la cartella corrente. Utilizzare il pulsante Cartella superiore per navigare fino alla cartella padre. Utilizzare il pulsante Nuova cartella per creare una cartella al livello corrente.

Nome file. Il nome con cui verrà archiviato l'oggetto.

Archivia. Archivia l'oggetto nella posizione corrente.

Aggiunta di informazioni sugli oggetti archiviati

Tutti i campi nella scheda Informazioni della finestra di dialogo Repository: Archivia sono facoltativi.

Autore. Il nome utente di chi ha creato l'oggetto nel repository. Per default, mostra il nome utente utilizzato per la connessione al repository, ma è possibile cambiare il nome qui.

Etichetta versione. Selezionare un'etichetta dall'elenco per indicare la versione dell'oggetto, oppure fare clic su **Aggiungi** per creare una nuova etichetta. Evitare di usare il carattere "[" nell'etichetta. Se non si desidera assegnare un'etichetta a questa versione dell'oggetto, non selezionare alcuna casella. Consultare l'argomento "Visualizzazione e modifica delle proprietà degli oggetti" a pagina 129 per ulteriori informazioni.

Descrizione. Una descrizione dell'oggetto. Gli utenti possono effettuare ricerche di oggetti in base alla descrizione (vedere la nota).

Parole chiave. Una o più parole chiave che sono correlate all'oggetto e che possono essere utilizzate a scopo di ricerca (vedere la nota).

Scadenza. La data trascorsa la quale l'oggetto non è più visibile agli utenti generici, ma solo al proprietario e all'amministratore del repository. Per impostare una data di scadenza, selezionare l'opzione **Data** e inserire una data, oppure sceglierne una utilizzando il pulsante del calendario.

Archivia. Archivia l'oggetto nella posizione corrente.

Nota: le informazioni contenute nei campi **Descrizione** e **Parole chiave** sono considerate separatamente rispetto ai dati inseriti in SPSS Modeler nella scheda Annotazioni dell'oggetto. Una ricerca per descrizione o parola chiave nel repository non restituisce le informazioni contenute nella scheda Annotazioni. Consultare l'argomento "Ricerca di oggetti nel repository" a pagina 125 per ulteriori informazioni.

Assegnazione di argomenti a un oggetto archiviato

Gli argomenti sono un sistema di classificazione gerarchico per il contenuto archiviato nel repository. È possibile scegliere tra gli argomenti disponibili durante l'archiviazione degli oggetti e gli utenti possono inoltre cercare gli oggetti per argomento. L'elenco degli argomenti disponibili è impostato dagli utenti del repository che dispongono dei privilegi appropriati (per maggiori informazioni, vedere il manuale *Deployment Manager User's Guide*).

Per assegnare un argomento all'oggetto, dalla scheda Argomenti della finestra di dialogo Repository: Archivio:

1. Fare clic sul pulsante **Aggiungi**.
2. Fare clic sul nome di un argomento dall'elenco degli argomenti disponibili.
3. Fare clic su **OK**.

Per rimuovere un'assegnazione di argomento:

4. Selezionare l'argomento nell'elenco degli argomenti assegnati.
5. Fare clic su **Elimina**.

Impostazione delle opzioni di protezione per gli oggetti archiviati

È possibile impostare o modificare una serie di opzioni di sicurezza per un oggetto archiviato nella scheda Sicurezza della finestra di dialogo Repository: Archivio. Per uno o più **principali** (ovvero, utenti o gruppi di utenti), è possibile:

- Assegnare i diritti di accesso all'oggetto
- Modificare i diritti di accesso all'oggetto
- Rimuovere i diritti di accesso all'oggetto

Principale. Il nome utente repository dell'utente o del gruppo che ha diritti di accesso su quell'oggetto.

Autorizzazioni. I diritti di accesso di cui questo utente o gruppo dispone per l'oggetto.

Aggiungi. Consente di aggiungere uno o più utenti o gruppi all'elenco di quelli che dispongono di diritti di accesso su questo oggetto. Consultare l'argomento "Aggiunta di un utente all'elenco delle autorizzazioni" per ulteriori informazioni.

Modifica. Consente di modificare i diritti di accesso dell'utente o gruppo selezionato per questo oggetto. L'accesso in lettura è garantito per default. Questa opzione consente di garantire ulteriori diritti di accesso, nello specifico i diritti Proprietario, Scrittura, Eliminazione e Modifica autorizzazioni.

Elimina. Elimina l'utente o il gruppo selezionato dall'elenco delle autorizzazioni per questo oggetto.

Aggiunta di un utente all'elenco delle autorizzazioni: Quando si seleziona **Aggiungi** nella scheda Sicurezza della finestra di dialogo Repository: Archivio sono disponibili i seguenti campi:

Seleziona provider. Scegliere un provider di protezione per l'autenticazione. Il repository può essere configurato in modo da utilizzare diversi provider di protezione. Se sono necessarie ulteriori informazioni, rivolgersi all'amministratore di sistema.

Trova. Inserire il nome utente repository dell'utente o del gruppo che si desidera aggiungere, quindi fare clic su **Cerca** per visualizzare quel nome nell'elenco degli utenti. Per aggiungere più di un nome utente alla volta, lasciare vuoto questo campo e fare semplicemente clic su **Cerca** per visualizzare un elenco di tutti i nomi utente repository.

Elenco utenti. Selezionare uno o più nomi utente dall'elenco e fare clic su OK per aggiungerli all'elenco delle autorizzazioni.

Modifica dei diritti di accesso per un oggetto: Quando si seleziona **Modifica** nella scheda Sicurezza della finestra di dialogo Repository: Archivio sono disponibili i seguenti campi:

Proprietario. Selezionare questa opzione per assegnare a questo utente o gruppo i diritti di accesso all'oggetto. Il proprietario ha il pieno controllo sull'oggetto, inclusi i diritti di accesso Eliminazione e Modifica.

Lettura. Per default, un utente o un gruppo che non è il proprietario dell'oggetto dispone solo dei diritti di accesso Lettura sull'oggetto. Selezionare le caselle di controllo appropriate per aggiungere i diritti di accesso Scrittura, Eliminazione e Modifica autorizzazioni per questo utente o gruppo.

Archiviazione di flussi

È possibile archiviare un flusso come file *.str* nel repository, da dove altri utenti possono accedervi.

Nota: per informazioni sul deployment di un flusso per sfruttare appieno altre funzioni del repository, vedere "Distribuzione di flussi" a pagina 130.

Per archiviare il flusso corrente:

1. Nel menu principale, fare clic su:
File > Archivia > Archivia come flusso...
2. Specificare le impostazioni relative alla connessione al repository, se necessario. Consultare l'argomento "Connessione al repository" a pagina 119 per ulteriori informazioni. Per informazioni specifiche sulla porta, la password e altri dettagli per la connessione, rivolgersi all'amministratore del sistema locale.
3. Nella finestra di dialogo Repository: Archivia, scegliere la cartella in cui archiviare l'oggetto, specificare le eventuali altre informazioni da registrare e fare clic sul pulsante **Archivia**. Consultare l'argomento "Impostazione delle proprietà degli oggetti" a pagina 120 per ulteriori informazioni.

Archiviazione di progetti

È possibile archiviare un progetto IBM SPSS Modeler completo come file *.cpj* nel repository, in modo che altri utenti possano accedervi.

Poiché un file di progetto è un contenitore per altri oggetti IBM SPSS Modeler, è necessario dire a IBM SPSS Modeler di archiviare gli oggetti del progetto nel repository. Questo parametro può essere impostato nella finestra di dialogo Proprietà progetto. Consultare l'argomento "Impostazione delle proprietà del progetto" a pagina 144 per ulteriori informazioni.

Ogni volta che si aggiunge un nuovo oggetto a un progetto, IBM SPSS Modeler chiede automaticamente di archivarlo se il progetto è configurato per l'archiviazione degli oggetti nel repository.

Al termine della sessione di IBM SPSS Modeler è necessario archiviare una nuova versione del file di progetto che comprenda i nuovi oggetti aggiunti. Il file di progetto contiene e richiama automaticamente

le ultime versioni degli oggetti. Se durante una sessione di IBM SPSS Modeler non sono stati aggiunti nuovi oggetti, non è necessario archiviare nuovamente il file di progetto. Occorre però archiviare le nuove versioni degli oggetti del progetto (stream, output e così via) che sono stati modificati.

Per archiviare un progetto

1. Selezionare il progetto nella scheda CRISP-DM o Classi nel riquadro dei manager in IBM SPSS Modeler e dal menu principale fare clic su:
File > Progetti > Archivia progetto...
2. Specificare le impostazioni relative alla connessione al repository, se necessario. Consultare l'argomento "Connessione al repository" a pagina 119 per ulteriori informazioni. Per informazioni specifiche sulla porta, la password e altri dettagli per la connessione, rivolgersi all'amministratore del sistema locale.
3. Nella finestra di dialogo Repository: Archivia, scegliere la cartella in cui archiviare l'oggetto, specificare le eventuali altre informazioni da registrare e fare clic sul pulsante **Archivia**. Consultare l'argomento "Impostazione delle proprietà degli oggetti" a pagina 120 per ulteriori informazioni.

Archiviazione di nodi

È possibile archiviare una definizione di nodo singolo dal flusso corrente come file *.nod* nel repository, da dove altri utenti possono accedervi.

Per archiviare un nodo:

1. Fare clic sul nodo con il tasto destro del mouse nell'area del flusso e fare clic su **Archivia nodo**.
2. Specificare le impostazioni relative alla connessione al repository, se necessario. Consultare l'argomento "Connessione al repository" a pagina 119 per ulteriori informazioni. Per informazioni specifiche sulla porta, la password e altri dettagli per la connessione, rivolgersi all'amministratore del sistema locale.
3. Nella finestra di dialogo Repository: Archivia, scegliere la cartella in cui archiviare l'oggetto, specificare le eventuali altre informazioni da registrare e fare clic sul pulsante **Archivia**. Consultare l'argomento "Impostazione delle proprietà degli oggetti" a pagina 120 per ulteriori informazioni.

Archiviazione di oggetti di output

È possibile archiviare un oggetto di output dal flusso corrente come file *.cou* nel repository, da dove altri utenti possono accedervi.

Per archiviare un oggetto di output:

1. Selezionare l'oggetto nella scheda Output del riquadro dei manager in SPSS Modeler e dal menu principale fare clic su:
File > Output > Archivia output...
2. In alternativa, fare clic con il tasto destro del mouse su un oggetto nella scheda Output e fare clic su **Archivia**.
3. Specificare le impostazioni relative alla connessione al repository, se necessario. Consultare l'argomento "Connessione al repository" a pagina 119 per ulteriori informazioni. Per informazioni specifiche sulla porta, la password e altri dettagli per la connessione, rivolgersi all'amministratore del sistema locale.
4. Nella finestra di dialogo Repository: Archivia, scegliere la cartella in cui archiviare l'oggetto, specificare le eventuali altre informazioni da registrare e fare clic sul pulsante **Archivia**. Consultare l'argomento "Impostazione delle proprietà degli oggetti" a pagina 120 per ulteriori informazioni.

Archiviazione di modelli e palette di modelli

È possibile archiviare un modello singolo come file *.gm* nel repository, da dove altri utenti possono accedervi. È anche possibile archiviare il contenuto completo della palette Modelli come file *.gen* nel repository.

Archiviazione di un modello:

1. Fare clic sull'oggetto nella palette Modelli in SPSS Modeler e nel menu principale fare clic su:
File > Modelli > Archivia modello...
2. In alternativa, fare clic con il tasto destro del mouse su un oggetto nella palette Modelli e fare clic su **Archivia modello**.
3. Continuare da "Completamento della procedura di archiviazione".

Archiviazione di una tavolozza dei Modelli

1. Fare clic con il tasto destro del mouse sullo sfondo della palette Modelli.
2. Nel menu di scelta rapida, fare clic su **Archivia palette**.
3. Continuare da "Completamento della procedura di archiviazione".

Completamento della procedura di archiviazione:

1. Specificare le impostazioni relative alla connessione al repository, se necessario. Consultare l'argomento "Connessione al repository" a pagina 119 per ulteriori informazioni. Per informazioni specifiche sulla porta, la password e altri dettagli per la connessione, rivolgersi all'amministratore del sistema locale.
2. Nella finestra di dialogo Repository: Archivia, scegliere la cartella in cui archiviare l'oggetto, specificare le eventuali altre informazioni da registrare e fare clic sul pulsante **Archivia**. Consultare l'argomento "Impostazione delle proprietà degli oggetti" a pagina 120 per ulteriori informazioni.

Recupero di oggetti da repository

È possibile recuperare i flussi, i modelli, le palette di modelli, i nodi, i progetti e gli oggetti di output archiviati nel repository.

Nota: oltre ad utilizzare le opzioni di menu descritte qui, è anche possibile recuperare i flussi, gli oggetti di output, i modelli e le tavolozza dei modelli facendo clic con il tasto destro del mouse sulla scheda appropriata del riquadro Manager nella parte superiore destra della finestra di SPSS Modeler.

1. Per recuperare un flusso, nel menu principale di IBM SPSS Modeler fare clic su:
File > Recupera flusso...
2. Per recuperare un modello, una palette di modelli, un progetto o un oggetto di output, nel menu principale di IBM SPSS Modeler fare clic su:
File > Modelli > Recupera modello...
o
File > Modelli > Recupera tavolozza dei modelli...
o
File > Progetti > Recupera progetto...
o
File > Output > Recupera output...
3. In alternativa, fare clic con il tasto destro del mouse nel riquadro dei manager o del progetto e fare clic su **Recupera** dal menu di scelta rapida.
4. Per recuperare un nodo, nel menu principale di IBM SPSS Modeler fare clic su:
Inserisci > Nodo (o Supernodo) dal repository...

- a. Specificare le impostazioni relative alla connessione al repository, se necessario. Consultare l'argomento "Connessione al repository" a pagina 119 per ulteriori informazioni. Per informazioni specifiche sulla porta, la password e altri dettagli per la connessione, rivolgersi all'amministratore del sistema locale.
5. Dalla finestra di dialogo Repository: Recupera, navigare fino all'oggetto, selezionarlo e fare clic sul pulsante **Recupera**. Consultare l'argomento per ulteriori informazioni.

Scelta di un oggetto da recuperare

I seguenti campi sono disponibili nella finestra di dialogo Repository: Recupera/Cerca:

Cerca in. Mostra la gerarchia delle cartelle per la cartella corrente. Per navigare a una cartella diversa, selezionarne una da questo elenco per arrivare direttamente a quella cartella, oppure navigare utilizzando l'elenco di oggetti sotto a questo campo.

Pulsante Cartella superiore. Si sposta di un livello verso l'alto rispetto alla cartella corrente nella gerarchia.

Pulsante Nuova cartella. Crea una nuova cartella al livello corrente nella gerarchia.

Nome file. Il nome file del repository dell'oggetto selezionato. Per recuperare quell'oggetto, fare clic su **Recupera**.

File di tipo. Il tipo di oggetto che si è scelto di recuperare. Solo gli oggetti di questo tipo, insieme alle cartelle, vengono mostrati nell'elenco degli oggetti. Per visualizzare gli oggetti di un tipo diverso di recupero, selezionare il tipo di oggetto dall'elenco.

Apri come bloccato. Per default, quando un oggetto viene recuperato, è bloccato nel repository in modo che altri non possano aggiornarlo. Se non si desidera che l'oggetto venga bloccato al recupero, deselezionare questa casella.

Descrizione, Parole chiave. Se erano stati definiti ulteriori dettagli sull'oggetto al momento dell'archiviazione, tali dettagli verranno visualizzati qui. Consultare l'argomento "Aggiunta di informazioni sugli oggetti archiviati" a pagina 120 per ulteriori informazioni.

Versione. Per recuperare una versione dell'oggetto diversa dall'ultima, fare clic su questo pulsante. Vengono visualizzate informazioni dettagliate su tutte le versioni ed è possibile scegliere la versione desiderata.

Selezione della versione di un oggetto

Per selezionare una versione specifica di un oggetto del repository, nella finestra di dialogo Repository: Seleziona versione:

1. (Facoltativo) Ordinare l'elenco per versione, etichetta, dimensione, data di creazione o autore della creazione, facendo doppio clic sull'intestazione della colonna appropriata.
2. Selezionare la versione dell'oggetto con cui si desidera lavorare.
3. Fare clic su Continua.

Ricerca di oggetti nel repository

È possibile effettuare una ricerca degli oggetti in base al nome, alla cartella, al tipo, all'etichetta o in base ad altri criteri.

Ricerca per nome

Per cercare gli oggetti in base al nome:

1. Nel menu principale di IBM SPSS Modeler, fare clic su:
Strumenti > Repository > Esplora...
 - a. Specificare le impostazioni relative alla connessione al repository, se necessario. Consultare l'argomento "Connessione al repository" a pagina 119 per ulteriori informazioni. Per informazioni specifiche sulla porta, la password e altri dettagli per la connessione, rivolgersi all'amministratore del sistema locale.
2. Fare clic sulla scheda **Ricerca**.
3. Nel campo **Cerca oggetti con nome**, specificare il nome dell'oggetto che si desidera trovare.

Quando si cercano oggetti in base al nome è possibile utilizzare l'asterisco (*) come carattere jolly al posto di qualsiasi stringa di caratteri e il punto interrogativo (?) al posto di un singolo carattere. Per esempio, con la chiave di ricerca **cluster** saranno trovati tutti gli oggetti il cui nome contiene la stringa *cluster*. La stringa di ricerca *m0?_** restituirà *M01_cluster.str* e *M02_cluster.str*, ma non *M01a_cluster.str*. La funzione di ricerca non fa distinzione tra caratteri maiuscoli/minuscoli (*cluster* restituirà *Cluster* e anche *CLUSTER*).

Nota: se gli oggetti sono molto numerosi, le ricerche possono richiedere più tempo.

Ricerca in base ad altri criteri

È possibile eseguire una ricerca in base a titolo, etichetta, data, autore, parole chiave, contenuto dell'indice o descrizione. Verranno trovati solo gli oggetti corrispondenti a *tutti* i criteri di ricerca specificati. Per esempio, si potrebbero individuare tutti i flussi contenenti uno o più modelli di cluster che hanno anche un'etichetta specifica e che sono stati modificati dopo una determinata data.

Tipi di oggetto. È possibile limitare la ricerca a modelli, stream, output, nodi, Supernodi, progetti, palette di modelli, scenari o altri tipi di oggetti.

- **Modelli.** È possibile cercare i modelli in base alla categoria (classificazione, approssimazione, raggruppamento cluster, ecc.) o in base a un algoritmo di modellazione specifico quale Kohonen. È possibile inoltre eseguire la ricerca in base ai campi utilizzati, per esempio cercare tutti i modelli che utilizzano un campo denominato *reddito* come campo di input o di output (obiettivo).
- **Flussi.** Per i flussi è possibile limitare la ricerca in base ai campi utilizzati o al tipo di modello (categoria o algoritmo) contenuti nel flusso.

Argomenti. È possibile eseguire la ricerca sui modelli associati ad argomenti specifici da un elenco impostato dagli utenti del repository che dispongono dei privilegi appropriati (per maggiori informazioni, vedere il manuale *Deployment Manager User's Guide*). Per ottenere tale elenco, selezionare questa casella, quindi fare clic sul pulsante *Aggiungi argomenti visualizzato*, selezionare uno o più argomenti dall'elenco e fare clic su *OK*.

Etichetta. Limita la ricerca a specifiche etichette delle versioni degli oggetti.

Date. È possibile specificare una data di creazione o di modifica e cercare gli oggetti creati o modificati prima, dopo o nell'intervallo di date specificato.

Autore. Limita la ricerca agli oggetti creati da un utente specifico.

Parole chiave. Consente di eseguire la ricerca in base alle parole chiave specificate. In IBM SPSS Modeler è possibile specificare le parole chiave nella scheda *Annotazione* di un flusso, modello o oggetto di output.

Descrizione. Consente di cercare termini specifici nel campo della descrizione. In IBM SPSS Modeler, la descrizione di un flusso, un modello o un oggetto di output è specificata nella scheda *Annotazione*. Se sono presenti più stringhe di ricerca, è possibile separarle con un punto e virgola, per esempio *reddito;*

tipo raccolto; valore richiesto. Si noti che, all'interno di una stringa di ricerca, gli spazi sono importanti. Per esempio, tipo di raccolto con uno spazio e tipo di raccolto con due spazi non si equivalgono.

Modifica di oggetti nel repository

È possibile modificare gli oggetti esistenti nel repository direttamente da SPSS Modeler. È possibile:

- Creare, rinominare o eliminare cartelle
- Bloccare o sbloccare oggetti
- Eliminare oggetti

Creazione, ridenominazione ed eliminazione delle cartelle

1. Per eseguire operazioni sulle cartelle nel repository, dal menu principale di SPSS Modeler fare clic su: **Strumenti > Repository > Esplora...**
 - a. Specificare le impostazioni relative alla connessione al repository, se necessario. Consultare l'argomento "Connessione al repository" a pagina 119 per ulteriori informazioni. Per informazioni specifiche sulla porta, la password e altri dettagli per la connessione, rivolgersi all'amministratore del sistema locale.
2. Assicurarsi che la scheda **Cartelle** sia attiva.
3. Per creare una nuova cartella, fare clic con il tasto destro del mouse sulla cartella padre e fare clic su **Nuova cartella**.
4. Per rinominare una cartella, fare clic sulla cartella con il tasto destro del mouse e fare clic su **Rinomina cartella**.
5. Per eliminare una cartella, fare clic sulla cartella con il tasto destro del mouse e fare clic su **Elimina cartella**.

Blocco e sblocco degli oggetti del repository

È possibile bloccare un oggetto per impedire qualsiasi aggiornamento delle sue versioni esistenti o la creazione di nuove versioni. Un oggetto bloccato è contrassegnato dal simbolo del lucchetto sulla relativa icona:



Figura 15. Oggetto bloccato

Per bloccare un oggetto

1. Nella finestra di navigazione del repository, fare clic sull'oggetto richiesto con il tasto destro del mouse.
2. Fare clic su **Blocca**.

Per sbloccare un oggetto

1. Nella finestra di navigazione del repository, fare clic sull'oggetto richiesto con il tasto destro del mouse.
2. Fare clic su **Sblocca**.

Eliminazione di oggetti dal repository

Prima di eliminare un oggetto dal repository occorre decidere se si desidera eliminare tutte le versioni o solo una determinata versione dell'oggetto.

Per eliminare tutte le versioni di un oggetto

1. Nella finestra di navigazione del repository, fare clic sull'oggetto richiesto con il tasto destro del mouse.
2. Fare clic su **Elimina oggetti**.

Per eliminare la versione più recente di un oggetto

1. Nella finestra di navigazione del repository, fare clic sull'oggetto richiesto con il tasto destro del mouse.
2. Fare clic su **Elimina**.

Per eliminare una versione precedente di un oggetto

1. Nella finestra di navigazione del repository, fare clic sull'oggetto richiesto con il tasto destro del mouse.
2. Fare clic su **Elimina versioni**.
3. Selezionare la versione o le versioni da eliminare e fare clic su **OK**.

Gestione delle proprietà degli oggetti di repository

Da SPSS Modeler è possibile controllare diverse proprietà degli oggetti. È possibile:

- Visualizzare le proprietà di una cartella
- Visualizzare e modificare le proprietà di un oggetto
- Creare, applicare ed eliminare le etichette di versione per un oggetto

Visualizzazione delle proprietà delle cartelle

Per visualizzare le proprietà di una cartella nella finestra del repository, fare clic con il tasto destro del mouse sulla cartella richiesta. Fare clic su **Proprietà cartella**.

scheda Generale

Questa scheda visualizza il nome della cartella e le date di creazione e di modifica.

Scheda Autorizzazioni

Questa scheda specifica le autorizzazioni di lettura e scrittura per la cartella. Sono elencati tutti gli utenti e i gruppi che dispongono dell'accesso alla cartella padre. Le autorizzazioni seguono una gerarchia. Per esempio, se non si dispone dell'autorizzazione alla lettura, non si potrà avere l'autorizzazione alla scrittura e se non si dispone dell'autorizzazione alla scrittura non si potrà avere l'autorizzazione all'eliminazione.

Utenti e gruppi. Elenca gli utenti e i gruppi del repository che hanno almeno l'accesso in lettura a questa cartella. Selezionare le caselle di controllo Scrittura ed Eliminazione per aggiungere questi diritti di accesso per la cartella a un utente o un gruppo particolare. Fare clic sull'icona **Aggiungi utenti o gruppi** nella parte destra della scheda Autorizzazioni per assegnare l'accesso a utenti e gruppi aggiuntivi. L'elenco degli utenti e dei gruppi disponibili è gestito dall'amministratore.

Sovrapponi autorizzazioni. Scegliere un'opzione per controllare il modo in cui le modifiche apportate alla cartella corrente si applicano alle rispettive cartelle secondarie, se presenti.

- **Sovrapponi tutte le autorizzazioni.** Sovrappone le autorizzazioni della cartella corrente a tutte le sottocartelle. In questo modo è possibile assegnare in modo rapido le autorizzazioni a più cartelle contemporaneamente. Impostare le autorizzazioni richieste per la cartella padre e quindi sovrapporle nel modo richiesto.
- **Sovrapponi solo modifiche.** Sovrappone solo le modifiche eseguite dall'ultima applicazione di modifiche. Per esempio, se è stato aggiunto un nuovo gruppo e si desidera assegnare a tale gruppo

l'accesso a tutte le cartelle nella filiale Vendite, è possibile assegnare al gruppo l'accesso alla cartella Vendite radice e quindi sovrapporre la modifica a tutte le sottocartelle. Tutte le altre autorizzazioni per le cartelle esistenti rimarranno invariate.

- **Non sovrapporre.** Qualsiasi modifica viene applicata solo alla cartella corrente e non viene sovrapposta alle cartelle figlio.

Visualizzazione e modifica delle proprietà degli oggetti

Nella finestra di dialogo Proprietà oggetto si possono visualizzare e modificare le proprietà. Alcune proprietà non possono essere modificate, ma è sempre possibile aggiornare un oggetto aggiungendo una nuova versione.

1. Nella finestra del repository, fare clic sull'oggetto richiesto con il tasto destro del mouse.
2. Fare clic su **Proprietà oggetto**.

Scheda Generale

Nome. Nome dell'oggetto visualizzato nel repository.

Creato il. Data in cui l'oggetto (non la versione) è stato creato.

Ultima modifica. Data in cui è stata modificata la versione più recente.

Autore. Il nome di login dell'utente.

Descrizione. Per default, la descrizione corrisponde a quella specificata nella scheda Annotazione dell'oggetto in SPSS Modeler.

Argomenti collegati. Nel repository è possibile organizzare per argomenti i modelli e gli oggetti correlati. L'elenco degli argomenti disponibili è impostato dagli utenti del repository che dispongono dei privilegi appropriati (per maggiori informazioni, vedere il manuale *Deployment Manager User's Guide*).

Parole chiave. Le parole chiave sono specificate nella scheda Annotazione di un flusso, un modello o un oggetto di output. Se sono presenti più parole chiave, devono essere separate da spazi. Le parole chiave possono contenere al massimo 255 caratteri. (Se le parole chiave contengono spazi, separarle mediante le virgolette).

Scheda Versioni

Gli oggetti archiviati nel repository possono avere più versioni. Nella scheda Versioni sono visualizzate le informazioni relative a ogni versione.

Per le versioni specifiche di un oggetto memorizzato è possibile definire o modificare le proprietà seguenti:

Versione. Identificatore univoco della versione, creato in base all'ora in cui la versione è stata archiviata.

Etichetta. Etichetta corrente della versione, se esistente. A differenza dell'identificatore di versione, le etichette possono essere spostate da una versione dell'oggetto a un'altra.

Per ogni versione vengono inoltre visualizzati la dimensione, la data di creazione e l'autore del file.

Modifica etichette. Fare clic sull'icona **Modifica etichette** nella parte superiore destra della scheda Versioni per definire, applicare o rimuovere le etichette per gli oggetti memorizzati. Consultare l'argomento "Gestione delle etichette di versione degli oggetti" a pagina 130 per ulteriori informazioni.

Scheda Autorizzazioni

La scheda Autorizzazioni consente di impostare le autorizzazioni di lettura e scrittura per l'oggetto. Sono elencati tutti gli utenti e i gruppi che dispongono dell'accesso all'oggetto corrente. Le autorizzazioni seguono una gerarchia. Per esempio, se non si dispone dell'autorizzazione alla lettura, non si potrà avere l'autorizzazione alla scrittura e se non si dispone dell'autorizzazione alla scrittura non si potrà avere l'autorizzazione all'eliminazione.

Utenti e gruppi. Elenca gli utenti e i gruppi del repository che hanno almeno l'accesso in lettura a questo oggetto. Selezionare le caselle di controllo Scrittura ed Eliminazione per aggiungere questi diritti di accesso per questo oggetto a un utente o un gruppo particolare. Fare clic sull'icona **Aggiungi utenti o gruppi** nella parte destra della scheda Autorizzazioni per assegnare l'accesso a utenti e gruppi aggiuntivi. L'elenco degli utenti e dei gruppi disponibili è gestito dall'amministratore.

Gestione delle etichette di versione degli oggetti

La finestra di dialogo Modifica etichette versione consente di:

- Applicare etichette all'oggetto selezionato
- Rimuovere etichette dall'oggetto selezionato
- Definire una nuova etichetta e applicarla all'oggetto

Per applicare etichette all'oggetto

1. Selezionare una o più etichette nell'elenco **Etichette disponibili**.
2. Fare clic sul pulsante freccia destra per spostare le etichette selezionate nell'elenco **Etichette applicate**.
3. Fare clic su **OK**.

Per rimuovere etichette dall'oggetto

1. Selezionare una o più etichette nell'elenco **Etichette applicate**.
2. Fare clic sul pulsante freccia sinistra per spostare le etichette selezionate nell'elenco **Etichette disponibili**.
3. Fare clic su **OK**.

Per definire una nuova etichetta e applicarla all'oggetto

1. Digitare il nome dell'etichetta nel campo **Nuova etichetta**.
2. Fare clic sul pulsante freccia destra per spostare la nuova etichetta nell'elenco **Etichette applicate**.
3. Fare clic su **OK**.

Distribuzione di flussi

Per consentire l'utilizzo di un flusso con l'applicazione thin client IBM SPSS Modeler Advantage, esso deve essere implementato come flusso (file *.str*) nel repository.

Indipendentemente che la distribuzione di un flusso avvenga come flusso (file *.str*) o come scenario (file *.scn*), l'oggetto può beneficiare al massimo delle funzioni applicabili a livello aziendale di IBM SPSS Collaboration and Deployment Services. Consultare l'argomento "Archiviazione e deployment di oggetti di repository" a pagina 118 per ulteriori informazioni.

Per eseguire la distribuzione del flusso corrente (metodo del menu File)

1. Nel menu principale, fare clic su:
File > Archivia > Distribuisci
2. Scegliere il tipo di distribuzione e compilare il resto della finestra di dialogo in funzione delle esigenze.

3. Fare clic su **Implementa come flusso** per eseguire la distribuzione del flusso da usare con IBM SPSS Modeler Advantage o IBM SPSS Collaboration and Deployment Services. Fare clic su **Implementa come scenario** per eseguire la distribuzione del flusso con IBM SPSS Collaboration and Deployment Services o Applicazioni predittive versione 5.
4. Fare clic su **Archivia**. Per maggiori informazioni, fare clic su **Aiuto**.
5. Continuare da "Completamento del processo di distribuzione".

Per eseguire la distribuzione del flusso corrente (metodo del menu Strumenti)

1. Nel menu principale, fare clic su:
Strumenti > Proprietà flusso > Distribuzione
2. Scegliere il tipo di distribuzione, completare la parte restante della scheda Distribuzione se necessario e fare clic su **Archivia**. Consultare l'argomento "Opzioni di distribuzione dei flussi" per ulteriori informazioni.

Completamento del processo di distribuzione

1. Specificare le impostazioni relative alla connessione al repository, se necessario. Consultare l'argomento "Connessione al repository" a pagina 119 per ulteriori informazioni. Per informazioni specifiche sulla porta, la password e altri dettagli per la connessione, rivolgersi all'amministratore del sistema locale.
2. Nella finestra di dialogo Repository: Archivia, scegliere la cartella in cui archiviare l'oggetto, specificare le eventuali altre informazioni da registrare e fare clic sul pulsante **Archivia**. Consultare l'argomento "Impostazione delle proprietà degli oggetti" a pagina 120 per ulteriori informazioni.

Opzioni di distribuzione dei flussi

La scheda Distribuzione nella finestra di dialogo Opzioni flusso consente di specificare le opzioni per la distribuzione del flusso, che può essere implementato come flusso o come scenario.

Quando si effettua la distribuzione come flusso è possibile aprire e modificare il flusso nell'applicazione thin client IBM SPSS Modeler Advantage. Lo stream viene archiviato nel repository come file con estensione *.str*.

Con la distribuzione come scenario, il flusso viene archiviato nel repository come file con estensione *.scn*. In questo modo, il flusso può essere utilizzato anche da Applicazioni predittive versione 5.

Sia che si esegua la distribuzione come flusso o che come scenario, è possibile sfruttare appieno le funzionalità aggiuntive disponibili con IBM SPSS Collaboration and Deployment Services, per esempio l'accesso a più utenti, il calcolo automatico del punteggio, l'aggiornamento del modello e l'analisi Campione/Sfidante.

Dalla scheda Distribuzione è possibile eseguire un'anteprima della descrizione del flusso creata da IBM SPSS Modeler. Consultare l'argomento "Descrizioni flusso" a pagina 52 per ulteriori informazioni.

Note: per assicurare un accesso regolare ai dati aziendali, un flusso sottoposto a deployment come scenario deve accedere ai propri dati di origine tramite IBM SPSS Collaboration and Deployment Services Enterprise View; in tal caso il flusso deve includere almeno un nodo origine Enterprise View all'interno di ogni ramo designato di calcolo del punteggio o di creazione del modello, a seconda dei casi.

Tipo di distribuzione. Scegliere come si desidera eseguire la distribuzione del flusso. Per poter essere sottoposti a distribuzione, tutti i flussi necessitano di un nodo di calcolo del punteggio designato; gli altri requisiti e opzioni dipendono dal tipo di distribuzione.

- **<nessuno>**. La distribuzione del flusso nel repository non sarà eseguito. Tutte le opzioni sono disattivate tranne l'anteprima descrizione flusso.

- **Solo calcolo del punteggio.** La distribuzione del flusso nel repository avviene quando si fa clic sul pulsante **Archivia**. Il punteggio dei dati si può calcolare mediante il nodo designato nel campo **Nodo Calcolo del punteggio**.
- **Aggiornamento modelli.** Uguale a Solo calcolo del punteggio, ma consente anche di aggiornare il modello nel repository utilizzando gli oggetti specificati nei campi **Nodo Modelli** e **Nugget del modello**. *Nota:* l'aggiornamento automatico del modello non è supportato per default in IBM SPSS Collaboration and Deployment Services, quindi è necessario scegliere questo tipo di distribuzione se si desidera utilizzare questa funzione quando si esegue un flusso dal repository. Consultare l'argomento "Aggiornamento modelli" a pagina 134 per ulteriori informazioni.

Nodo calcolo del punteggio. Selezionare un nodo grafico, di output o di esportazione per identificare il ramo del flusso da utilizzare per il calcolo del punteggio dei dati. Il flusso può contenere un numero qualsiasi di rami, modelli e nodi terminali validi, ma solo un ramo di calcolo del punteggio deve essere designato per la distribuzione. Questo è il requisito di base per la distribuzione di qualsiasi flusso.

Parametri calcolo del punteggio. Consente di specificare i parametri che possono essere modificati quando viene eseguito il ramo di calcolo del punteggio. Consultare l'argomento "Parametri di calcolo del punteggio e di creazione del modello" a pagina 133 per ulteriori informazioni.

Nodo modellazione. Per l'aggiornamento dei modelli, specifica il nodo Modelli utilizzato per rigenerare o aggiornare il modello nel repository. Deve essere un nodo Modelli dello stesso tipo di quello specificato per **Nugget del modello**.

Parametri di creazione del modello. Consente di specificare i parametri che possono essere modificati quando viene eseguito il nodo Modelli. Consultare l'argomento "Parametri di calcolo del punteggio e di creazione del modello" a pagina 133 per ulteriori informazioni.

Nugget del modello. Per l'aggiornamento dei modelli, specifica il nugget del modello da aggiornare o rigenerare ogni volta che viene aggiornato il flusso nel repository (generalmente all'interno di un lavoro pianificato). Il modello deve essere situato sul ramo di calcolo del punteggio. Sebbene un ramo di calcolo del punteggio possa includere più modelli, è possibile specificarne solo uno. Si noti che quando il flusso viene inizialmente creato, il modello potrebbe essere un semplice modello segnaposto che viene aggiornato o rigenerato non appena sono disponibili nuovi dati.

Distribuisci come flusso. Fare clic su questa opzione se si desidera utilizzare il flusso con IBM SPSS Modeler Advantage o IBM SPSS Collaboration and Deployment Services (vedere la nota seguente).

Distribuisci come scenario. Fare clic su questa opzione se si desidera utilizzare il flusso con IBM SPSS Collaboration and Deployment Services o Applicazioni predittive versione 5 (vedere la nota seguente).

Controllo. Fare clic su questo pulsante per controllare se il flusso è valido per la distribuzione. Per poter essere sottoposti a distribuzione, tutti i flussi necessitano di un nodo di calcolo del punteggio designato. Per sottoporre a distribuzione come scenario, il flusso deve inoltre contenere un nodo origine Enterprise View valido. Se queste condizioni non vengono soddisfatte, verranno visualizzati dei messaggi di errore.

Archivia. Sottopone a distribuzione il flusso, se valido. In caso contrario, viene visualizzato un messaggio di errore. Fare clic sul pulsante **Correggi**, correggere l'errore e riprovare.

Anteprima descrizione del flusso. Consente di visualizzare il contenuto della descrizione del flusso creata da IBM SPSS Modeler. Consultare l'argomento "Descrizioni flusso" a pagina 52 per ulteriori informazioni.

Nota: (Distribuisci come flusso o scenario) nel ramo di modellazione è possibile utilizzare più nodi Enterprise View. In questo caso, l'uso di un'unica connessione dati per tutti i nodi Enterprise View all'interno del ramo è preferibile nella maggior parte dei casi e necessario per l'analisi Campione/Sfidante.

- Se il supporto di Campione/Sfidante non è richiesto, è possibile utilizzare diverse connessioni Enterprise View all'interno dello stesso ramo, a condizione che le connessioni varino solo in termini di Definizione provider di dati.
- Queste limitazioni sono applicabili solo all'interno di un dato ramo. Tra i rami di calcolo del punteggio e creazione del modello, è possibile utilizzare diverse connessioni Enterprise View senza tali limitazioni.

Parametri di calcolo del punteggio e di creazione del modello

Quando si effettua la distribuzione di un flusso a IBM SPSS Collaboration and Deployment Services, è possibile scegliere i parametri da visualizzare o modificare ogni volta che il modello viene aggiornato o di cui viene calcolato il punteggio. Per esempio, è possibile specificare valori minimi e massimi o un altro valore soggetto a modifica ogni volta che viene eseguito il lavoro.

1. Per rendere visibile un parametro in modo che possa essere visualizzato o modificato dopo la distribuzione del flusso, selezionarlo dall'elenco nella finestra di dialogo Calcolo del punteggio dei parametri.

L'elenco di parametri disponibili è definito nella scheda Parametri della finestra di dialogo delle proprietà del flusso. Consultare l'argomento "Impostazione di parametri di flusso e di sessione" a pagina 48 per ulteriori informazioni.

Il ramo di calcolo del punteggio

Se si sta eseguendo la distribuzione di un flusso, è necessario designare un ramo del flusso come **ramo di calcolo del punteggio** (ovvero, il ramo che contiene il nodo di calcolo del punteggio). Quando si designa un ramo come ramo di calcolo del punteggio, quel ramo viene evidenziato nell'area del flusso, così come il collegamento del modello al nugget sul ramo di calcolo del punteggio. Questa rappresentazione grafica è particolarmente utile nel caso di flussi complessi che presentano più rami, dove il ramo di calcolo del punteggio potrebbe non essere immediatamente ovvio.

Nota: è possibile designare come ramo di calcolo del punteggio un solo ramo del flusso.

Se per il flusso era già stato definito un ramo di calcolo del punteggio, il nuovo ramo designato sostituisce quello esistente come ramo di calcolo del punteggio. È possibile impostare il colore dell'evidenziazione del ramo di calcolo del punteggio utilizzando un'opzione Colore personalizzato. Consultare l'argomento "Impostazione delle opzioni di visualizzazione" a pagina 152 per ulteriori informazioni.

È possibile mostrare o nascondere l'evidenziazione del ramo di calcolo del punteggio utilizzando il pulsante della barra degli strumenti Mostra/nascondi markup del flusso.



Figura 16. Pulsante della barra degli strumenti Mostra/nascondi markup del flusso

Identificazione del ramo di calcolo del punteggio per la distribuzione

È possibile designare il ramo di calcolo del punteggio dal menu di scelta rapida di un nodo terminale, oppure dal menu Strumenti. Se si utilizza il menu di scelta rapida, il nodo di calcolo del punteggio viene impostato automaticamente nella scheda Distribuzione delle proprietà del flusso.

Per designare un ramo come ramo di calcolo del punteggio (menu di scelta rapida)

1. Collegare il nugget del modello a un nodo terminale (un nodo di elaborazione o di output downstream rispetto all'insieme).
2. Fare clic con il tasto destro del mouse sul nodo terminale.

3. Nel menu, fare clic su **Utilizza come ramo di calcolo del punteggio**.

Per designare un ramo come ramo di calcolo del punteggio (menu Strumenti)

1. Collegare il nugget del modello a un nodo terminale (un nodo di elaborazione o di output downstream rispetto all'insieme).
2. Nel menu principale, fare clic su:
Strumenti > Proprietà flusso > Distribuzione
3. Nell'elenco **Tipo di distribuzione**, fare clic su **Solo calcolo del punteggio** o **Aggiornamento modelli**, in base alle necessità. Consultare l'argomento "Opzioni di distribuzione dei flussi" a pagina 131 per ulteriori informazioni.
4. Fare clic sul campo **Nodo Calcolo del punteggio** e selezionare un nodo terminale dall'elenco.
5. Fare clic su **OK**.

Aggiornamento modelli

L'aggiornamento modelli è il processo di ricostruzione di un modello esistente in un flusso utilizzando dati più recenti. Lo stream in sé e per sé non viene modificato nel repository. Per esempio, il tipo di algoritmo e le impostazioni specifiche del flusso rimangono invariate, ma il modello viene riaddestrato con nuovi dati e aggiornato se la nuova versione funziona meglio della precedente.

È possibile impostare per l'aggiornamento un solo nugget del modello in un flusso. Questa operazione è nota come **aggiornamento modelli**. Se si fa clic sull'opzione **Aggiornamento modelli** nella scheda Distribuzione delle proprietà del flusso (vedere "Opzioni di distribuzione dei flussi" a pagina 131), il nugget del modello designato in quel momento diventa il modello di aggiornamento. È possibile inoltre designare un modello come modello di aggiornamento dal menu di scelta rapida di un nugget del modello. Per fare questo, però, il nugget deve già trovarsi sul ramo di calcolo del punteggio.

Disattivare lo stato di "modello di aggiornamento" di un nugget equivale a impostare il tipo di distribuzione del flusso su Solo calcolo del punteggio, e in questo caso la scheda Distribuzione della finestra di dialogo delle proprietà del flusso verrebbe aggiornata di conseguenza. È possibile attivare e disattivare questo stato mediante l'opzione **Utilizza come modello di aggiornamento** nel menu di scelta rapida del nugget sul ramo di calcolo del punteggio corrente.

La rimozione del collegamento del modello di un nugget sul ramo di calcolo del punteggio rimuove anche lo stato di "modello di aggiornamento" del nugget. La rimozione del collegamento del modello si può annullare mediante il menu Modifica o la barra degli strumenti; in questo modo si ripristina inoltre lo stato di "modello di aggiornamento" del nugget.

Modalità di selezione del modello di aggiornamento

Così come avviene per il ramo di calcolo del punteggio, anche il collegamento al modello di aggiornamento appare evidenziato nel flusso. Il nugget del modello scelto come modello di aggiornamento, e di conseguenza il collegamento evidenziato, dipende dal numero di insiemi presenti nel flusso.

Modello singolo nel flusso

Se un nugget del modello singolo con collegamento si trova sul ramo di calcolo del punteggio quando è identificato come tale, quell'insieme diventa il modello di aggiornamento del flusso.

Modelli multipli nel flusso

Se nel flusso sono presenti più nugget con collegamento, il modello di aggiornamento viene scelto secondo le modalità descritte di seguito.

Se un nugget del modello è stato definito nella scheda Distribuzione della finestra di dialogo delle proprietà del flusso e si trova anche nel flusso, quell'insieme diventa il modello di aggiornamento.

Se nella scheda Distribuzione non è stato definito alcun nugget, oppure c'è un nugget definito ma non si trova sul ramo di calcolo del punteggio, allora il nugget più vicino al nodo terminale diventa il modello di aggiornamento.

Se successivamente si deseleggiano tutti i collegamenti del modello come collegamenti di aggiornamento, viene evidenziato solo il ramo di calcolo del punteggio, non i collegamenti. Il tipo di distribuzione è impostato su Solo calcolo del punteggio.

Nota: è possibile scegliere di impostare uno dei collegamenti sullo stato Sostituisci, ma non l'altro. In questo caso, il nugget del modello scelto come modello di aggiornamento è quello che presenta un collegamento di aggiornamento e che si trova più vicino al nodo terminale quando si designa il ramo di calcolo del punteggio.

Nessun modello nel flusso

Se non ci sono modelli nel flusso, oppure se ci sono solo modelli senza collegamenti, il tipo di distribuzione è impostato su Solo calcolo del punteggio.

Controllo di un ramo di calcolo del punteggio per verificare la presenza di errori

Quando si designa il ramo di calcolo del punteggio, viene eseguito un controllo alla ricerca di eventuali errori, per esempio la mancanza di un nodo Enterprise View quando si esegue la distribuzione del flusso come scenario.

Se viene trovato un errore, il ramo di calcolo del punteggio viene evidenziato con il colore riservato agli errori e viene visualizzato un messaggio. È possibile impostare il colore degli errori utilizzando un'opzione Colore personalizzato. Consultare l'argomento "Impostazione delle opzioni di visualizzazione" a pagina 152 per ulteriori informazioni.

Se viene trovato un errore, procedere come segue:

1. Correggere l'errore in base al contenuto del messaggio di errore.
2. Eseguire una delle seguenti operazioni:
 - Fare clic con il pulsante destro del mouse sul nodo terminale e fare clic su **Controlla scenario** nel menu di scelta rapida.
 - Nel menu principale, fare clic su:
Strumenti > Proprietà flusso > Distribuzione
e fare clic su **Seleziona**.
3. Se necessario, ripetere questo processo finché non viene più rilevato alcun errore.

Capitolo 10. Esportazione in applicazioni esterne

Informazioni sull'esportazione in applicazioni esterne

IBM SPSS Modeler consente di esportare l'intero processo di data mining in applicazioni esterne mediante diversi meccanismi, in modo che il lavoro svolto per la preparazione dei dati e la creazione dei modelli possa essere sfruttato anche fuori dall'ambiente IBM SPSS Modeler.

La precedente sezione ha illustrato come eseguire la distribuzione dei flussi in un repository IBM SPSS Collaboration and Deployment Services per sfruttarne la possibilità di accesso da parte di più utenti, di pianificazione dei lavori e altre funzioni. In modo analogo è possibile utilizzare i flussi di IBM SPSS Modeler con:

- IBM SPSS Modeler Advantage
- Applicazioni in grado di importare ed esportare file in formato PMML

Per ulteriori informazioni sull'utilizzo dei flussi con IBM SPSS Modeler Advantage, vedere "Apertura di un flusso in IBM SPSS Modeler Advantage".

Per informazioni sull'esportazione e importazione di modelli come file PMML per avere la possibilità di condividerli con qualsiasi altra applicazione in grado di supportare tale formato, vedere "Importazione ed esportazione di modelli come PMML" a pagina 138.

Apertura di un flusso in IBM SPSS Modeler Advantage

I flussi di IBM SPSS Modeler si possono utilizzare con l'applicazione thin client IBM SPSS Modeler Advantage. Benché sia possibile creare applicazioni personalizzate interamente all'interno di IBM SPSS Modeler Advantage, è possibile anche utilizzare un flusso già creato in IBM SPSS Modeler come base per il flusso di lavoro di un'applicazione.

Per aprire un flusso IBM SPSS Modeler Advantage:

1. Eseguire la distribuzione del flusso nel repository IBM SPSS Collaboration and Deployment Services facendo clic sull'opzione **Implementa come flusso**. Consultare l'argomento "Distribuzione di flussi" a pagina 130 per ulteriori informazioni.
2. Fare clic sul pulsante Apri in IBM SPSS Modeler Advantage nella barra degli strumenti o nel menu principale, fare clic su:

File > Apri in IBM SPSS Modeler Advantage

1. Specificare le impostazioni relative alla connessione al repository, se necessario. Consultare l'argomento "Connessione al repository" a pagina 119 per ulteriori informazioni. Per informazioni specifiche sulla porta, la password e altri dettagli per la connessione, rivolgersi all'amministratore del sistema locale.

Nota: il software IBM SPSS Modeler Advantage deve essere installato anche sul server del repository.

1. Nella finestra di dialogo Repository: Archivia, scegliere la cartella in cui archiviare l'oggetto, specificare le eventuali altre informazioni da registrare e fare clic sul pulsante **Archivia**. Consultare l'argomento "Impostazione delle proprietà degli oggetti" a pagina 120 per ulteriori informazioni.

In questo modo viene avviato IBM SPSS Modeler Advantage con il flusso già aperto. Lo stream viene chiuso in IBM SPSS Modeler.

Importazione ed esportazione di modelli come PMML

PMML, acronimo di Predictive Model Markup Language, è un formato XML per descrivere modelli statistici e di data mining, inclusi input ai modelli, trasformazioni utilizzate per la preparazione dei dati per il data mining e parametri che definiscono i modelli stessi. In IBM SPSS Modeler è possibile importare ed esportare PMML, rendendo possibile la condivisione di modelli con altre applicazioni che supportano questo formato, quali IBM SPSS Statistics.

Per ulteriori informazioni sul formato PMML, vedere il sito Web del gruppo di data mining all'indirizzo <http://www.dmg.org>.

Per esportare un modello:

L'esportazione in formato PMML è supportata per la maggior parte dei tipi di modelli generati in IBM SPSS Modeler. Consultare l'argomento "Tipi di modello che supportano il formato PMML" per ulteriori informazioni.

1. Fare clic con il pulsante destro del mouse su un nugget del modello nella palette Modelli. In alternativa, fare doppio clic su un nugget del modello nell'area e selezionare il menu File.
2. Nel menu, fare clic su **Esporta PMML**.
3. Nella finestra di dialogo Esporta (o Salva), specificare una directory di destinazione e un nome univoco per il modello.

Nota: è possibile modificare opzioni per l'esportazione in formato PMML nella finestra di dialogo Opzioni utente. Nel menu principale, fare clic su:

Strumenti > Opzioni > Opzioni utente

e selezionare la scheda PMML.

Consultare l'argomento "Impostazione delle opzioni di esportazione PMML" a pagina 153 per ulteriori informazioni.

Per importare un modello salvato come PMML

I modelli esportati come PMML da IBM SPSS Modeler o un'altra applicazione possono essere importati nella palette Modelli. Consultare l'argomento "Tipi di modello che supportano il formato PMML" per ulteriori informazioni.

1. Nella palette Modelli, fare clic con il pulsante destro del mouse sulla palette e selezionare **Importa PMML** dal menu.
2. Selezionare il file da importare e specificare le opzioni per le etichette di variabile in base alle proprie esigenze.
3. Fare clic su **Apri**.

Utilizza etichette variabili se presenti nel modello. Nel formato PMML è possibile specificare nomi ed etichette di variabile (per esempio ID referente per *RefID*) per le variabili contenute nel dizionario dati. Selezionare questa opzione per utilizzare le etichette di variabile, se presenti nel PMML esportato.

Se sono state selezionate le opzioni delle etichette di variabile ma non sono presenti etichette di variabile nel PMML, i nomi delle variabili verranno utilizzati normalmente.

Tipi di modello che supportano il formato PMML

Esportazione PMML

Modelli IBM SPSS Modeler. I seguenti modelli creati in IBM SPSS Modeler possono essere esportati come PMML 4.0:

- C&R Tree
- QUEST
- CHAID
- Regressione lineare
- Rete neurale
- C5.0
- Regressione logistica
- Genlin
- SVM
- Apriori
- Carma
- Medie K
- Kohonen
- TwoStep
- GLMM (il supporto è solo per modelli GLMM solo ad effetti fissi)
- Elenco di decisioni
- Cox
- Sequenza (calcolo del punteggio per modelli PLMM Sequenza non supportato)
- Modello Statistics

Modelli nativi di database. Per i modelli generati utilizzando algoritmi nativi di database, l'esportazione PMML è disponibile solo per i modelli IBM InfoSphere Warehouse. I modelli creati utilizzando Analysis Services di Microsoft o Oracle Data Miner non possono essere esportati. Si noti inoltre che i modelli IBM esportati come PMML non possono essere reimportati in IBM SPSS Modeler.

Importazione PMML

IBM SPSS Modeler può importare e calcolare il punteggio di modelli PMML generati dalle versioni correnti di tutti i prodotti IBM SPSS Statistics, compresi i modelli esportati da IBM SPSS Modeler nonché i modelli o le trasformazioni PMML generate da IBM SPSS Statistics 17.0 o versioni successive. In sostanza, sono importabili tutti i modelli PMML di cui il modulo di gestione del punteggio è in grado di calcolare il punteggio, con le seguenti eccezioni:

- Non è possibile importare modelli Apriori, CARMA, Rilevamento anomalie e Sequenza.
- Non è possibile sfogliare i modelli PMML dopo averli importati in IBM SPSS Modeler, anche se è possibile utilizzarli per il calcolo del punteggio. Questa eccezione riguarda anche i modelli che sono stati originariamente esportati da IBM SPSS Modeler. Per evitare tale limitazione, esportare il modello come file di modello generato (*.gm) anziché come PMML.
- I modelli IBM InfoSphere Warehouse esportati come PMML non possono essere importati.
- Si è verificata una convalida limitata all'importazione, ma la convalida completa viene eseguita al tentativo di calcolare il punteggio del modello. Pertanto è possibile che l'importazione riesca e che, invece, il calcolo del punteggio fallisca o produca risultati non corretti.

Capitolo 11. Progetti e report

Introduzione ai progetti

Un **progetto** è un gruppo di file correlati a un'attività di data mining e include i flussi di dati, i grafici, i modelli generati, i report e qualsiasi altro elemento creato in IBM SPSS Modeler. I progetti di IBM SPSS Modeler non forniscono semplicemente un metodo di organizzazione dell'output, ma anche numerose altre funzionalità. Grazie ai progetti, è possibile:

- Annotare ogni oggetto nel file di progetto.
- Utilizzare il modello CRISP-DM per l'elaborazione del data mining in base alle specifiche esigenze aziendali. I progetti includono anche un sistema di guida di CRISP-DM che fornisce informazioni dettagliate ed esempi pratici relativi al data mining con CRISP-DM.
- Aggiungere al progetto oggetti esterni a IBM SPSS Modeler, per esempio una presentazione di PowerPoint che illustra gli obiettivi del data mining o white paper relativi agli algoritmi che si prevede di utilizzare.
- Creare report completi o semplici report di aggiornamento basati sulle proprie annotazioni. I report possono essere creati in formato HTML, in modo da poter essere facilmente pubblicati nella rete Intranet dell'organizzazione.

Nota: se il riquadro dei progetti non è visibile nella finestra IBM SPSS Modeler, fare clic su **Progetto** nel menu Visualizza.

Gli oggetti aggiunti a un progetto possono essere visualizzati in due modi: mediante **visualizzazione Classi** e **visualizzazione CRISP-DM**. Tutti gli elementi aggiunti a un progetto vengono inseriti in entrambe le visualizzazioni ed è possibile passare da una visualizzazione all'altra per ottenere l'organizzazione ottimale.

Visualizzazione CRISP-DM

I progetti di IBM SPSS Modeler supportano lo standard CRISP-DM (Cross-Industry Standard Process for Data Mining) e consentono pertanto di organizzare l'elaborazione del data mining aziendale mediante un metodo consolidato del settore e non proprietario. Il modello CRISP-DM prevede sei fasi dell'elaborazione, dalla fase iniziale (raccolta dei requisiti di business) a quella finale (deployment dei risultati). Alcune fasi non prevedono in genere l'utilizzo di IBM SPSS Modeler, tuttavia il riquadro dei progetti le include tutte e sei allo scopo di fornire una posizione centralizzata per l'archiviazione e la registrazione di tutti i materiali associati al progetto. Per esempio, la fase Business Understanding è in genere incentrata sulla raccolta dei requisiti di business e sulla definizione degli obiettivi insieme ai colleghi, anziché sull'utilizzo dei dati in IBM SPSS Modeler. Il riquadro dei progetti consente di archiviare le note delle riunioni nella cartella *Business Understanding*, in modo da poterle riutilizzare e inserire nei report in momenti successivi.

Nella visualizzazione di CRISP-DM del riquadro dei progetti è disponibile un sistema di guida che fornisce informazioni sull'intero ciclo di vita del data mining. Da IBM SPSS Modeler è possibile accedere a questa Guida in linea facendo clic su **Guida in linea di CRISP-DM** nel menu Aiuto.

Nota: se il riquadro dei progetti non è visibile nella finestra, fare clic su **Progetto** nel menu Visualizza.

Impostazione della fase di default del progetto

Gli oggetti aggiunti a un progetto vengono inseriti in una fase di default di CRISP-DM. Ciò implica che è necessario organizzare manualmente gli oggetti in base alla fase del data mining in cui vengono utilizzati. È consigliabile impostare come cartella di default la fase di lavoro corrente.

Per selezionare la fase di default da utilizzare:

1. Nella visualizzazione CRISP-DM, fare clic con il pulsante destro del mouse sulla cartella corrispondente alla fase di default che si desidera impostare.
2. Fare clic su **Imposta come default** nel menu.

La cartella di default verrà visualizzata in grassetto.

Visualizzazione Classi

Nella visualizzazione Classi del riquadro dei progetti, le operazioni di IBM SPSS Modeler sono organizzate per categorie in base al tipo di oggetti creati. È possibile inserire gli oggetti salvati nelle seguenti categorie:

- Flussi
- Nodi
- Modelli
- Tabelle, grafici, report
- Altro (file non di IBM SPSS Modeler, per esempio presentazioni o white paper relativi ai processi di data mining)

Gli oggetti aggiunti nella visualizzazione Classi vengono inseriti anche nella cartella della fase di default della visualizzazione CRISP-DM.

Nota: se il riquadro dei progetti non è visibile nella finestra, fare clic su **Progetto** nel menu Visualizza.

Creazione di un progetto

Un progetto non è altro che un file contenente i riferimenti a tutti i file associati al progetto. Gli elementi di un progetto vengono pertanto salvati sia singolarmente, sia come un riferimento nel file di progetto (.cpj). Questa struttura di riferimenti comporta quanto segue:

- Gli elementi di un progetto devono essere salvati singolarmente prima di essere aggiunti a un progetto. In caso contrario, verrà richiesto di salvarli prima di aggiungerli al progetto corrente.
- Gli oggetti aggiornati singolarmente, per esempio i flussi, vengono aggiornati anche nel file di progetto.
- Se si spostano o si eliminano manualmente gli oggetti (per esempio i flussi, i nodi e gli oggetti di output) dal file system, i collegamenti nel file di progetto non saranno più validi.

Creazione di un nuovo progetto

La procedura di creazione di un nuovo progetto nella finestra di IBM SPSS Modeler è molto semplice. È possibile iniziare subito la creazione se non è aperto alcun progetto oppure chiudere il progetto corrente e iniziare da un progetto vuoto.

Nel menu principale, fare clic su:

File > Progetti > Nuovo progetto...

Aggiunta di oggetti a un progetto

Dopo avere creato o aperto un progetto, è possibile aggiungere oggetti quali stream di dati, nodi e report utilizzando diversi metodi.

Aggiunta di oggetti dallo strumento Manager

È possibile aggiungere stream o output mediante lo strumento Manager disponibile nella parte in alto a destra della finestra di IBM SPSS Modeler.

1. Selezionare un oggetto, per esempio una tabella o un flusso, da una delle schede della finestra dei manager.
2. Fare clic con il pulsante destro del mouse e selezionare **Aggiungi a progetto**.
Se l'oggetto è stato salvato in precedenza, verrà aggiunto automaticamente alla cartella degli oggetti appropriata (nella visualizzazione Classi) o alla cartella della fase di default (nella visualizzazione CRISP-DM).
3. In alternativa, è possibile trascinare gli oggetti dai manager al riquadro dei progetti.

Nota: è possibile che venga richiesto di salvare prima l'oggetto. Durante il salvataggio, assicurarsi che l'opzione **Aggiungi file a progetto** della finestra di dialogo Salva sia selezionata. In questo modo, l'oggetto verrà aggiunto automaticamente al progetto dopo il salvataggio.

Aggiunta di nodi dall'area

Tramite la finestra di dialogo Salva, è possibile aggiungere i singoli nodi dall'area del flusso.

1. Selezionare un nodo nell'area.
2. Fare clic con il pulsante destro del mouse e selezionare **Salva nodo**. In alternativa, nel menu principale, fare clic su:
Modifica > Nodo > Salva nodo...
3. Nella finestra di dialogo Salva selezionare **Aggiungi file a progetto**.
4. Specificare un nome per il nodo e fare clic su **Salva**.

Il file verrà salvato e aggiunto al progetto. I nodi vengono aggiunti nella cartella *Nodi* della visualizzazione Classi e nella cartella della fase di default nella visualizzazione CRISP-DM.

Aggiunta di file esterni

È possibile aggiungere a un progetto un'ampia gamma di oggetti esterni a IBM SPSS Modeler, che possono risultare utili per la gestione dell'intero processo di data mining all'interno di IBM SPSS Modeler. Per esempio, è possibile archiviare in un progetto collegamenti a dati, note, presentazioni e grafici. Nella visualizzazione CRISP-DM è possibile aggiungere i file esterni nella cartella desiderata, mentre nella visualizzazione Classi questi possono essere salvati solo nella cartella *Altro*.

Per aggiungere file esterni a un progetto:

1. Trascinare i file dal desktop nel progetto.
o
2. Con il pulsante destro del mouse, fare clic sulla cartella di destinazione nella visualizzazione CRISP-DM o Classi.
3. Fare clic su **Aggiungi a cartella** nel menu di scelta rapida.
4. Selezionare un file nella finestra di dialogo e fare clic su **Apri**.

Nei progetti di IBM SPSS Modeler verrà aggiunto un riferimento all'oggetto selezionato.

Trasferimento di progetti in IBM SPSS Collaboration and Deployment Services Repository

È possibile trasferire un intero progetto (compresi tutti i file dei componenti) nel IBM SPSS Collaboration and Deployment Services Repository in una sola operazione. Gli oggetti che sono già archiviati nella cartella di destinazione non saranno spostati. Questa funzionalità opera anche in direzione contraria, ovvero consente di trasferire interi progetti da IBM SPSS Collaboration and Deployment Services Repository al file system locale.

Nota: per l'accesso ad un repository IBM SPSS Collaboration and Deployment Services è richiesta una licenza separata. Per ulteriori informazioni, consultare <http://www.ibm.com/software/analytics/spss/products/deployment/cds/>

Trasferimento di un progetto

Verificare che il progetto da trasferire sia aperto nel riquadro dei progetti.

Per trasferire un progetto:

1. Con il pulsante destro del mouse, fare clic sulla cartella dei progetti principale e selezionare **Trasferisci progetto**.
2. Se richiesto, accedere a IBM SPSS Collaboration and Deployment Services Repository.
3. Specificare la nuova cartella per il progetto e fare clic su **OK**.

Impostazione delle proprietà del progetto

È possibile personalizzare il contenuto e la documentazione di un progetto nella finestra di dialogo Proprietà progetto. Per accedere alle proprietà del progetto:

1. Fare clic con il pulsante destro del mouse su un oggetto o su una cartella nel riquadro dei progetti e selezionare **Proprietà progetto**.
2. Fare clic sulla scheda **Progetto** per inserire le informazioni di base del progetto.

Creato. Mostra la data di creazione del progetto (non modificabile).

Riepilogo. È possibile creare un riepilogo del progetto di data mining, che verrà visualizzato nel report del progetto.

Contenuti. Elenca il tipo e il numero dei componenti a cui viene fatto riferimento nel file di progetto (non modificabile).

Salva oggetti non salvati come. Specifica se gli oggetti non salvati devono essere salvati nel file system locale o archiviati nel repository. Consultare l'argomento "Informazioni su IBM SPSS Collaboration and Deployment Services Repository" a pagina 117 per ulteriori informazioni.

Aggiorna i riferimenti a oggetti quando si carica il progetto. Selezionare questa opzione per aggiornare i riferimenti del progetto ai relativi componenti. *Nota:* i file aggiunti a un progetto non vengono salvati nel file di progetto, bensì vengono archiviati nel progetto sotto forma di riferimenti ai file e pertanto lo spostamento o l'eliminazione di un file comporterà la rimozione del relativo oggetto dal progetto.

Annotazione di un progetto

Il riquadro dei progetti offre diversi metodi per l'annotazione dei processi di data mining. Le annotazioni a livello di progetto consentono di tenere traccia degli obiettivi e delle scelte generali, mentre le annotazioni a livello di cartella o di nodo forniscono informazioni più dettagliate. Nella scheda Annotazioni è possibile inserire informazioni a livello di progetto, per esempio l'esclusione dei dati con dati mancanti irrecuperabili oppure le ipotesi formulate durante l'esplorazione dati.

Per annotare un progetto:

1. Selezionare la cartella del progetto nella visualizzazione CRISP-DM o Classi.
2. Fare clic sulla cartella con il pulsante destro del mouse e selezionare **Proprietà progetto**.
3. Fare clic sulla scheda **Annotazioni**.
4. Digitare le parole chiave e il testo per descrivere il progetto.

Proprietà cartella e Annotazioni

È possibile annotare le singole cartelle del progetto, sia nella visualizzazione CRISP-DM che nella visualizzazione Classi. Le annotazioni all'interno della visualizzazione CRISP-DM consentono di documentare in modo estremamente efficace gli obiettivi dell'organizzazione relativi a ogni fase del data mining. Per esempio, mediante lo strumento di annotazione è possibile inserire per la cartella *Business Understanding* un testo del tipo "L'obiettivo di business di questo studio è quello di ridurre il tasso di abbandono dei clienti più importanti". Il testo potrà quindi essere aggiunto automaticamente al report di progetto selezionando l'opzione **Includi nel report**.

Per annotare una cartella:

1. Selezionare una cartella nel riquadro dei progetti.
2. Fare clic sulla cartella con il pulsante destro del mouse e selezionare **Proprietà cartella**.

Nella visualizzazione CRISP-DM, le annotazioni delle cartelle includono un riepilogo dell'obiettivo associato a ogni fase e indicazioni sull'esecuzione delle principali attività del data mining. Le annotazioni possono essere rimosse o modificate.

Nome. In quest'area è visualizzato il nome del campo selezionato.

Testo di suggerimento. Crea i testi descrittivi personalizzati che saranno visualizzati quando si posiziona il puntatore del mouse su una cartella del progetto. Questi testi sono utili nella visualizzazione CRISP-DM, per esempio se si desidera fornire una breve panoramica degli obiettivi di ogni fase o contrassegnare lo stato di una fase come "In corso" o "Completata".

Campo annotazione. In questo campo è possibile inserire annotazioni più lunghe che possono essere raccolte nel report del progetto. Le annotazioni della visualizzazione CRISP-DM includono una descrizione di ogni fase del data mining, che può tuttavia essere personalizzata in base alle specifiche esigenze del progetto.

Includi nel report. Per includere l'annotazione nei report, selezionare l'opzione **Includi nel report**.

Proprietà oggetto

È possibile visualizzare le proprietà degli oggetti e decidere se includere i singoli oggetti nel report del progetto. Per accedere alle proprietà dell'oggetto:

1. Fare clic con il pulsante destro del mouse su un oggetto nel riquadro dei progetti.
2. Fare clic su **Proprietà oggetto** nel menu.

Nome. Visualizza il nome dell'oggetto salvato.

Percorso. Visualizza la posizione dell'oggetto salvato.

Includi nel report. Selezionare questa opzione per inserire le informazioni sull'oggetto in un report generato.

Chiusura di un progetto

Quando si esce da IBM SPSS Modeler o si apre un nuovo progetto, il file di progetto esistente (.cpj) viene chiuso.

Alcuni file associati al progetto, quali flussi, nodi o grafici, potrebbero essere ancora aperti. Se si desidera lasciare questi file aperti, rispondere No al messaggio ... **Si desidera salvare e chiudere questi file?**

Se dopo avere chiuso un progetto si salva o si modifica un file associato, la versione aggiornata del file verrà inserita nel progetto alla successiva apertura. Per evitare che ciò accada, rimuovere il file dal progetto o salvarlo con un nome di file diverso.

Generazione di report

La possibilità di creare report basati sugli elementi e sulle annotazioni del progetto è una delle funzioni più utili dei progetti e rappresenta un componente essenziale per l'efficacia del data mining, come illustrato nella metodologia CRISP-DM. Il report può essere generato direttamente in uno dei vari formati di file o in una finestra di output sullo schermo per consentire la visualizzazione immediata. Il report può essere quindi stampato, salvato o visualizzato in un browser Web. I report salvati possono essere distribuiti ad altri utenti della propria organizzazione.

Accade di frequente che i report vengano generati più volte dai file di progetto durante il processo di data mining, allo scopo di distribuirli alle persone coinvolte nel progetto. I report includono le informazioni sugli oggetti ai quali viene fatto riferimento nel file di progetto e le eventuali annotazioni create ed è possibile generarli in base alla visualizzazione Classi o CRISP-DM.

Per generare un report:

1. Selezionare la cartella del progetto nella visualizzazione CRISP-DM o Classi.
2. Fare clic sulla cartella con il pulsante destro del mouse e selezionare **Report progetto**.
3. Impostare le opzioni per il report e fare clic su **Genera report**.

Le opzioni presenti nella finestra di dialogo del report consentono di generare il tipo di report desiderato in diversi modi:

Nome output. Indicare il nome della finestra di output se si decide di inviare l'output del report allo schermo. È possibile specificare un nome personalizzato o lasciare che IBM SPSS Modeler assegni automaticamente un nome alla finestra.

Output su schermo. Selezionare questa opzione per generare il report e visualizzarlo in una finestra di output. Si noti che dalla finestra di output è possibile esportare il report in vari formati di file.

Output su file. Selezionare questa opzione per generare il report e salvarlo come file del tipo specificato nell'elenco Tipo di file.

Nome file. Specificare un nome di file per il report generato. I file vengono salvati per default nella directory `\bin` di IBM SPSS Modeler. Utilizzare il pulsante con i puntini di sospensione (...) per specificare una diversa posizione.

Tipo di file. I tipi di file disponibili sono:

- **Documento HTML.** Il report viene salvato in un unico file HTML. Se contiene dei grafici, questi vengono salvati sotto forma di file PNG a cui il file HTML fa riferimento. Quando si pubblica il report su Internet, accertarsi di avere caricato sia il file HTML che le immagini a cui fa riferimento.
- **Documento di testo.** Il report viene salvato in un unico file di testo. Se contiene dei grafici, nel report vengono inclusi solo dei riferimenti al nome del file e al percorso.
- **Documento Microsoft Word.** Il report viene salvato come un singolo documento, con gli eventuali grafici incorporati direttamente nel documento.
- **Documento Microsoft Excel.** Il report viene salvato come un singolo foglio di calcolo, con gli eventuali grafici incorporati direttamente nel foglio.
- **Documento Microsoft PowerPoint.** Ogni fase viene mostrata su una nuova diapositiva. Gli eventuali grafici sono incorporati direttamente nelle diapositive di PowerPoint.
- **Oggetto di output.** Quando viene aperto in IBM SPSS Modeler, questo file (`.cou`) genera una finestra di output identica a quella generata con l'opzione **Output su schermo** nel **Formato del report**.

Nota: per esportare in un file di Microsoft Office, è necessario che l'applicazione corrispondente sia installata.

Titolo. Indicare un titolo per il report.

Struttura report. Selezionare **CRISP-DM** o **Classi**. La visualizzazione CRISP-DM fornisce un report di stato che contiene informazioni sintetiche di carattere generale, nonché dettagli su ogni fase del data mining. La visualizzazione Classi è basata su oggetti ed è più appropriata per la registrazione interna di dati e stream.

Autore. Viene visualizzato il nome utente di default, che può tuttavia essere modificato.

Il report include. Selezionare un metodo per includere gli oggetti nel report. Selezionare **tutte le cartelle e tutti gli oggetti** per inserire tutti gli elementi aggiunti al file di progetto. È inoltre possibile inserire gli elementi in base alla selezione di **Includi nel report** nelle proprietà degli oggetti. In alternativa, per verificare gli elementi non inseriti nel report, specificare l'inserimento unicamente degli elementi contrassegnati per l'esclusione (nei quali **Includi nel report** non è selezionato).

Seleziona. Questa opzione consente di aggiornare il progetto selezionando solo **elementi recenti** per il report oppure di tenere traccia dei vecchi problemi a volte non risolti impostando i parametri per **elementi vecchi**. Selezionare **tutti gli elementi** per non considerare il parametro del tempo per il report.

Ordina per. È possibile selezionare una combinazione delle seguenti caratteristiche degli oggetti per ordinarli all'interno di una cartella:

- **Tipo.** Raggruppa gli oggetti in base al tipo.
- **Nome.** Dispone gli oggetti in ordine alfabetico.
- **Data aggiunta.** Ordina gli oggetti in base alla data in cui sono stati aggiunti al progetto.

Salvataggio ed esportazione dei report generati

I report generati sullo schermo vengono visualizzati in una nuova finestra di output. Gli eventuali grafici inclusi nel report vengono visualizzati come immagini in linea.

Terminologia utilizzata nei report

Nel report è riportato il numero totale dei nodi presenti in ogni stream. I numeri sono elencati sotto le seguenti intestazioni, che utilizzano la terminologia di IBM SPSS Modeler anziché quella di CRISP-DM:

- **Lettori di dati.** Nodi origine.
- **Scrittori di dati.** Nodi di esportazione.
- **Builder modello.** Nodi Modelli o di creazione.
- **Applicatore del modello.** Modelli generati, detti anche nugget.
- **Builder di output.** Nodi Grafico o di output.
- **Altro.** Tutti gli altri nodi relativi al progetto. Per esempio, quelli disponibili nella scheda Oper su campi o Oper su record nella palette nodi.

Per salvare un report:

1. Nel menu File, fare clic su **Salva**.
2. Digitare un nome per il file.
Il report viene salvato come oggetto di output.

Per esportare un report:

3. Nel menu File, fare clic su **Esporta** e sul tipo di file per l'esportazione.
4. Digitare un nome per il file.

Il report viene salvato nel formato prescelto.

I tipi di file per l'esportazione sono i seguenti:

- HTML
- Testo
- Microsoft Word
- MS Excel
- MS PowerPoint

Nota: per esportare in un file di Microsoft Office, è necessario che l'applicazione corrispondente sia installata.

I pulsanti nella parte superiore della finestra consentono di:

- Stampare il report.
- Visualizzare il report in formato HTML in un browser Web esterno.

Capitolo 12. Personalizzazione di IBM SPSS Modeler

Personalizzazione delle opzioni di IBM SPSS Modeler

È possibile eseguire diverse operazioni per personalizzare IBM SPSS Modeler in base alle esigenze. La personalizzazione consiste soprattutto nell'impostazione di specifiche opzioni utente, quali l'allocazione di memoria, le directory di default e l'utilizzo di suono e colore. È inoltre possibile personalizzare la palette dei nodi situata nella parte inferiore della finestra di IBM SPSS Modeler.

Impostazione delle opzioni di IBM SPSS Modeler

Le opzioni per IBM SPSS Modeler possono essere personalizzate e impostate in vari modi:

- Per impostare le opzioni di sistema, per esempio l'utilizzo della memoria e le opzioni internazionali, fare clic su **Opzioni di sistema** nel menu **Strumenti > Opzioni**.
- Per impostare le opzioni utente, per esempio i caratteri e i colori visualizzati, fare clic su **Opzioni utente** nel menu **Strumenti > Opzioni**.
- Per specificare la posizione delle applicazioni che interagiscono con IBM SPSS Modeler, scegliere **Applicazioni di supporto** dal menu **Strumenti > Opzioni**
- Per specificare le directory di default utilizzate in IBM SPSS Modeler, fare clic su **Imposta directory o Imposta directory server** nel menu File.

?È inoltre possibile impostare opzioni valide per tutti i flussi o alcuni di essi. Consultare l'argomento "Impostazione delle opzioni per i flussi" a pagina 41 per ulteriori informazioni.

Opzioni di sistema

Per specificare la lingua preferita o le impostazioni internazionali per IBM SPSS Modeler, fare clic su **Opzioni di sistema** nel menu **Strumenti > Opzioni**. In questa finestra è inoltre possibile impostare la memoria massima utilizzabile da IBM SPSS Modeler. Si noti che le modifiche apportate in questa finestra di dialogo diventeranno effettive solo al successivo riavvio di IBM SPSS Modeler.

Memoria massima. Selezionare questa opzione per stabilire un limite, in megabyte, per l'utilizzo della memoria da parte di IBM SPSS Modeler. In alcune piattaforme, IBM SPSS Modeler limita le dimensioni delle elaborazioni per ridurre l'utilizzo di risorse in computer con risorse limitate o forti carichi di lavoro. Se si stanno elaborando grandi quantità di dati, potrebbe derivarne un errore di "memoria esaurita". Per migliorare il carico di memoria, specificare una nuova soglia.

Utilizzare locale di sistema. Questa opzione è selezionata per default e impostata su inglese (Stati Uniti). Deselezionarla per specificare un'altra lingua dall'elenco delle lingue e delle impostazioni internazionali disponibili.

Gestione della memoria

Oltre all'impostazione **Memoria massima** specificata nella finestra di dialogo Opzioni di sistema, sono disponibili svariate soluzioni per ottimizzare l'utilizzo della memoria:

- Configurare l'opzione **Numero di membri massimo per campi nominali** nella finestra di dialogo Proprietà stream. Questa opzione consente di specificare il numero massimo di membri per i campi nominali dopo il quale il livello di misurazione diventa *Senza tipo*. Consultare l'argomento "Impostazione delle opzioni generali per i flussi" a pagina 42 per ulteriori informazioni.
- Liberare memoria facendo clic nell'angolo inferiore destro della finestra di IBM SPSS Modeler, nel punto in cui vengono visualizzate la memoria utilizzata da IBM SPSS Modeler e la quantità di memoria allocata (xxMB/xxMB). Quando si fa clic al suo interno, tale area assume un colore più scuro,

quindi i valori relativi all'allocazione diminuiscono. Quando il colore dell'area torna normale, in IBM SPSS Modeler è stata liberata tutta la memoria disponibile.

Impostazione delle directory di default

Per specificare la directory di default utilizzata per i browser dei file e l'output, scegliere **Imposta directory** o **Imposta directory server** dal menu File.

- **Imposta directory.** Utilizzare questa opzione per impostare la directory che viene utilizzata. La directory di default in uso dipende dal percorso di installazione della versione di IBM SPSS Modeler o dal percorso della riga di comando utilizzato per avviare IBM SPSS Modeler. In modalità locale, la directory di lavoro corrisponde al percorso utilizzato per tutte le operazioni e i file di output lato client (se ad essi viene fatto riferimento con percorsi relativi).
- **Imposta directory server.** L'opzione Imposta directory server, del menu File, viene attivata a ogni connessione del server remoto. Utilizzarla per specificare la directory di default per tutti i file del server e i file di dati specificati per l'input o l'output. La directory di default del server è *\$CLEO/data*, dove *\$CLEO* è la directory di installazione della versione Server di IBM SPSS Modeler. Tramite la riga di comando, è possibile ignorare questa impostazione di default utilizzando il flag `-server_directory` con l'argomento della riga di comando `modelerclient`.

Impostazione delle opzioni utente

Per impostare le opzioni generali per IBM SPSS Modeler, scegliere **Opzioni utente** dal menu **Strumenti > Opzioni**. Queste opzioni si applicano a tutti i flussi utilizzati in IBM SPSS Modeler.

Per impostare i seguenti tipi di opzioni, fare clic sulla scheda corrispondente:

- **Notifiche**, per impostare per esempio opzioni relative alla sovrascrittura di modelli e ai messaggi di errore.
- **Visualizza**, per impostare per esempio opzioni relative ai colori di sfondo e dei grafici.
- **Opzioni di visualizzazione del colore della sintassi.**
- **Opzioni di esportazione PMML** utilizzate per l'esportazione di modelli in PMML (Predictive Model Markup Language).
- **Informazioni su autore o utente**, quali nome, iniziali e indirizzo di posta elettronica. Queste informazioni possono essere visualizzate nella scheda Annotazioni per i nodi e gli altri oggetti che vengono creati.

Per impostare opzioni specifiche per i flussi, per esempio i separatori decimali, i formati di data e ora, l'ottimizzazione, il layout e gli script dei flussi, utilizzare la finestra di dialogo Proprietà stream disponibile dai menu File e Strumenti.

Impostazione delle opzioni di notifica

La scheda Notifiche della finestra di dialogo Opzioni utente consente di impostare varie opzioni relative all'occorrenza e al tipo di avvisi e finestre di conferma in IBM SPSS Modeler. È anche possibile specificare il comportamento delle schede Output e Modelli nel riquadro dei manager, al momento della generazione di nuovi output e modelli.

Mostra finestra di dialogo con feedback di esecuzione del flusso. Selezionare questa opzione per visualizzare una finestra di dialogo con un indicatore di avanzamento dopo tre secondi dall'inizio dell'esecuzione di un flusso. La finestra di dialogo comprende anche i dettagli degli oggetti di output creati dal flusso.

- **Chiudi finestra di dialogo al completamento.** Per default, la finestra di dialogo viene chiusa al termine dell'esecuzione del flusso. Deselezionare questa casella di controllo per continuare a visualizzare la finestra di dialogo una volta terminato il flusso.

Avvisa quando un nodo sovrascrive un file. Selezionare questa opzione se si desidera che venga visualizzato un messaggio di errore quando le operazioni dei nodi sovrascrivono un file esistente.

Avvisa quando un nodo sovrascrive una tabella database. Selezionare questa opzione se si desidera che venga visualizzato un messaggio di errore quando le operazioni dei nodi sovrascrivono una tabella di database esistente.

Notifica acustica

Mediante l'elenco, specificare se utilizzare i suoni per notificare all'utente un evento o un errore. Sono disponibili numerosi suoni. Utilizzare il pulsante Esegui (altoparlante) per riprodurre il suono selezionato. Utilizzare il pulsante con i puntini di sospensione (...) per ricercare e selezionare un suono.

Nota: i file *.wav* utilizzati per creare i suoni in IBM SPSS Modeler sono memorizzati nella directory */media/sounds* dell'installazione.

- **Disattiva tutti i suoni.** Selezionare questa opzione per disattivare la notifica acustica per tutti gli eventi.

Notifica visiva

Le opzioni riportate in questo gruppo sono utilizzate per specificare il comportamento delle schede Output e Modelli del riquadro dei manager nella parte superiore destra dello schermo quando vengono generati nuovi elementi. Selezionare **Nuovo modello** o **Nuovo output** dall'elenco per specificare il comportamento della scheda corrispondente.

Per **Nuovo modello** sono disponibili le seguenti opzioni:

Aggiungi modello al flusso. Se selezionata (impostazione predefinita), aggiunge un nuovo modello al flusso e alla scheda Modelli non appena viene creato il modello. Nel flusso, il modello compare con un collegamento al nodo Modelli a partire dal quale è stato creato. Se si deseleziona questa casella, il modello viene aggiunto solo alla scheda Modelli.

Sostituisci modello precedente. Se selezionata (impostazione predefinita), sovrascrive un modello esistente del flusso nella scheda Modelli e nell'area del flusso. Se la casella è deselezionata, il modello viene aggiunto ai modelli esistenti sulla scheda e nell'area. Si noti che questa impostazione viene sovrascritta dall'impostazione di sostituzione del modello nel collegamento a un modello.

Per **Nuovo output** sono disponibili le seguenti opzioni:

Avvisa quando gli output superano [n]. Selezionare se visualizzare un avviso quando il numero di elementi nella scheda Output supera una quantità predefinita. La quantità di default è 20, che può essere modificata se necessario.

Le seguenti opzioni sono disponibili in tutti i casi:

Seleziona scheda. Scegliere se passare alla scheda Output o Modelli quando l'oggetto corrispondente viene generato durante l'esecuzione del flusso.

- Selezionare **Sempre** per passare alla scheda corrispondente nel riquadro dei manager.
- Selezionare **Se generato dal flusso corrente** per attivare la scheda corrispondente per gli oggetti generati dal flusso attualmente visibile nell'area.
- Selezionare **Mai** per impedire che il software passi alla scheda corrispondente per notificare gli output o i modelli generati.

Scheda lampeggiante. Selezionare se far lampeggiare la scheda Output o Modelli nel riquadro dei manager ogni volta che vengono generati nuovi modelli o output.

- Selezionare **Se non selezionato** per far lampeggiare la scheda corrispondente (se non è già selezionata) ogni volta che nel riquadro dei manager vengono generati nuovi oggetti.

- Selezionare **Mai** per impedire che il software faccia lampeggiare la scheda corrispondente per notificare l'utente in merito alla generazione di oggetti.

Scorri tavolozza per renderla visibile (solo Nuovo modello). Specificare se far scorrere automaticamente la scheda Modelli nel riquadro dei manager per rendere visibile il modello più recente.

- Selezionare **Sempre** per attivare lo scorrimento.
- Selezionare **Se generato dal flusso corrente** per far scorrere solo gli oggetti generati dal flusso attualmente visibile nell'area.
- Selezionare **Mai** per impedire che il software faccia scorrere automaticamente la scheda Modelli.

Apri finestra (solo Nuovo output). Selezionare se aprire automaticamente una finestra di output alla generazione.

- Selezionare **Sempre** per aprire sempre una nuova finestra di output.
- Selezionare **Se generato dal flusso corrente** per aprire una nuova finestra per l'output generato dal flusso attualmente visibile nell'area.
- Selezionare **Mai** per impedire che il software apra automaticamente nuove finestre per l'output generato.

Fare clic su **Valori di default** per ripristinare le impostazioni predefinite del sistema relative a questa scheda.

Impostazione delle opzioni di visualizzazione

Tramite la scheda Visualizza della finestra di dialogo Opzioni utente, è possibile impostare le opzioni di visualizzazione dei caratteri e dei colori in IBM SPSS Modeler.

Mostra finestra di dialogo di benvenuto all'avvio. Selezionare per visualizzare la finestra di dialogo di benvenuto all'avvio. Tale finestra contiene le opzioni per avviare l'esercitazione Esempi di applicazioni, aprire un flusso di esempio o un flusso o progetto esistente oppure creare un nuovo flusso.

Mostra markup flusso e supernodo. Selezionare per far sì che l'eventuale markup sui flussi e i Supernodi venga visualizzato per default. Il markup comprende i commenti ai flussi, i collegamenti a modelli e l'evidenziazione dei rami di calcolo del punteggio.

Caratteri e colori standard (effettivi al riavvio). Le opzioni presenti in questa casella di controllo sono utilizzate per specificare le caratteristiche dello schermo e lo schema di colori di IBM SPSS Modeler nonché la dimensione dei caratteri visualizzati. Le opzioni selezionate verranno applicate solo previa chiusura e riavvio di IBM SPSS Modeler.

- **Aspetto.** Consente di scegliere uno schema di colori e caratteristiche dello schermo standard. È possibile scegliere fra:
SPSS Standard (default), una struttura comune a tutti i prodotti IBM SPSS.
SPSS Classic, una struttura ben nota agli utenti delle precedenti versioni di IBM SPSS Modeler.
Windows, un layout Windows che può essere utile per aumentare il contrasto delle palette e dell'area del flusso.
- **Dimensione carattere di default per i nodi.** Specificare la dimensione del carattere da utilizzare nelle palette dei nodi e per i nodi visualizzati nell'area del flusso.

Nota: è possibile impostare le dimensioni delle icone dei nodi di un flusso nel riquadro Layout della scheda Opzioni della finestra di dialogo delle proprietà del flusso. Dal menu principale, scegliere

Strumenti > Proprietà flusso > Opzioni > Layout.

Colori personalizzati. Questa tabella elenca i colori selezionati utilizzati per diversi elementi di visualizzazione. Per ciascun elemento elencato nella tabella, è possibile cambiare il colore corrente

facendo doppio clic sulla riga corrispondente nella colonna **Colore** e selezionando un colore dall'elenco. Per specificare un colore personalizzato, scorrere l'elenco fino in fondo e fare clic sulla voce **Colore...**

Ordine colori categorie grafici. In questa tabella sono riportati i colori correntemente selezionati e utilizzati per la visualizzazione nei grafici appena creati. L'ordine dei colori riflette l'ordine in cui verranno utilizzati nel grafico. Se, per esempio, un campo nominale utilizzato come sovrapposizione colore contiene quattro valori univoci, verranno utilizzati solo i primi quattro colori elencati qui. Per ciascun elemento elencato nella tabella, è possibile cambiare il colore corrente facendo doppio clic sulla riga corrispondente nella colonna **Colore** e selezionando un colore dall'elenco. Per specificare un colore personalizzato, scorrere l'elenco fino in fondo e fare clic sulla voce **Colore...** Le modifiche apportate non influenzano i grafici creati in precedenza.

Fare clic su **Valori di default** per ripristinare le impostazioni predefinite del sistema relative a questa scheda.

Impostazione delle opzioni di visualizzazione della sintassi

Tramite la scheda Sintassi della finestra di dialogo Opzioni utente, è possibile impostare le opzioni degli attributi dei caratteri e dei colori degli script creati in IBM SPSS Modeler.

Evidenziazione sintassi. Questa tabella elenca i colori attualmente selezionati, utilizzati per diversi elementi della sintassi, compresi il carattere e la finestra in cui vengono visualizzati. Per ciascun elemento elencato nella tabella, è possibile cambiare il colore facendo clic sull'elenco a discesa corrispondente nella riga e selezionando un colore dall'elenco. Inoltre, per gli elementi del carattere, è possibile selezionare di aggiungere un enfasi in grassetto o corsivo.

Anteprima. Questa tabella mostra un esempio di sintassi che utilizza gli attributi dei colori e dei caratteri selezionati nella tabella **Evidenziazione sintassi**. Questa anteprima si aggiorna appena si modifica una selezione.

Fare clic su **Valori di default** per ripristinare le impostazioni predefinite del sistema relative a questa scheda.

Impostazione delle opzioni di esportazione PMML

Dalla scheda PMML è possibile controllare le modalità di esportazione dei modelli in PMML (Predictive Model Markup Language) di IBM SPSS Modeler. Consultare l'argomento "Importazione ed esportazione di modelli come PMML" a pagina 138 per ulteriori informazioni.

Esporta PMML. Qui è possibile configurare le variazioni di PMML che funzionano meglio con l'applicazione di destinazione.

- Selezionare **Con estensioni** per consentire le estensioni PMML nei casi speciali in cui non esiste un equivalente PMML standard. Si noti che nella maggior parte dei casi questo produrrà gli stessi risultati di PMML standard.
- Selezionare **Come PMML standard...** per esportare PMML nella modalità più vicina possibile allo standard PMML.

Opzioni PMML Standard. Se viene selezionata l'opzione **Come PMML V3.2 standard**, è possibile scegliere uno dei due metodi validi per esportare modelli di regressione lineare e logistica:

- Come **Modelli <Regressione generale> PMML**
- Come **Modelli <Regressione> PMML**

Per ulteriori informazioni sul formato PMML, vedere il sito Web del gruppo di data mining all'indirizzo <http://www.dmg.org>.

Impostazione delle informazioni sull'utente

Informazioni su utente/autore. Le informazioni immesse qui possono essere visualizzate nella scheda Annotazioni dei nodi e degli altri oggetti creati.

Personalizzazione della palette dei nodi

I flussi vengono creati mediante i nodi. La palette dei nodi disponibile nella parte inferiore della finestra di IBM SPSS Modeler contiene tutti i nodi utilizzabili per la creazione dei flussi. Consultare l'argomento "Palette nodi" a pagina 16 per ulteriori informazioni.

La palette dei nodi può essere riorganizzata in due modi:

- Con la personalizzazione del Manager palette. Consultare l'argomento "Personalizzazione del Manager palette" per ulteriori informazioni.
- Con la modifica del modo in cui le schede della palette che contengono sottopalette vengono visualizzate nella palette dei nodi. Consultare l'argomento "Creazione di una sottopalette" a pagina 156 per ulteriori informazioni.

Personalizzazione del Manager palette

È possibile personalizzare il Manager palette per soddisfare esigenze specifiche di utilizzo di IBM SPSS Modeler. Per esempio, se si analizzano di frequente dati di serie temporali di un database, è possibile decidere di tenere contemporaneamente disponibili in un'unica scheda della palette il nodo origine del database, il nodo Intervalli di tempo, il nodo Serie temporali e il nodo Grafico tempo. Il Manager palette consente di effettuare con facilità queste modifiche, creando schede della palette personalizzate nella palette dei nodi.

Il Manager palette consente di effettuare varie attività:

- Controllare quali schede della tavolozza vengono visualizzate nella tavolozza dei nodi sotto l'area.
- Controllare l'ordine in cui le schede della tavolozza vengono visualizzate nella palette dei nodi.
- Creare e modificare le proprie schede della tavolozza e le eventuali sotto-tavolozze associate.
- Modificare le selezioni di nodi di default nella scheda Preferiti.

Per accedere a Gestore tavolozze, nel menu Strumenti, fare clic su **Gestisci tavolozze**.

Nome tavolozza. Viene elencata ogni scheda della palette disponibile, visualizzata o meno nella palette dei nodi, incluse le schede della palette create dall'utente. Consultare l'argomento "Creazione di una scheda della palette" a pagina 155 per ulteriori informazioni.

N. di nodi. Il numero di nodi visualizzato in ogni scheda della palette. Un numero elevato indica che potrebbe essere più utile creare sotto-tavolozze per suddividere i nodi sulla scheda. Consultare l'argomento "Creazione di una sottopalette" a pagina 156 per ulteriori informazioni.

Visibile?. Selezionare questo campo per visualizzare la scheda della palette nella palette dei nodi. Consultare l'argomento "Visualizzazione delle schede della palette nella palette dei nodi" a pagina 155 per ulteriori informazioni.

Sotto- tavolozza. Per selezionare le sotto-tavolozza da visualizzare su una scheda della tavolozza, evidenziare il **Nome palette** richiesto e fare clic su questo pulsante per visualizzare la finestra di dialogo Sotto-tavolozza. Consultare l'argomento "Creazione di una sottopalette" a pagina 156 per ulteriori informazioni.

Ripristina default. Per rimuovere completamente tutte le modifiche e aggiunte effettuate a palette e sottopalette e ripristinare le impostazioni di default delle palette, fare clic su questo pulsante.

Creazione di una scheda della palette

Per creare una scheda della palette personalizzata:

1. Scegliere Manager palette dal menu Strumenti.
2. A destra della colonna *Visibile?* fare clic sul pulsante Aggiungi palette a progetto. Viene visualizzata la finestra di dialogo Crea/Modifica palette.
3. Digitare un **Nome palette** univoco.
4. Nell'area **Nodi disponibili**, selezionare il nodo da aggiungere alla scheda della palette.
5. Fare clic sul pulsante con la freccia the destra di aggiunta nodo per spostare il nodo evidenziato nell'area **Nodi selezionati**. Ripetere la procedura fino ad avere aggiunto tutti i nodi desiderati.
Al termine dell'aggiunta, è possibile modificare l'ordine in cui i nodi selezionati appaiono nella scheda della palette:
6. Utilizzare i pulsanti con le frecce per spostare un nodo verso l'alto o verso il basso di una riga.
7. Utilizzare i tasti direzionali per spostare un nodo all'inizio o alla fine dell'elenco.
8. Per rimuovere un nodo da una palette, evidenziarlo e fare clic sul pulsante Elimina a destra dell'area **Nodi selezionati**.

Visualizzazione delle schede della palette nella palette dei nodi

Potrebbero essere disponibili delle opzioni di IBM SPSS Modeler che non vengono mai utilizzate; in tal caso, è possibile utilizzare il Manager palette per nascondere le schede che contengono questi nodi.

Per selezionare le schede da visualizzare nella palette dei nodi:

1. Scegliere Manager palette dal menu Strumenti.
2. Utilizzare le caselle di controllo disponibili nella colonna *Visibile?* per scegliere se includere o nascondere ogni scheda della palette.

Per rimuovere in modo permanente una scheda della palette dalla palette dei nodi, evidenziarla e fare clic sul pulsante Elimina a destra della colonna *Visibile?*. Una volta eliminata, una scheda della palette non può essere recuperata.

Nota: non è possibile eliminare le schede della palette di default fornite con IBM SPSS Modeler, fatta eccezione per la scheda Preferiti.

Modifica dell'ordine di visualizzazione nella palette dei nodi

Dopo avere selezionato le schede della palette da visualizzare, è possibile modificare l'ordine in cui vengono visualizzate nella palette dei nodi:

1. Utilizzare i pulsanti con le frecce per spostare una scheda della palette verso l'alto o verso il basso di una riga. Lo spostamento in alto corrisponde alla sinistra della palette dei nodi e viceversa.
2. Utilizzare i tasti direzionali per spostare una scheda della palette all'inizio o alla fine dell'elenco. Le schede all'inizio dell'elenco appariranno a sinistra nella palette dei nodi.

Visualizzazione delle sottopalette su una scheda della palette

Così come è possibile controllare quali schede della palette vengono visualizzate nella palette dei nodi, è possibile controllare quali sottopalette sono disponibili per la relativa scheda della palette padre.

Per selezionare le sottopalette da visualizzare in una scheda della palette:

1. Scegliere Manager palette dal menu Strumenti.
2. Selezionare la palette desiderata.
3. Fare clic sul pulsante Sottopalette. Viene visualizzata la finestra di dialogo Sottopalette.
4. Utilizzare le caselle di controllo disponibili nella colonna *Visibile?* per scegliere se includere o meno ogni sottopalette nella scheda della palette. La sottopalette **Tutte** viene sempre visualizzata e non è possibile eliminarla.

5. Per rimuovere in modo permanente una sottopalette dalla scheda della palette, evidenziare la sottopalette e fare clic sul pulsante Elimina a destra della colonna *Visibile?*.

Nota: non è possibile eliminare le sottopalette di default fornite con la scheda della palette Modelli.

Modifica dell'ordine di visualizzazione nella scheda della palette

Dopo avere selezionato le sottopalette da visualizzare, è possibile modificare l'ordine in cui vengono visualizzate nella scheda della palette padre:

1. Utilizzare i pulsanti con le frecce semplici per spostare una sottopalette verso l'alto o verso il basso di una riga.
2. Utilizzare i tasti direzionali per spostare una palette secondaria all'inizio o alla fine dell'elenco.

Le sottopalette create vengono visualizzate nella palette dei nodi quando si seleziona la relativa scheda della palette padre. Consultare l'argomento "Modifica della visualizzazione di una scheda della palette" per ulteriori informazioni.

Creazione di una sottopalette

Poiché è possibile aggiungere tutti i nodi esistenti alle schede della palette personalizzate create, potrebbe accadere di selezionare un numero di nodi tale da richiedere lo scorrimento dello schermo per visualizzarli tutti. Per evitare lo scorrimento, è possibile creare delle sottopalette in cui collocare i nodi scelti per la scheda della palette. Per esempio, se si è creata una scheda della palette contenente i nodi utilizzati più frequentemente per creare i flussi, è possibile creare quattro sottopalette che suddividono le selezioni per nodo origine, Operazioni su campi, Modelli e Output.

Nota: è possibile selezionare nodi per le sottopalette solo tra i nodi aggiunti alla scheda della tavolozza padre.

Per creare una sottopalette:

1. Scegliere Manager palette dal menu Strumenti.
2. Selezionare la palette a cui si desidera aggiungere le sottopalette.
3. Fare clic sul pulsante Sottopalette. Viene visualizzata la finestra di dialogo Sottopalette.
4. A destra della colonna *Visibile?* fare clic sul pulsante Aggiungi sottopalette. Viene visualizzata la finestra di dialogo Crea/Modifica sottopalette.
5. Digitare un **Nome sottopalette** univoco.
6. Nell'area **Nodi disponibili**, selezionare il nodo da aggiungere alla sottopalette.
7. Fare clic sul pulsante con la freccia a destra di aggiunta nodo per spostare il nodo selezionato nell'area **Nodi selezionati**.
8. Al termine, scegliere **OK** per tornare alla finestra di dialogo Sottopalette.

Le sottopalette create vengono visualizzate nella palette dei nodi quando si seleziona la relativa scheda della palette padre. Consultare l'argomento "Modifica della visualizzazione di una scheda della palette" per ulteriori informazioni.

Modifica della visualizzazione di una scheda della palette

A causa dell'elevato numero di nodi disponibili in IBM SPSS Modeler, questi potrebbero non essere tutti visibili su schermi di piccole dimensioni se non scorrendo a destra o sinistra la palette dei nodi, particolarmente nel caso della scheda della palette Modelli. Per ridurre la necessità di scorrimento, è possibile scegliere di visualizzare solo i nodi contenuti in una sottopalette, se disponibile. Consultare l'argomento "Creazione di una sottopalette" per ulteriori informazioni.

Per cambiare i nodi visualizzati in una scheda della palette, selezionare la scheda e, dal menu a sinistra, selezionare se visualizzare tutti i nodi o solo i nodi contenuti in una specifica sottopalette.

Gestione dei nodi CEMI

CEMI non è attualmente più supportato ed è stato sostituito da CLEF, che offre una serie di funzioni molto più versatili e semplici da usare. Per ulteriori informazioni, consultare *IBM SPSS Modeler 16 CLEF Developer's Guide* presente in questa release.

Capitolo 13. Considerazioni sulle performance dei flussi e dei nodi

È possibile progettare i flussi per ottimizzare le performance disponendo i nodi nella configurazione più efficiente e attivandone la cache quando è necessario, nonché prestando attenzione alle altre considerazioni esposte in questa sezione.

A parte le considerazioni illustrate in questa sede, in genere è possibile ottenere ulteriori e sensibili miglioramenti delle performance utilizzando il database in modo efficiente, in particolare mediante l'ottimizzazione SQL.

Ordine dei nodi

Anche quando non si fa uso dell'ottimizzazione SQL, l'ordine dei nodi in un flusso può incidere sulle prestazioni. L'obiettivo generale è ridurre al minimo l'elaborazione a valle; pertanto, i nodi che riducono la quantità di dati devono essere collocati nella parte iniziale del flusso. IBM SPSS Modeler Server può applicare automaticamente alcune regole di riordinamento in fase di compilazione per spostare in avanti alcuni nodi quando si ha la garanzia che l'operazione è priva di rischi. Questa funzione è attivata di default. Contattare l'amministratore di sistema per verificare che sia attivata nella propria installazione di Clementine.

Quando si utilizza l'ottimizzazione SQL, è consigliabile migliorarne la disponibilità e l'efficienza. Poiché l'ottimizzazione si interrompe quando il flusso contiene un'operazione che non può essere eseguita nel database, è preferibile raggruppare le operazioni ottimizzate da SQL nella parte iniziale del flusso. Questa strategia mantiene una parte più consistente dell'elaborazione nel database, trasferendo una mole di dati inferiore a IBM SPSS Modeler.

Le seguenti operazioni possono essere eseguite nella maggior parte dei database. Tentare di raggrupparle all'*inizio* del flusso:

- Unione per chiave (unione)
- Select
- Aggrega
- Ordina
- Campione
- Accoda
- Operazioni distinte in modalità *includi*, in cui vengono selezionati tutti i campi
- Operazioni di riempimento
- Operazioni di derivazione elementari mediante calcoli aritmetici standard o manipolazione delle stringhe (a seconda delle operazioni supportate dal database)
- Creazione di flag

Le seguenti operazioni non possono essere eseguite nella maggior parte dei database e devono quindi essere collocate all'interno del flusso *dopo* le operazioni riportate nell'elenco precedente:

- Operazioni su dati non contenuti nel database, per esempio file flat
- Unione per ordine
- Balance
- Operazioni distinte in modalità *scarta* o in cui solo un sottoinsieme di campi è selezionato come distinto

- Tutte le operazioni che richiedono di accedere ai dati di record diversi da quello in corso di elaborazione
 - Derivazioni dei campi stato e conteggio
 - Operazioni del nodo Cronologia
 - Operazioni con le funzioni "@" (di serie temporali)
 - Modalità di verifica del tipo *Avvisa e Termina*
 - Costruzione, applicazione e analisi di modelli
- Nota:* i modelli di strutture ad albero delle decisioni, insiemi di regole, regressione lineare e generati da fattori possono generare stringhe SQL e possono quindi essere spostati nuovamente nel database.
- Output di dati a qualunque destinazione diversa dal database che sta elaborando i dati

Cache dei nodi

Per ottimizzare l'esecuzione del flusso, è possibile impostare una **cache** in qualsiasi nodo diverso da un nodo terminale. La cache configurata per un nodo viene riempita con i dati che passano attraverso il nodo alla successiva esecuzione del flusso di dati. Da questo momento in poi, i dati vengono letti dalla cache, che è archiviata in una directory temporanea, anziché dalla sorgente dati.

La memorizzazione dei dati nella cache è utile soprattutto dopo un'operazione che richiede tempo, per esempio un ordinamento, un'unione o un'aggregazione. Si supponga, per esempio, che esista un nodo origine impostato in modo da leggere i dati relativi alle vendite da un database e un nodo Aggregazione che riepiloga le vendite in base alla posizione. È possibile impostare una cache nel nodo Aggregazione anziché nel nodo origine, in modo tale da memorizzarvi i dati aggregati anziché l'intero insieme di dati.

Nota: l'impostazione di una cache nei nodi origine, che consente semplicemente di archiviare una copia dei dati originali non appena vengono letti in IBM SPSS Modeler, nella maggior parte dei casi non migliorerà le prestazioni.

I nodi con la cache attivata sono indicati da una piccola icona di documento nell'angolo in alto a destra. Quando i dati vengono memorizzati nella cache del nodo, l'icona del documento diventa di colore verde.

Per attivare una cache:

1. Nell'area del flusso, fare clic con il pulsante destro del mouse sul nodo e fare clic su **Cache** nel menu.
2. Fare clic su **Attiva** nel sottomenu.
3. Per disattivare la cache, fare clic con il pulsante destro del mouse sul nodo e fare clic su **Disattiva** nel sottomenu.

Memorizzazione nella cache di nodi nel database

Per i flussi eseguiti nel database, i dati possono essere memorizzati nella cache midstream in una tabella temporanea nel database anziché nel file system. Insieme all'ottimizzazione SQL, ciò consente di ottenere significativi vantaggi in termini di prestazioni. Per esempio, l'output di un flusso che unisce più tabelle per creare una visualizzazione data mining può essere memorizzato nella cache e riutilizzato in un secondo tempo in base alle esigenze. Generando automaticamente codice SQL per tutti i nodi downstream, è possibile migliorare ulteriormente le prestazioni.

Se si memorizzano dati nella cache del database con stringhe di lunghezza superiore a 255 caratteri, assicurarsi che esista un nodo tipologia a monte del nodo di cache e che i valori di campo vengano letti oppure impostare la lunghezza della stringa utilizzando il parametro `default_sql_string_length` nel file `options.cfg`. In questo modo la colonna corrispondente nella tabella temporanea viene impostata sulla larghezza corretta per adattarsi alle stringhe.

Per sfruttare la memorizzazione nella cache del database, è necessario che siano attivate l'ottimizzazione SQL e la memorizzazione nella cache del database. Si noti che le impostazioni di ottimizzazione del Server prevalgono su quelle del Client. Consultare l'argomento "Impostazione delle opzioni di ottimizzazione per i flussi" a pagina 44 per ulteriori informazioni.

Con la memorizzazione nella cache del database attivata, è sufficiente fare clic con il pulsante destro del mouse su qualsiasi nodo non terminale per memorizzare i dati nella cache in quel punto e la cache verrà creata automaticamente direttamente nel database alla successiva esecuzione del flusso. Se la memorizzazione nella cache del database o l'ottimizzazione SQL non sono attivate, la cache verrà scritta nel file system.

Nota: i seguenti database supportano tabelle temporanee per la memorizzazione di dati nella cache: DB2, Netezza, Oracle, SQL Server e and Teradata. Altri database utilizzano una tabella normale per la memorizzazione di dati nella cache. È possibile personalizzare il codice SQL per database specifici - contattare l'assistenza.

Performance: nodi di elaborazione

Ordina. Il nodo Ordina deve leggere tutto l'insieme di dati di input prima di poterlo ordinare. I dati vengono salvati in memoria fino a un certo limite e i dati eccedenti vengono riversati su disco. L'algoritmo di ordinamento è di tipo combinato: i dati vengono letti e inseriti in memoria fino al limite e ordinati con un algoritmo di ordinamento rapido di tipo ibrido. Se i dati rientrano tutti nella memoria, l'ordinamento è terminato. In caso contrario viene applicato un algoritmo di ordinamento di unione. I dati ordinati vengono scritti su file e il successivo blocco di dati viene letto e inserito in memoria, ordinato e scritto su disco. L'operazione viene ripetuta fino a quando tutti i dati sono stati letti; i blocchi ordinati vengono quindi uniti. L'unione può richiedere ripetuti passaggi sui dati archiviati su disco. Al culmine dell'elaborazione, il nodo Ordina avrà due copie complete su disco del dataset, una ordinata e una non ordinata.

Il tempo di esecuzione complessivo dell'algoritmo è nell'ordine di $N \cdot \log(N)$, in cui N è il numero di record. L'ordinamento in memoria è più veloce dell'unione da disco: pertanto, il tempo di esecuzione effettivo può essere ridotto allocando una maggiore quantità di memoria all'ordinamento. L'algoritmo alloca a sé stesso una porzione della RAM fisica controllata dall'opzione di configurazione di IBM SPSS Modeler Server *Moltiplicatore utilizzo memoria*. Per aumentare la memoria utilizzata per l'ordinamento, espandere la RAM fisica o incrementare il valore di questa opzione. Si noti che, quando la porzione di memoria utilizzata supera la quantità allocata per l'esecuzione del processo provocando il paging di parte della memoria su disco, si ha una diminuzione delle prestazioni perché l'algoritmo di ordinamento accede alla memoria secondo uno schema diretto che può determinare un paging eccessivo. L'algoritmo di ordinamento è utilizzato da vari nodi oltre al nodo Ordina, ma per tutti valgono le stesse regole riguardo alle prestazioni.

Raccolta. Il nodo Raccolta legge tutto l'insieme dei dati di input per calcolare i limiti dei raccoglitori prima di assegnare i record ai vari raccoglitori. Durante il calcolo dei limiti l'insieme di dati viene memorizzato nella cache e in seguito viene letto nuovamente per l'allocazione. Quando il metodo di raccolta è a *larghezza fissa* o *media+deviazione standard*, l'insieme di dati viene memorizzato direttamente su disco. Questi metodi hanno un tempo di esecuzione lineare e richiedono spazio su disco sufficiente per salvare l'intero insieme di dati. Quando il metodo di raccolta è per *ranghi* o *blocchi*, l'insieme di dati viene ordinato con l'algoritmo di ordinamento descritto in precedenza e, una volta ordinato, viene utilizzato come cache. A causa dell'ordinamento, il tempo di esecuzione di questi metodi è pari a $M \cdot N \cdot \log(N)$, dove M è il numero di campi nei raccoglitori e N è il numero dei record; lo spazio su disco necessario è pari al doppio delle dimensioni dell'insieme di dati.

Generando un nodo Ricava in base ai raccoglitori generati è possibile migliorare le performance nei passaggi successivi. Le operazioni di derivazione sono molto più rapide della raccolta.

Unione per chiave (unione). Il nodo Unione, quando il metodo di unione è per *chiave* (equivalente all'unione di un database), ordina ogni insieme di dati di input in base ai campi chiave. Questa parte della procedura ha un tempo di esecuzione di $M*N*\log(N)$, dove M è il numero degli input e N è il numero di record nell'input più grande; l'operazione richiede una quantità di spazio su disco sufficiente per salvare tutti gli insiemi di dati di input più una seconda copia dell'insieme di dati più grande. Il tempo di esecuzione dell'operazione di unione è proporzionale alle dimensioni dell'insieme dei dati di output, che dipende dalla frequenza della corrispondenza delle chiavi. Nel peggiore dei casi, in cui l'output è il prodotto Cartesiano degli input, il tempo di esecuzione può quasi avvicinarsi a NM . Si tratta di un'eventualità rara: quasi tutte le operazioni di unione hanno un numero molto inferiore di corrispondenze delle chiavi. Se un insieme di dati è relativamente più grande degli altri, o se i dati in ingresso sono già ordinati in base a un campo chiave, è possibile migliorare le prestazioni di questo nodo mediante la scheda Ottimizzazione.

Aggregazione. Quando l'opzione *Chiavi contigue* non è impostata, questo nodo legge (senza memorizzare) tutto l'insieme dei dati di input prima di generare l'output aggregato. Nei casi più estremi, in cui le dimensioni dei dati aggregati raggiungono un limite (determinato dall'opzione di configurazione *Moltiplicatore utilizzo memoria* di IBM SPSS Modeler Server), il resto dell'insieme di dati viene ordinato ed elaborato come se l'opzione *Chiavi contigue* fosse impostata. Quando questa opzione è impostata non vi è alcun salvataggio dei dati, perché i record di output aggregati vengono generati durante la lettura dei dati di input.

Elimina duplicati. Il nodo Elimina duplicati memorizza tutti i campi chiave univoci dell'insieme dei dati di input; nei casi in cui tutti i campi sono campi chiave e tutti i record sono univoci, il nodo memorizza tutto l'insieme di dati. Per default, il nodo Elimina duplicati ordina i dati nei campi chiave e quindi seleziona (o scarta) il primo record distinto di ogni gruppo. Per gli insiemi di dati più piccoli con un numero ridotto di chiavi distinte o per quelli che sono stati già ordinati, è possibile selezionare delle opzioni per migliorare la velocità e l'efficienza dell'elaborazione.

Tipo. In alcuni casi, il nodo Tipo memorizza i dati di input nella cache durante la lettura dei valori; la cache viene utilizzata per l'elaborazione a valle. Benché la cache richieda spazio su disco sufficiente per archiviare tutto l'insieme di dati, il suo utilizzo riduce i tempi di elaborazione.

Valutazione. Il nodo Valutazione deve ordinare i dati di input per calcolare i blocchi. L'operazione di ordinamento viene ripetuta per tutti i modelli valutati poiché i punteggi e l'ordine dei record che ne deriva sono diversi caso per caso. Il tempo di esecuzione è pari a $M*N*\log(N)$, dove M è il numero dei modelli e N è il numero di record.

Performance: nodi Modelli

Rete neurale e Kohonen. Gli algoritmi di addestramento delle reti neurali (compreso l'algoritmo Kohonen) effettuano molti passaggi sui dati di addestramento. I dati vengono salvati in memoria fino a un certo limite e i dati eccedenti vengono riversati su disco. L'accesso ai dati di addestramento dal disco è dispendioso in termini di risorse perché il metodo di accesso è diretto e può provocare un'eccessiva attività del disco. È possibile disattivare l'uso dell'archiviazione su disco per questi algoritmi, forzando il salvataggio di tutti i dati in memoria, selezionando l'opzione **Ottimizza per velocità** nella scheda Modello della finestra di dialogo del nodo. Si noti che, se la quantità di memoria necessaria per archiviare i dati supera la memoria di lavoro del processo server, viene eseguito il paging di una parte della memoria su disco e le prestazioni ne risentiranno di conseguenza.

Quando è selezionata l'**ottimizzazione per la memoria**, una percentuale della RAM fisica viene allocata all'algoritmo secondo il valore dell'opzione di configurazione di IBM SPSS Modeler Server *Percentuale limite memoria di modellazione*. Per utilizzare una maggiore quantità di memoria per l'addestramento delle reti neurali, espandere la RAM o incrementare il valore di questa opzione; tenere presente, tuttavia, che l'impostazione di un valore troppo alto causerà il paging su disco.

Il tempo di esecuzione degli algoritmi delle reti neurali dipende dal livello di precisione richiesto e può essere controllato impostando una condizione di arresto nella finestra di dialogo del nodo.

Medie K. L'algoritmo di cluster Medie K dispone delle stesse opzioni di controllo dell'utilizzo della memoria degli algoritmi delle reti neurali. Le prestazioni sui dati salvati su disco sono tuttavia migliori perché l'accesso ai dati è sequenziale.

Performance: espressioni CLEM

Le funzioni sequenza CLEM (“@ funzioni”) che esaminano a ritroso il flusso di dati devono memorizzare una quantità di dati tale da consentire di risalire quanto più possibile all'indietro nel flusso. Per le operazioni in cui il grado di ricerca a ritroso è illimitato, è necessario salvare tutti i valori del campo. Per operazione illimitata si intende un'operazione in cui il valore dell'offset non è un numero intero vero e proprio, per esempio @OFFSET(Vendite, Mese). Il valore dell'offset è il nome del campo *Mese*, il cui valore è sconosciuto fino al momento dell'esecuzione. Il server deve salvare tutti i valori del campo *Vendite* per poter garantire risultati precisi. Quando si conosce il limite superiore è consigliabile specificarlo come argomento aggiuntivo, per esempio come @OFFSET(Vendite, Mese, 12). Questa operazione comunica al server di archiviare solo i 12 valori più recenti del campo *Vendite*. Le funzioni di sequenza, siano esse limitate o meno, impediscono quasi sempre la generazione di stringhe SQL.

Capitolo 14. Accesso facilitato in IBM SPSS Modeler

Panoramica delle funzionalità per l'accesso facilitato in IBM SPSS Modeler

IBM SPSS Modeler offre accesso facilitato a tutti gli utenti nonché supporto specifico per gli utenti con disabilità visive e motorie. In questa sezione viene descritto come utilizzare funzionalità di accesso facilitato quali i lettori di schermo e i tasti di scelta rapida.

Tipi di funzionalità per l'accesso facilitato

Questa applicazione per il data mining può essere utilizzata in numerosi modi, per far fronte alle esigenze di tutti gli utenti. Per esempio, è possibile creare stream, specificare opzioni e leggere l'output senza dover utilizzare il mouse. Negli argomenti seguenti verranno descritti i tasti di scelta rapida disponibili per le varie operazioni. Inoltre, IBM SPSS Modeler supporta un'ampia gamma di lettori di schermo, per esempio JAWS per Windows. È possibile, inoltre, ottimizzare le combinazioni di colori in modo da aumentare il contrasto sullo schermo. Questi argomenti verranno illustrati più avanti in questa sezione.

Accesso facilitato per utenti con problemi di vista

In IBM SPSS Modeler sono disponibili numerose proprietà che consentono di ottimizzare e semplificare l'uso del software.

Opzioni di visualizzazione

È possibile selezionare i colori per la visualizzazione dei grafici. È possibile, inoltre, decidere di utilizzare le impostazioni di Windows anche per Clementine. Questo accorgimento, infatti, può servire ad aumentare il contrasto sullo schermo.

1. Per impostare le opzioni di visualizzazione, fare clic su **Opzioni utente** nel menu Strumenti.
2. Fare clic sulla scheda **Visualizza**. Le opzioni disponibili in questa scheda riguardano la combinazione di colori del software, i colori dei grafici e le dimensioni dei caratteri per i nodi.

Utilizzo di suoni per la notifica

L'attivazione o la disattivazione di suoni consente di controllare la modalità di notifica per determinate operazioni nel software. Per esempio, è possibile attivare suoni per eventi quali la creazione o eliminazione di nodi o la generazione di nuovi modelli o output.

1. Per impostare le opzioni di notifica, fare clic su **Opzioni utente** nel menu Strumenti.
2. Fare clic sulla scheda **Notifiche**.

Controllo dell'apertura automatica di nuove finestre

La scheda Notifiche della finestra di dialogo Opzioni utente viene inoltre utilizzata per specificare se si desidera che i nuovi output generati, quali tabelle e grafici, vengano aperti in una finestra separata. Può essere più conveniente disattivare questa opzione e aprire una finestra di output solo se richiesto.

1. Per impostare le opzioni, fare clic su **Opzioni utente** nel menu Strumenti.
2. Fare clic sulla scheda **Notifiche**.
3. Nella finestra di dialogo selezionare **Nuovo output** dall'elenco nel gruppo **Notifiche visive**.
4. In **Apri finestra**, selezionare **Mai**.

Dimensioni dei nodi

È possibile visualizzare i nodi con le dimensioni standard o con dimensioni ridotte. Se necessario, è possibile regolare queste dimensioni in base alle proprie esigenze.

1. Per impostare le opzioni per le dimensioni dei nodi, fare clic su **Proprietà flusso** nel menu File.
2. Fare clic sulla scheda **Layout**.
3. Nell'elenco **Dimensione icona**, selezionare **Standard**.

Accesso facilitato per utenti non vedenti

Gli utenti non vedenti possono avvalersi di un lettore di schermo, quale JAWS per Windows. In IBM SPSS Modeler sono disponibili varie impostazioni che consentono di ottimizzare l'uso del lettore di schermo.

Opzioni di visualizzazione

In genere le prestazioni dei lettori di schermo sono migliori se il contrasto dello schermo è elevato. Se è già stato impostato un contrasto elevato per Windows, è possibile scegliere di utilizzare le impostazioni di Windows anche per Clementine.

1. Per impostare le opzioni di visualizzazione, fare clic su **Opzioni utente** nel menu Strumenti.
2. Fare clic sulla scheda **Visualizza**.

Utilizzo di suoni per la notifica

L'attivazione o disattivazione di suoni consente di controllare la modalità di notifica per determinate operazioni nel software. Per esempio, è possibile attivare suoni per eventi quali la creazione o eliminazione di nodi o la generazione di nuovi modelli o output.

1. Per impostare le opzioni di notifica, fare clic su **Opzioni utente** nel menu Strumenti.
2. Fare clic sulla scheda **Notifiche**.

Controllo dell'apertura automatica di nuove finestre

La scheda Notifiche della finestra di dialogo Opzioni utente viene inoltre utilizzata per specificare se si desidera che i nuovi output generati vengano aperti in una finestra separata. Può essere più conveniente disattivare questa opzione e aprire una finestra di output quando necessario.

1. Per impostare le opzioni, fare clic su **Opzioni utente** nel menu Strumenti.
2. Fare clic sulla scheda **Notifiche**.
3. Nella finestra di dialogo selezionare **Nuovo output** dall'elenco nel gruppo **Notifiche visive**.
4. In **Apri finestra**, selezionare **Mai**.

Accesso facilitato mediante tastiera

La funzionalità del prodotto è accessibile da tastiera. Di base, è possibile premere Alt più il tasto pertinente per attivare i menu delle finestre (per esempio, premendo Alt+F si accede al menu File) oppure premere il tasto Tab per spostarsi fra i controlli delle finestre di dialogo. Tuttavia, ogni finestra principale del prodotto presenta problemi speciali ma sono anche disponibili utili suggerimenti per l'esplorazione delle finestre di dialogo.

In questa sezione verranno descritte le funzionalità per l'accesso facilitato mediante tastiera, grazie alle quali è possibile svolgere con efficienza ogni operazione, dall'apertura di un flusso all'utilizzo delle finestre di dialogo dei nodi, fino agli interventi sull'output. Inoltre, verranno forniti elenchi dei tasti di scelta rapida, per rendere ancora più efficiente l'esplorazione.

Tasti di scelta rapida per l'esplorazione della finestra principale

La finestra principale di IBM SPSS Modeler è il luogo in cui si svolge la maggior parte del lavoro di data mining. L'area principale è detta **area di disegno del flusso** e serve per creare ed eseguire i flussi di dati. Nella parte inferiore della finestra sono presenti le **palette dei nodi**, che contengono tutti i nodi disponibili. Le palette sono strutturate in schede, che corrispondono al tipo di operazione di data mining per ogni gruppo di nodi. Per esempio, i nodi che consentono di portare i dati in IBM SPSS Modeler sono raggruppati nella scheda Input, mentre i nodi che consentono di derivare campi o applicarvi filtri sono raggruppati nella scheda Oper su campi (abbreviazione di Operazioni sui campi).

Nel lato destro della finestra sono presenti vari strumenti per la gestione dei flussi, dell'output e dei progetti. La metà superiore a destra contiene i **manager** e presenta tre schede, rispettivamente per la gestione dei flussi, dell'output e dei modelli generati. È possibile accedere a questi oggetti selezionando la scheda e un oggetto dall'elenco. La metà inferiore a destra contiene il **riquadro dei progetti**, che consente di organizzare il lavoro in progetti. In quest'area sono presenti due schede, che consentono di accedere a due visualizzazioni diverse di un progetto. Nella **visualizzazione Classi** gli oggetti sono ordinati per tipo, mentre nella **visualizzazione CRISP-DM** gli oggetti sono ordinati in base alla fase di data mining pertinente, per esempio Data Preparation o Modeling. Questi aspetti della finestra di IBM SPSS Modeler sono trattati in dettaglio nella guida in linea e nel Manuale dell'utente.

Di seguito è riportata una tabella con i tasti di scelta rapida che consentono di spostarsi nella finestra principale di IBM SPSS Modeler e di creare stream. I tasti di scelta rapida per le finestre di dialogo e l'output sono riportati negli argomenti successivi. Questi tasti di scelta rapida sono disponibili solo nella finestra principale.

Tabella 32. Tasti di scelta rapida per la finestra principale

Tasto di scelta rapida	Funzione
Ctrl+F5	Consente di spostare lo stato attivo sulle palette di nodi.
Ctrl+F6	Consente di spostare lo stato attivo sull'area di disegno del flusso.
Ctrl+F7	Consente di spostare lo stato attivo nel riquadro dei manager.
Ctrl+F8	Consente di spostare lo stato attivo nel riquadro dei progetti.

Tabella 33. Tasti di scelta rapida per nodi e flussi

Tasto di scelta rapida	Funzione
Ctrl+N	Consente di creare una nuova area di disegno del flusso vuota.
Ctrl+O	Consente di visualizzare la finestra di dialogo Apri, da cui è possibile selezionare e aprire un flusso esistente.
Ctrl+tasti numerici	Consente di spostare lo stato attivo sulla scheda corrispondente in una finestra o riquadro. In una finestra o in un riquadro con schede, per esempio, Ctrl+1 consente di passare alla prima scheda partendo da sinistra, Ctrl+2 alla seconda e così via.
Ctrl+Freccia giù	Nella palette dei nodi, consente di spostare lo stato attivo da una scheda della palette al primo nodo sotto a tale scheda.
Ctrl+Freccia su	Nella palette dei nodi, consente di spostare lo stato attivo da un nodo alla scheda della palette corrispondente.
Invio	Quando è selezionato un nodo in una palette dei nodi (inclusi i modelli rifiniti nella palette dei modelli generati), consente di aggiungere il nodo all'area di disegno del flusso. Se nell'area è già selezionato un nodo, premendo Invio verrà aperta la finestra di dialogo del nodo.
Ctrl+Invio	Quando nella palette è selezionato un nodo, consente di aggiungere tale nodo all'area di disegno del flusso senza selezionarlo, spostando lo stato attivo sul primo nodo della palette.

Tabella 33. Tasti di scelta rapida per nodi e flussi (Continua)

Tasto di scelta rapida	Funzione
Alt+Invio	Quando nella palette è selezionato un nodo, consente di aggiungere tale nodo all'area di disegno del flusso e di selezionarlo, spostando lo stato attivo sul primo nodo della palette.
Maiusc+Barra spaziatrice	Quando un nodo o un commento ha lo stato attivo nella palette, consente di alternare tra la selezione e la deselection del nodo o del commento. Se vengono selezionati anche altri nodi o commenti, questo comporta la loro deselection.
Ctrl+Maiusc+Barra spaziatrice	Quando un nodo o commento ha lo stato attivo nel flusso o nella palette, consente di alternare tra la selezione e la deselection del nodo o del commento. Questo non incide su altri eventuali nodi o commenti selezionati.
Freccia destra/sinistra	Se l'area di disegno del flusso ha lo stato attivo, consente di spostare l'intero stream orizzontalmente sullo schermo. Se una scheda della palette ha lo stato attivo, consente di spostarsi tra le schede. Se un nodo della palette ha lo stato attivo, consente di spostarsi tra i nodi della palette.
Freccia su/giù	Se l'area di disegno del flusso ha lo stato attivo, consente di spostare l'intero stream verticalmente sullo schermo. Se un nodo della palette ha lo stato attivo, consente di spostarsi tra i nodi della palette. Se una palette secondaria ha lo stato attivo, consente di spostarsi tra le altre palette secondarie della scheda della palette.
Alt+Freccia destra/sinistra	Consente di spostare i nodi e i commenti selezionati nell'area di disegno del flusso orizzontalmente nella direzione della freccia.
Alt+Freccia su/giù	Consente di spostare i nodi e i commenti selezionati nell'area di disegno del flusso verticalmente nella direzione della freccia.
Ctrl+A	Consente di selezionare tutti i nodi di un flusso.
Ctrl+Q	Quando un nodo ha lo stato attivo, consente di selezionare il nodo stesso e tutti i nodi downstream e di deselectionare tutti i nodi upstream.
Ctrl+W	Quando un nodo ha lo stato attivo, consente di deselectionare il nodo stesso e tutti i nodi downstream selezionati.
Ctrl+Alt+D	Consente di duplicare il nodo selezionato.
Ctrl+Alt+L	Quando nel flusso è selezionato un nugget del modello, apre una finestra di dialogo Inserisci che consente di caricare un modello salvato da un file .nod nel flusso.
Ctrl+Alt+R	Visualizza la scheda Annotazioni per un nodo selezionato per consentire di rinominare il nodo.
Ctrl+Alt+U	Consente di creare un nodo origine Input utente.
Ctrl+Alt+C	Consente di attivare/disattivare la cache per un nodo.
Ctrl+Alt+F	Consente di svuotare la cache di un nodo.
Scheda	Nell'area di disegno del flusso, consente di spostarsi tra tutti i commenti e i nodi origine del flusso corrente. In una palette di nodi, consente di spostarsi tra i nodi della palette. In una sottopalette selezionata, consente di spostarsi al primo nodo della sottopalette.
Maiusc+Tab	Esegue la medesima operazione del tasto Tab ma in ordine inverso.
Ctrl+Tab	Con lo stato attivo nel riquadro dei manager o nel riquadro dei progetti, sposta lo stato attivo sull'area di disegno del flusso. Con lo stato attivo in una palette di nodi, sposta lo stato attivo dal nodo alla scheda della palette e viceversa.

Tabella 33. Tasti di scelta rapida per nodi e flussi (Continua)

Tasto di scelta rapida	Funzione
Qualsiasi tasto alfabetico	Con lo stato attivo su un nodo del flusso corrente, consente di assegnare lo stato attivo e di passare al nodo successivo il cui nome inizia con la lettera del tasto premuto.
F1	Consente di aprire la guida in linea in corrispondenza di un argomento relativo all'elemento attivo.
F2	Consente di iniziare il processo di connessione per un nodo selezionato nell'area. Utilizzare il tasto Tab per passare al nodo richiesto nell'area e premere Maiusc+Barra spaziatrice per terminare la connessione.
F3	Consente di eliminare tutte le connessioni per il nodo selezionato nell'area.
F6	Consente di spostare lo stato attivo tra il riquadro dei manager, il riquadro dei progetti e le palette dei nodi.
F10	Apri il menu File.
Maiusc+F10	Apri il menu di scelta rapida per il nodo o il flusso.
Canc	Consente di eliminare il nodo selezionato dall'area.
Esc	Chiude il menu di scelta rapida o la finestra di dialogo.
Ctrl+Alt+X	Consente di espandere un Supernodo.
Ctrl+Alt+Z	Consente di eseguire lo Zoom avanti in un Supernodo.
Ctrl+Alt+Maiusc+Z	Consente di eseguire lo Zoom indietro dal Supernodo
Ctrl+E	Con lo stato attivo nell'area di disegno del flusso, consente di eseguire il flusso corrente.

In IBM SPSS Modeler sono inoltre disponibili vari tasti di scelta rapida standard, come Ctrl+C, che consente di effettuare una copia. Consultare l'argomento "Utilizzo dei tasti di scelta rapida" a pagina 22 per ulteriori informazioni.

Tasti di scelta rapida per finestre di dialogo e tabelle

Sono disponibili numerosi tasti di scelta rapida e tasti per il lettore di schermo, che agevolano l'utilizzo di finestre di dialogo, tabelle e tabelle all'interno di finestre di dialogo. Di seguito è riportato un elenco completo.

Tabella 34. Tasti di scelta rapida per le finestre di dialogo e per il builder di espressioni

Tasto di scelta rapida	Funzione
Alt+4	Consente di chiudere tutte le finestre di dialogo o le finestre di output aperte. L'output può essere recuperato dalla scheda Output nel riquadro dei manager.
Ctrl+Fine	Se lo stato attivo è su un controllo nel builder di espressioni, consente di spostare il punto di inserimento alla fine dell'espressione.
Ctrl+1	Nel builder di espressioni consente di spostare lo stato attivo sul controllo di modifica dell'espressione.
Ctrl+2	Nel builder di espressioni consente di spostare lo stato attivo sull'elenco di funzioni.
Ctrl+3	Nel builder di espressioni consente di spostare lo stato attivo sull'elenco di campi.

Tasti di scelta rapida per le tabelle

Questi tasti di scelta rapida funzionano per le tabelle di output e per i controlli tabella in finestre di dialogo di nodi quali Tipo, Filtro e Unione. Di norma si utilizza il tasto Tab per spostarsi tra le celle delle tabelle e Ctrl+Tab per uscire. *Nota:* in alcuni casi può accadere che il lettore di schermo non legga subito il contenuto di una cella. Premendo una o due volte i tasti freccia, si reimposta il software e si avvia la lettura.

Tabella 35. Tasti di scelta rapida per le tabelle

Tasto di scelta rapida	Funzione
Ctrl+W	Per le tabelle, consente di leggere la descrizione breve della riga (row) selezionata. Per esempio, "I valori della riga selezionata 2 sono sesso, flag, m/f, e così via."
Ctrl+Alt+W	Per le tabelle, consente di leggere la descrizione lunga della riga (row) selezionata. Per esempio, "I valori della riga selezionata 2 sono campo = sesso, tipo = flag, sesso = m/f, e così via."
Ctrl+D	Per le tabelle, consente di leggere la descrizione breve dell'area selezionata. Per esempio, "La selezione è composta da una riga per sei colonne."
Ctrl+Alt+D	Per le tabelle, consente di leggere la descrizione lunga dell'area selezionata. Per esempio, "La selezione è composta da una riga per sei colonne. Le colonne selezionate sono Campo, Tipo, Mancante. La riga selezionata è 1".
Ctrl+T	Per le tabelle, consente di leggere una descrizione breve delle colonne selezionate. Per esempio, "Campi, Tipo, Mancante".
Ctrl+Alt+T	Per le tabelle, consente di leggere una descrizione lunga delle colonne selezionate. Per esempio, "Le colonne selezionate sono Campi, Tipo, Mancante".
Ctrl+R	Per le tabelle, consente di conoscere il numero di record nella tabella.
Ctrl+Alt+R	Per le tabelle, consente di conoscere il numero di record nella tabella, nonché i nomi delle colonne.
Ctrl+I	Per le tabelle, consente di leggere le informazioni, ovvero il contenuto, della cella con lo stato attivo.
Ctrl+Alt+I	Per le tabelle, consente di leggere la descrizione lunga delle informazioni della cella con lo stato attivo, ovvero nome della colonna e contenuto della cella.
Ctrl+G	Per le tabelle, consente di leggere informazioni brevi, di carattere generale, sulla selezione.
Ctrl+Alt+G	Per le tabelle, consente di leggere informazioni lunghe, di carattere generale, sulla selezione.
Ctrl+Q	Per le tabelle, consente di attivare e disattivare rapidamente (quick) le descrizioni delle celle delle tabelle. Ctrl+Q consente di leggere descrizioni lunghe, per esempio "Sesso=Femmina" mentre ci si sposta nella tabella utilizzando i tasti freccia. Se si seleziona nuovamente Ctrl+Q, si passa alle descrizioni brevi (contenuto della cella).

Tasti di scelta rapida per i commenti

Quando si opera con i commenti su schermo, è possibile utilizzare i seguenti tasti di scelta rapida.

Tabella 36. Tasti di scelta rapida di commento

Tasto di scelta rapida	Funzione
Alt+C	Consente di mostrare/nascondere la funzione commenti.
Alt+M	Consente di inserire un nuovo commento se i commenti sono visualizzati o di visualizzare i commenti se sono nascosti.

Tabella 36. Tasti di scelta rapida di commento (Continua)

Tasto di scelta rapida	Funzione
Scheda	Nell'area di disegno del flusso, consente di spostarsi tra tutti i commenti e i nodi origine del flusso corrente.
Invio	Quando un commento ha lo stato attivo, indica l'inizio della modifica.
Alt+Invio o Ctrl+Tab	Termina la modifica e salva le modifiche apportate.
Esc	Annulla la modifica. I cambiamenti apportati durante la modifica vengono persi.
Alt+Maiusc+Freccia su	Riduce l'altezza dell'area di testo di una cella di griglia (o un pixel) se è attivato (o disattivato) il blocco sulla griglia.
Alt+Maiusc+Freccia giù	Aumenta l'altezza dell'area di testo di una cella di griglia (o un pixel) se è attivato (o disattivato) il blocco sulla griglia.
Alt+Maiusc+Freccia sinistra	Riduce la larghezza dell'area di testo di una cella di griglia (o un pixel) se è attivato (o disattivato) il blocco sulla griglia.
Alt+Maiusc+Freccia destra	Aumenta la larghezza dell'area di testo di una cella di griglia (o un pixel) se è attivato (o disattivato) il blocco sulla griglia.

Tasti di scelta rapida per il Visualizzatore cluster e il Visualizzatore del modello

Per spostarsi all'interno delle finestre del Visualizzatore cluster e del Visualizzatore del modello è possibile utilizzare dei tasti di scelta rapida.

Tabella 37. Tasti di scelta rapida generali - Visualizzatore cluster e il Visualizzatore del modello

Tasto di scelta rapida	Funzione
Scheda	Consente di spostare lo stato attivo sul comando successivo di quella schermata.
Maiusc+Tab	Consente di spostare lo stato attivo sul comando precedente di quella schermata.
Freccia giù	Se un elenco a discesa ha lo stato attivo, consente di aprire l'elenco o di spostarsi sulla voce successiva al suo interno. Se un menu ha lo stato attivo, consente di spostarsi sulla voce di menu seguente. Se il grafico in miniatura ha lo stato attivo, consente di spostarsi al grafico successivo (o al primo grafico se lo stato attivo è sull'ultima anteprima).
Freccia su	Se un elenco a discesa è aperto, consente di spostarsi sulla voce precedente al suo interno. Se un menu ha lo stato attivo, consente di spostarsi sulla voce di menu precedente. Se il grafico in miniatura ha lo stato attivo, consente di spostarsi al grafico precedente (o all'ultimo grafico se lo stato attivo è sulla prima anteprima).
Invio	Consente di chiudere un elenco a discesa aperto o di effettuare una selezione in un menu aperto.
F6	Consente di alternare lo stato attivo tra il riquadro destro e sinistro della finestra.

Tabella 37. Tasti di scelta rapida generali - Visualizzatore cluster e il Visualizzatore del modello (Continua)

Tasto di scelta rapida	Funzione
Freccia destra e sinistra	Se una scheda ha lo stato attivo, consente di spostarsi sulla scheda precedente o successiva. Se un menu ha lo stato attivo, consente di spostarsi sul menu precedente o successivo.
Alt+lettera	Consente di selezionare il pulsante o il menu nel cui nome è sottolineata la lettera.
Esc	Consente di chiudere un menu o un elenco a discesa aperto.

Solo Visualizzatore cluster

Il Visualizzatore cluster dispone di una visualizzazione Cluster che contiene una griglia cluster-per-funzioni.

Per scegliere la visualizzazione Cluster anziché la visualizzazione Riepilogo del modello:

1. Premere più volte il tasto Tab fino a quando viene selezionato il pulsante **Visualizza**.
2. Premere due volte il tasto Freccia giù per selezionare **Cluster**.
Da qui è possibile selezionare una singola cella all'interno della griglia:
3. Premere più volte il tasto Tab fino ad arrivare sull'ultima icona della barra degli strumenti di visualizzazione.



Figura 17. Icona Mostra struttura di visualizzazione

4. Premere un'altra volta il tasto Tab, seguito dalla Barra spaziatrice e da un tasto freccia.

Attualmente sono disponibili i seguenti tasti di scelta rapida:

Tabella 38. Tasti di scelta rapida del Visualizzatore cluster

Tasto di scelta rapida	Funzione
Tasto freccia	Consente di spostare lo stato attivo tra le singole celle della griglia. La visualizzazione della distribuzione delle celle nel riquadro di destra cambia via via che lo stato attivo passa da una cella all'altra.
Ctrl+, (virgola)	Consente di selezionare o deselezionare tutta la colonna nella griglia in cui una cella ha lo stato attivo. Per aggiungere una colonna alla selezione, spostarsi su una cella in quella colonna con i tasti freccia e premere nuovamente Ctrl+.
Scheda	Consente di spostare lo stato attivo al di fuori della griglia e sul comando successivo della schermata.
Maiusc+Tab	Consente di spostare lo stato attivo al di fuori della griglia e sul comando precedente della schermata.
F2	Consente di accedere alla modalità di modifica (solo celle di descrizione e delle etichette).
Invio	Consente di salvare le modifiche apportate e di uscire dalla modalità di modifica (solo celle di descrizione e delle etichette).
Esc	Consente di uscire dalla modalità di modifica senza salvare le modifiche (solo celle di descrizione e delle etichette).

Esempio di utilizzo dei tasti di scelta rapida: Creazione di flussi

Per rendere più chiara la procedura per la creazione di stream mediante la tastiera o un lettore di schermo, di seguito è riportato un esempio che illustra come creare un flusso senza utilizzare il mouse. In questo esempio verrà creato un flusso contenente un nodo Testo Variabile, un nodo Deriva e un nodo Istogramma:

1. **Avviare IBM SPSS Modeler.** Quando IBM SPSS Modeler viene avviato, lo stato attivo è sulla scheda Preferiti della palette dei nodi.
2. **Ctrl+Freccia giù.** Consente di spostarsi dalla scheda nel corpo della scheda.
3. **Freccia a destra.** Consente di spostare lo stato attivo sul nodo Testo Variabile.
4. **Barra spaziatrice.** Consente di selezionare il nodo Testo Variabile.
5. **Ctrl+Invio.** Consente di aggiungere il nodo Testo variabile all'area di disegno del flusso. Inoltre, la selezione viene mantenuta sul nodo Testo Variabile, in modo che il nodo che verrà aggiunto successivamente risulti connesso ad esso.
6. **Tabulazione.** Consente di spostare lo stato attivo sulla tavolozza dei nodi.
7. **Freccia a destra 4 volte.** Consente di passare al nodo Deriva.
8. **Barra spaziatrice.** Consente di selezionare il nodo Deriva.
9. **Alt+Invio.** Consente di aggiungere il nodo Deriva all'area e di spostare la selezione sul nodo Deriva. Ora questo nodo è pronto per essere connesso al successivo nodo che verrà aggiunto.
10. **Tabulazione.** Consente di spostare lo stato attivo sulla tavolozza dei nodi.
11. **Freccia a destra 5 volte.** Consente di spostare lo stato attivo sul nodo Istogramma nella palette.
12. **Barra spaziatrice.** Consente di selezionare il nodo Istogramma.
13. **Invio.** Consente di aggiungere il nodo al flusso e di spostare lo stato attivo sull'area di disegno del flusso.

Passare all'esempio successivo o salvare il flusso se si desidera provare il prossimo esempio in un secondo tempo.

Esempio di utilizzo dei tasti di scelta rapida: Modifica dei nodi

In questo esempio verrà utilizzato il flusso creato nell'esempio precedente. Lo stream è costituito da un nodo Testo Variabile, da un nodo Deriva e da un nodo Istogramma. La procedura parte con lo stato attivo sul terzo nodo del flusso, vale a dire sul nodo Istogramma.

1. **Ctrl+Freccia sinistra 2 volte.** Consente di spostare nuovamente lo stato attivo sul nodo Testo Variabile.
2. **Invio.** Consente di aprire la finestra di dialogo Testo Variabile. Con il tasto Tab spostarsi fino al campo File e digitare il percorso e il nome di un file di testo per selezionare il file. Con la combinazione di tasti Ctrl+Tab spostarsi nella parte inferiore della finestra di dialogo fino a raggiungere il pulsante OK e premere Invio per chiudere la finestra di dialogo.
3. **Ctrl+Freccia destra.** Consente di spostare lo stato attivo sul secondo nodo, vale a dire il nodo Deriva.
4. **Invio.** Consente di aprire la finestra di dialogo del nodo Deriva. Premere Tab per selezionare i campi e specificare le condizioni per la derivazione. Premere Ctrl+Tab per spostarsi sul pulsante OK e premere Invio per chiudere la finestra di dialogo.
5. **Ctrl+Freccia destra.** Consente di spostare lo stato attivo sul terzo nodo, vale a dire il nodo Istogramma.
6. **Invio.** Consente di aprire la finestra di dialogo del nodo Istogramma. Premere Tab per selezionare i campi e specificare le opzioni per il grafico. Per gli elenchi a discesa, premere il tasto Freccia giù per aprire l'elenco ed evidenziare una voce, quindi premere Invio per selezionarla. Con il tasto Tab spostarsi sul pulsante OK e premere Invio per chiudere la finestra di dialogo.

A questo punto è possibile aggiungere altri nodi o eseguire il flusso corrente. Di seguito sono riportati alcuni suggerimenti da tenere presente quando si creano stream:

- Quando si collegano manualmente i nodi, utilizzare F2 per creare il punto di inizio di una connessione, il tasto Tab per spostarsi sul punto di fine e Maiusc+Barra spaziatrice per finalizzare la connessione.
- Utilizzare F3 per eliminare tutte le connessioni per il nodo selezionato nell'area.
- Dopo aver creato un flusso, utilizzare Ctrl+E per eseguire il flusso corrente.

È disponibile un elenco completo dei tasti di scelta rapida. Consultare l'argomento "Tasti di scelta rapida per l'esplorazione della finestra principale" a pagina 167 per ulteriori informazioni.

Uso di un lettore di schermo

In commercio sono disponibili numerosi lettori di schermo. IBM SPSS Modeler è configurato per supportare JAWS per Windows tramite il Java Access Bridge che viene installato insieme a IBM SPSS Modeler. Se JAWS è installato, per utilizzare il prodotto è sufficiente avviare JAWS prima di avviare IBM SPSS Modeler.

Per la natura specifica e unica della rappresentazione grafica del processo di data mining in IBM SPSS Modeler, la rappresentazione visiva di grafici e diagrammi è eccellente. È possibile, tuttavia, analizzare i dati e prendere decisioni sulla base di output e modelli anche visualizzandoli in forma testuale, tramite un lettore di schermo.

Nota: sui computer client a 64 bit alcune funzioni delle tecnologie per l'accesso facilitato non funzionano, in quanto Java Access Bridge non è progettato per il funzionamento a 64 bit.

Utilizzo del file di dizionario di IBM SPSS Modeler

In IBM SPSS Modeler è disponibile un file di dizionario (*Awt.JDF*) da utilizzare con JAWS. Per utilizzare questo file:

1. Passare alla sottodirectory */accessibility* dell'installazione di IBM SPSS Modeler e copiare il file di dizionario (*Awt.JDF*).
2. Copiarlo nella directory con gli script per JAWS.

Se nel sistema sono presenti altre applicazioni JAVA, è possibile che esista già un altro file denominato *Awt.JDF*. In questo caso, per poter utilizzare il file di dizionario potrebbe essere necessario modificarlo manualmente.

Utilizzo di un lettore di schermo con output HTML

In IBM SPSS Modeler possono verificarsi alcune difficoltà per quanto riguarda la visualizzazione di output HTML mediante un lettore di schermo. I problemi si verificano con i seguenti tipi di output:

- Output visualizzato nella scheda Avanzate dei nodi Regressione, Regressione logistica e Fattoriale
- Output del nodo Report

In ognuna di queste finestre o finestre di dialogo è presente uno strumento sulla barra degli strumenti che consente di visualizzare l'output nel browser predefinito, in cui è disponibile il supporto per il lettore di schermo. È quindi possibile utilizzare il lettore di schermo per trasmettere le informazioni relative all'output.

Accesso facilitato nella finestra Struttura ad albero interattiva

La visualizzazione standard di un modello di struttura ad albero delle decisioni nella finestra Struttura ad albero interattiva potrebbe causare problemi ai lettori di schermo. Per accedere a una versione di accesso facilitato, nei menu Struttura ad albero interattiva fare clic su:

Visualizza > Finestra Accessibile

Viene visualizzata una vista simile alla mappa della struttura ad albero standard che JAWS può leggere correttamente. È possibile spostarsi in alto, in basso, verso destra, verso sinistra utilizzando i tasti freccia standard. Durante l'esplorazione della finestra accessibile, si sposta anche lo stato attivo nella finestra Struttura ad albero interattiva. Utilizzare la barra spaziatrice per modificare la selezione, oppure utilizzare Ctrl+Barra spaziatrice per estendere la selezione corrente.

Suggerimenti per l'utilizzo

Esistono vari accorgimenti per rendere più facilmente accessibile l'ambiente di IBM SPSS Modeler. Di seguito sono riportati alcuni suggerimenti di carattere generale per l'utilizzo di IBM SPSS Modeler.

- **Uscita da caselle di testo estese.** Utilizzare Ctrl+Tab per uscire da caselle di testo estese. *Nota:* Ctrl+Tab consente, inoltre, di uscire dai controlli delle tabelle.
- **Utilizzo di Tab anziché dei tasti freccia.** Quando si selezionano le opzioni di una finestra di dialogo è necessario utilizzare Tab per spostarsi tra i pulsanti d'opzione. In questo contesto, infatti, i tasti freccia non funzionano.
- **Elenchi a discesa.** Negli elenchi a discesa delle finestre di dialogo è possibile utilizzare il tasto Esc o la Barra spaziatrice per selezionare una voce e chiudere l'elenco. È possibile, inoltre, utilizzare il tasto Esc per chiudere gli elenchi a discesa che rimangono aperti dopo che si è passati a un altro controllo mediante il tasto Tab.
- **Stato dell'esecuzione.** Quando si esegue un flusso o un database di grandi dimensioni, può verificarsi un ritardo di JAWS per quanto riguarda la lettura dello stato del flusso. Premere Ctrl periodicamente per aggiornare la lettura dello stato.
- **Utilizzo delle tavolozze di nodi.** La prima volta che si accede a una scheda della palette di nodi, può accadere che JAWS non proponga il nome del nodo. In questo caso è possibile utilizzare Ctrl+Freccia destra, quindi Ctrl+Freccia sinistra per reimpostare il lettore di schermo e ottenere il nome del nodo.
- **Letture dei menu.** Può accadere, la prima volta che si apre un menu, che JAWS salti il primo comando. Se si sospetta un'evenienza simile, utilizzare Freccia giù, quindi Freccia su per conoscere il primo comando del menu.
- **Menu in cascata.** JAWS non è in grado di leggere il primo livello di un menu in cascata. Se si avverte un'interruzione della lettura mentre si scorre un menu, premere il tasto Freccia destra per conoscere i comandi del menu secondario.

Inoltre, se è stato installato IBM SPSS Modeler Text Analytics, i seguenti suggerimenti consentono di rendere l'interfaccia workbench interattiva ancora più accessibile.

- **Accesso alle finestre di dialogo.** Potrebbe essere necessario premere il tasto Tab per spostare lo stato attivo sul primo controllo di una finestra di dialogo.
- **Uscita da caselle di testo estese.** Utilizzare Ctrl+Tab per uscire da caselle di testo estese e passare al controllo successivo. *Nota:* Ctrl+Tab consente, inoltre, di uscire dai controlli delle tabelle.
- **Digitare la prima lettera per trovare un elemento in un elenco a struttura ad albero.** Quando si cerca un elemento nel riquadro delle categorie, nel riquadro dei risultati estratti o nella struttura ad albero delle librerie, è possibile digitare la prima lettera dell'elemento quando lo stato attivo si trova nel riquadro. In questo modo verrà selezionata l'occorrenza successiva di un elemento che inizia con la lettera immessa.
- **Elenchi a discesa.** Negli elenchi a discesa delle finestre di dialogo è possibile utilizzare la Barra spaziatrice per selezionare una voce e chiudere l'elenco.

Ulteriori suggerimenti per l'utilizzo vengono discussi ampiamente nei seguenti argomenti.

Interferenze con altri programmi

Durante il test di IBM SPSS Modeler con lettori di schermo quali JAWS, il team di sviluppo ha scoperto che l'utilizzo di un server SMS (Systems Management Server) nell'organizzazione potrebbe interferire con la capacità di JAWS di leggere applicazioni basate su Java, quali IBM SPSS Modeler. Per risolvere il problema è sufficiente disattivare SMS. Per ulteriori informazioni su SMS, visitare il sito Web Microsoft.

JAWS e Java

Il livello di supporto per le applicazioni basate su Java è diverso nelle varie versioni di JAWS. Sebbene IBM SPSS Modeler supporti tutte le versioni recenti di JAWS, alcune versioni potrebbero presentare problemi di lieve entità se utilizzate in combinazione con sistemi basati su Java. Visitare il sito Web di JAWS per Windows all'indirizzo <http://www.FreedomScientific.com>.

Utilizzo dei grafici in IBM SPSS Modeler

Per i lettori di schermo è difficile interpretare le visualizzazioni grafiche delle informazioni, per esempio istogrammi, grafici di valutazione, multiplot e grafici a dispersione. Si tenga presente, tuttavia, che è possibile visualizzare grafici Web e nodi di distribuzione utilizzando il riepilogo in formato testo disponibile nella finestra di output.

Capitolo 15. Supporto Unicode

Supporto Unicode in IBM SPSS Modeler

IBM SPSS Modeler dispone del supporto completo per Unicode, sia nella versione IBM SPSS Modeler che in quella IBM SPSS Modeler Server. Consente quindi lo scambio di dati con altre applicazioni che supportano Unicode, inclusi i database multilingue, senza alcun rischio di perdita delle informazioni dovuta alla conversione in o da uno schema di codifica specifico di determinate opzioni internazionali.

- IBM SPSS Modeler archivia i dati Unicode internamente ed è in grado di leggere e scrivere dati multilingue memorizzati nei database in formato Unicode senza alcun rischio di perdita.
- IBM SPSS Modeler può leggere e scrivere file di testo con codifica UTF-8. Le operazioni di importazione ed esportazione di file di testo verranno eseguite di default nella codifica specifica delle opzioni internazionali, ma supporteranno la codifica alternativa UTF-8. Questa impostazione può essere specificata nei nodi di importazione ed esportazione file oppure è possibile modificare la codifica di default nella finestra di dialogo delle proprietà del flusso. Consultare l'argomento "Impostazione delle opzioni generali per i flussi" a pagina 42 per ulteriori informazioni.
- I file di dati Statistics, SAS e di testo archiviati nella codifica specifica delle opzioni internazionali verranno convertiti in UTF-8 durante l'importazione e riconvertiti in fase di esportazione. Se, durante la scrittura in un file, sono presenti caratteri Unicode non disponibili nel set di caratteri specifico delle opzioni internazionali, tali caratteri verranno sostituiti e verrà visualizzato un avviso. Questa situazione dovrebbe verificarsi solo nel caso in cui i dati siano stati importati da una sorgente dati che supporta Unicode (un database o un file di testo UTF-8) e contengano caratteri di opzioni internazionali diverse o di più opzioni internazionali o set di caratteri.
- Le immagini IBM SPSS Modeler Solution Publisher sono codificate UTF-8 e possono essere perfettamente trasferite tra piattaforme e opzioni internazionali diverse.

Informazioni su Unicode

L'obiettivo dello standard Unicode consiste nel fornire un metodo uniforme per la codifica di testo multilingue, in modo da garantire una facile condivisione delle informazioni tra confini, opzioni internazionali e applicazioni. Lo standard Unicode, giunto ora alla versione 4.0.1, definisce un set di caratteri che rappresenta un superinsieme di tutti i set di caratteri comunemente utilizzati oggi e assegna a ciascun carattere un nome univoco e un punto di codice. I caratteri e i relativi punti di codice sono identici a quelli UCS (Universal Character Set), definiti dalla norma ISO-10646. Per ulteriori informazioni, fare riferimento alla Home page Unicode.

Note

Queste informazioni sono state preparate per prodotti e servizi offerti in tutto il mondo.

È possibile che IBM non offra i prodotti, servizi o funzioni illustrati in questa documentazione. Consultare il rappresentante locale IBM per le informazioni sui prodotti e servizi attualmente disponibili nella propria zona. Qualsiasi riferimento a un prodotto, programma o servizio IBM non implica o intende dichiarare che può essere utilizzato solo quel prodotto, programma o servizio IBM. In sostituzione a quelli forniti da IBM è possibile utilizzare qualsiasi prodotto, programma o servizio funzionalmente equivalente che non comporti la violazione dei diritti di proprietà intellettuale IBM o altri diritti. Tuttavia, è responsabilità dell'utente valutare e verificare il funzionamento di qualsiasi prodotto, programma o servizio non IBM.

IBM potrebbe avere brevetti o domande di brevetti in corso relativi ad argomenti discussi nella presente pubblicazione. La consegna del presente documento non conferisce alcuna licenza rispetto a questi brevetti. Rivolgere per iscritto i quesiti sulle licenze a:

IBM Director of Licensing
IBM Europe
Schoenaicher Str.220
D-7030 Boeblingen
Deutschland

Per richieste di licenze relative ad informazioni double-byte (DBCS) contattare il Dipartimento di Proprietà Intellettuale IBM nel proprio paese o inviare richieste per iscritto a:

Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan Ltd.
1623-14, Shimotsuruma, Yamato-shi
Kanagawa 242-8502 Japan

Il seguente paragrafo non è valido nel Regno Unito o per tutti i paesi le cui leggi nazionali siano in contrasto con le disposizioni in esso contenute INTERNATIONAL BUSINESS MACHINES CORPORATION FORNISCE QUESTA PUBBLICAZIONE "NELLO STATO IN CUI ESSA SI TROVA" SENZA ALCUNA GARANZIA ESPLICITA O IMPLICITA IVI INCLUSE EVENTUALI GARANZIE DI COMMERCIALIZZABILITÀ ED IDONEITÀ AD UNO SCOPO PARTICOLARE Alcuni stati non consentono limitazioni di garanzie espresse o implicite in determinate transazioni, pertanto quanto sopra potrebbe non essere applicabile.

Le presenti informazioni possono includere imprecisioni tecniche o errori tipografici. Le modifiche periodiche apportate alle informazioni contenute in questa pubblicazione verranno inserite nelle nuove edizioni della pubblicazione. IBM si riserva il diritto di apportare miglioramenti e/o modifiche al prodotto o programma descritto in questa pubblicazione in qualsiasi momento e senza preavviso.

Qualsiasi riferimento nelle presenti informazioni a siti Web non IBM viene fornito esclusivamente per facilitare la consultazione e non rappresenta in alcun modo un'approvazione o sostegno da parte nostra di tali siti Web. I materiali disponibili sui siti Web non fanno parte di questo prodotto IBM e l'utilizzo di questi è a discrezione dell'utente.

IBM può utilizzare o distribuire qualsiasi informazione fornita dall'utente nel modo che ritiene più idoneo senza incorrere in alcun obbligo nei confronti dell'utente stesso.

Coloro che detengono la licenza su questo programma e desiderano avere informazioni su di esso allo scopo di consentire: (i) lo scambio di informazioni tra programmi indipendenti ed altri (compreso questo) e (ii) l'uso reciproco di tali informazioni dovrebbero contattare:

IBM Software Group
ATTN: Licensing
200 W. Madison St.
Chicago, IL; 60606
U.S.A.

Tali informazioni saranno fornite in conformità ai termini e alle condizioni in vigore e, in alcuni casi, dietro pagamento.

Il programma su licenza descritto in questa documentazione e tutto il materiale su licenza ad esso relativo vengono forniti da IBM nei termini del Customer Agreement IBM IBM International Program License Agreement o di eventuali accordi equivalenti intercorsi tra le parti.

Tutti i dati sulle prestazioni qui contenuti sono stati elaborati in ambiente controllato. Di conseguenza, i risultati ottenuti con sistemi operativi diversi possono variare in modo significativo. Alcune misurazioni potrebbero essere state effettuate su sistemi in corso di sviluppo e non c'è garanzia che tali misurazioni coincidano con quelle effettuate sui sistemi comunemente disponibili. Inoltre, alcune misurazioni potrebbero essere stime elaborate tramite l'estrapolazione. I risultati effettivi potrebbero variare. Gli utenti di questo documento devono verificare i dati relativi al proprio ambiente specifico.

Le informazioni relative a prodotti non IBM sono state ottenute dai fornitori di tali prodotti, da loro annunci pubblicati e da altre fonti disponibili al pubblico. IBM non ha testato quei prodotti e non può garantire l'accuratezza delle prestazioni, la compatibilità o qualsiasi altra dichiarazione relativa a prodotti non-IBM. Eventuali domande in merito alle funzionalità dei prodotti non IBM vanno indirizzate ai fornitori di tali prodotti.

Qualsiasi affermazione relativa agli obiettivi e alla direzione futura di IBM è soggetta a modifica o revoca senza preavviso e concerne esclusivamente gli scopi dell'azienda.

Le presenti informazioni includono esempi di dati e report utilizzati in operazioni di business quotidiane. Per fornire una descrizione il più possibile esaustiva, gli esempi includono nomi di persone, società, marchi e prodotti. Tutti questi nomi sono fittizi e ogni somiglianza a nomi e indirizzi utilizzati da aziende reali è puramente casuale.

Per chi visualizza queste informazioni a video: le fotografie e le illustrazioni a colori potrebbero non essere disponibili.

Marchi

IBM, il logo IBM e ibm.com sono marchi o marchi registrati di International Business Machines Corp., registrati in molte giurisdizioni nel mondo. Altri nomi di prodotti e servizi possono essere marchi di IBM o altre società. Un elenco aggiornato di marchi IBM è disponibile sul sito Web "Copyright and trademark information" all'indirizzo www.ibm.com/legal/copytrade.shtml.

Intel, il logo Intel, Intel Inside, il logo Intel Inside, Intel Centrino, il logo Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium e Pentium sono marchi o marchi registrati di Intel Corporation o delle sue consociate negli Stati Uniti e in altri paesi.

Linux è un marchio registrato di Linus Torvalds negli Stati Uniti e/o negli altri paesi.

Microsoft, Windows, Windows NT e il logo Windows sono marchi di Microsoft Corporation negli Stati Uniti e/o negli altri paesi.

UNIX è un marchio registrato di The Open Group negli Stati Uniti e in altri paesi.

Java e tutti i marchi e logo basati su Java sono marchi o marchi registrati di Oracle e/o suoi affiliati.

Altri nomi di prodotti e servizi possono essere marchi commerciali di IBM o di altre aziende.

Indice analitico

A

- accesso facilitato 165, 176
 - esempio 173
 - funzioni in IBM SPSS Modeler 165
 - suggerimenti in IBM SPSS Modeler 175
- aggiornamento
 - nodi origine 42
- aggiornamento modelli 131, 134
- aggiunta
 - a un progetto 142
- aggiunta di connessioni IBM SPSS Modeler Server 13
- analisi basata su strutture ad albero
 - applicazioni comuni 25
- Analisi Campione/Sfidante 118, 131
- and, operatore 95
- annotazione
 - flussi 55, 59
 - nodi 55, 59
- annotazioni
 - cartella 145
 - conversione in commenti 59
 - progetto 144
- annulla 19
- annullamento della mappatura dei campi 62
- anteprima
 - dati nodo 40
- apertura
 - flussi 62
 - modelli 62
 - nodi 62
 - output 62
 - progetti 142
 - stati 62
- applicazioni 25
- applicazioni comuni 25
- applicazioni del data mining 26
- Applicazioni predittive 131
- apprendimento automatico 25
- archiviazione di oggetti in IBM SPSS Collaboration and Deployment Services Repository 120
- area 15
- area del flusso
 - impostazioni 46
- assegnazione di un nome a nodi e stream 59
- attivazione nodi 37
- attribute 25
- automazione 71
- avvertenze
 - impostazione delle opzioni 150
- avvisi 47

B

- barra degli strumenti 19
- barra rovesciata nelle espressioni CLEM 86
- blocco di nodi 41
- blocco oggetti IBM SPSS Collaboration and Deployment Services Repository 127
- Builder di espressioni
 - accesso 79
 - cenni generali 79
 - utilizzo 79

C

- cache
 - attivazione 149
 - impostazione di una cache 38
 - salvataggio 39
 - svuotamento 39, 42
- calcolo del punteggio
 - ramo 54, 133, 135
- cambiamento giorni 43
- campi 25, 85, 87
 - in espressioni CLEM 80
 - visualizzazione di valori 80
- campi del modello 64
- campi essenziali 62, 64
- campi obbligatori 64
- campi ora
 - conversione 108
- caratteri 85, 86, 152, 153
- caratteri speciali
 - rimozione da stringhe 75
- caricamento
 - nodi 62
 - stati 62
- caricamento in IBM SPSS Modeler Server 12
- cartelle, IBM SPSS Collaboration and Deployment Services Repository 127, 128
- caso 25
- classi 19, 141, 142
- CLEM 79
 - controllo di espressioni 81
 - esempi 71
 - espressioni 74, 85
 - funzioni 79
 - generazione di espressioni 79
 - introduzione 24, 71
 - lingua 85
 - tipi di dati 85, 86, 87
- client
 - directory di default 150
- codifica 42, 177
- codifica di default del flusso 42
- codifica di testo 42
- codifica UTF-8 42, 177

colori

- impostazioni 152
- colori script
 - impostazioni 153
- commenti
 - elenco di tutti in un flusso 58
 - scelte rapide da tastiera 170, 171
 - su nodi e stream 55
- concatenazione di stringhe 92
- condizioni 74
- Conformità 508 165
- connessioni
 - a IBM SPSS Collaboration and Deployment Services Repository 119
 - a IBM SPSS Modeler Server 12, 13
 - cluster di server 13
- convenzioni 91
- Coordinator of Processes 13
- COP 13
- copia 19
- creazione di palette personalizzate 155
- creazione di sottopalette 156
- CRISP-DM 19, 141
 - visualizzazione progetti 141

D

- data mining 25
 - esempi di applicazioni 34
 - strategia 27
- database
 - funzioni 79
- date
 - conversione 108
 - gestione 108
- dati
 - anteprima 40
- decimali
 - formati di visualizzazione 44
- default
 - fase del progetto 141
 - deployment di scenari 131
- descrizioni
 - annotazione di nodi 59
- descrizioni flusso 52
- descrizioni stream 54
- directory
 - default 150
- directory temp 14
- disattiva nodi 37, 38
- distribuzione 118
- distribuzione chi-quadrato
 - funzioni di probabilità 97
- distribuzione f
 - funzioni di probabilità 97
- distribuzione normale
 - funzioni di probabilità 97
- distribuzione t
 - funzioni di probabilità 97
- diverso da, operatore 93

documentazione 3
DTD 138

E

elenchi 85, 86
elenco di tutti i commenti di un flusso 58
esecuzione dei flussi 54
esempi
 cenni generali 5
 Guida alle applicazioni 3
esempi di applicazioni 3
esplorazione
 scelte rapide da tastiera 166
esportazione
 descrizioni stream 54
 PMML 138
espressioni 85
espressioni CLEM
 performance 163
etichette
 valore 138
 variabile 138
 visualizzazione 42
etichette, oggetti IBM SPSS Collaboration and Deployment Services Repository 130
etichette versione, oggetti IBM SPSS Collaboration and Deployment Services Repository 130

F

fattore 174
file di dati testuali
 codifica 177
file di dizionario 174
file di output
 salvataggio 61
file di PowerPoint 142
file di registro
 visualizzazione SQL generato 46
file di stream di backup
 ripristino 60
file SAS
 codifica 177
File statistiche
 codifica 177
finestra di dialogo all'avvio 152
finestra di dialogo di benvenuto 152
finestra principale 15
Finestra Struttura ad albero interattiva
 accesso facilitato 174
flussi 11
 aggiunta a progetti 142
 aggiunta di commenti 55
 aggiunta di nodi 36, 38
 annotazione 55, 59
 archiviazione in IBM SPSS Collaboration and Deployment Services Repository 122
 caricamento 62
 connessione di nodi 36
 creazione 35
 disattivazione nodi 37

flussi (*Continua*)
 esecuzione 54
 file di backup 60
 ignorare i nodi 37
 opzioni 41, 42, 43, 44, 46
 opzioni di distribuzione 131
 ridenominazione 52, 59
 ridimensionamento 22
 salvataggio 60
 visualizzazione tempi di esecuzione 48
flusso 15
formati di data 43, 87, 88
formati di ora 43, 87, 88
formati di visualizzazione
 decimali 44
 numeri 44
 scientifico 44
 simbolo di raggruppamento 44
 valuta 44
formati di visualizzazione numeri 44
formato di visualizzazione valute 44
funzione @BLANK 69, 92, 114
funzione @DIFF 109
funzione @FIELD 69, 115
funzione @FIELDS_BETWEEN 69, 77, 115
funzione @FIELDS_MATCHING 69, 77, 115
funzione @INDEX 109
funzione @LAST_NON_BLANK 109, 114
funzione @MAX 109
funzione @MEAN 109
funzione @MIN 109
funzione @MULTI_RESPONSE_SET 78, 115
funzione @NULL 69, 92, 114
funzione @OFFSET 109
 considerazioni sulle performance 163
funzione @PARTITION_FIELD 115
funzione @PREDICTED 115
funzione @SDEV 109
funzione @SINCE 109
funzione @SUM 109
funzione @TARGET 115
funzione @TESTING_PARTITION 115
funzione @THIS 109
funzione @TODAY 105
funzione @TRAINING_PARTITION 115
funzione
 @VALIDATION_PARTITION 115
funzione abs 96
funzione allbutfirst 99
funzione allbutlast 99
funzione alphabefore 99
funzione arccos 97
funzione arccosh 97
funzione arcsin 97
funzione arcsinh 97
funzione arctan 97
funzione arctan2 97
funzione arctanh 97
funzione cdf_chisq 97
funzione cdf_f 97
funzione cdf_normal 97
funzione cdf_t 97

funzione cos 97
funzione cosh 97
funzione count_equal 77, 93
funzione count_greater_than 77, 93
funzione count_less_than 77, 93
funzione count_non_nulls 93
funzione count_not_equal 77, 93
funzione count_nulls 69, 77, 93
funzione count_substring 99
funzione date_before 93
funzione datetime_date 92
funzione DIFF 109
funzione div 96
funzione endstring 99
funzione exp 96
funzione first_index 78, 93
funzione first_non_null 78, 93
funzione first_non_null_index 78, 93
funzione fracof 96
funzione hasendstring 99
funzione hasmidstring 99
funzione hasstartstring 99
funzione INDEX 109
funzione integer_bitcount 98
funzione integer_leastbit 98
funzione integer_length 98
funzione intof 96
funzione is_date 92
funzione is_datetime 92
funzione is_integer 92
funzione is_number 92
funzione is_real 92
funzione is_string 92
funzione is_time 92
funzione is_timestamp 92
funzione isalphacode 99
funzione isendstring 99
funzione islowercode 99
funzione ismidstring 99
funzione isnumbercode 99
funzione isstartstring 99
funzione issubstring 99
funzione issubstring_count 99
funzione issubstring_lim 99
funzione isuppercode 99
funzione last_index 78, 93
funzione LAST_NON_BLANK 109
funzione last_non_null 78, 93
funzione last_non_null_index 78, 93
funzione length 99
funzione locchar 99
funzione locchar_back 99
funzione log 96
funzione log10 96
funzione lowertoupper 99
funzione matches 99
funzione max 93
funzione MAX 109
funzione max_index 78, 93
funzione max_n 77, 93
funzione MEAN 109
funzione mean_n 77, 96
funzione membro 93
funzione min 93
funzione MIN 109
funzione min_index 78, 93
funzione min_n 77, 93

- funzione mod 96
- funzione negate 96
- funzione OFFSET 109
- funzione oneof 99
- funzione pi 97
- funzione potenza (esponenziale) 96
- funzione random 99
- funzione random0 99
- funzione rem 96
- funzione replace 99
- funzione replicate 99
- funzione round 96
- funzione SDEV 109
- funzione sdev_n 77, 96
- funzione sign 96
- funzione sin 97
- funzione SINCE 109
- funzione sinh 97
- funzione skipchar 99
- funzione skipchar_back 99
- funzione soundex 104
- funzione soundex_difference 104
- funzione sqrt 96
- funzione startstring 99
- funzione stripchar 99
- funzione strmember 99
- funzione subscrs 99
- funzione substring 99
- funzione substring_between 99
- funzione SUM 109
- funzione sum_n 77, 96
- funzione tan 97
- funzione tanh 97
- funzione testbit 98
- funzione THIS 109
- funzione time_before 93
- funzione to_date 92, 105
- funzione to_dateline 105
- funzione to_datetime 92
- funzione to_integer 92
- funzione to_number 92
- funzione to_string 92
- funzione to_time 92, 105
- funzione to_timestamp 92, 105
- funzione trim 99
- funzione trim_start 99
- funzione trimend 99
- funzione undef 114
- funzione unicode_char 99
- funzione unicode_value 99
- funzione uppertolower 99
- funzione value_at 78, 93
- funzioni 87, 88, 91, 92, 109
 - @BLANK 69
 - @FIELD 79, 115
 - @GLOBAL_MAX 113
 - @GLOBAL_MEAN 113
 - @GLOBAL_MIN 113
 - @GLOBAL_SDEV 113
 - @GLOBAL_SUM 113
 - @PARTITION 115
 - @PREDICTED 79, 115
 - @TARGET 79, 115
 - database 79
 - esempi 71
 - funzioni definite dall'utente (UDF) 79

- funzioni (*Continua*)
 - gestione valori mancanti 69
 - in espressioni CLEM 79
- funzioni bitwise 98
- funzioni CLEM
 - bitwise 98
 - casuali 99
 - conversione 92
 - datetime 105
 - elenco funzioni disponibili 90
 - funzioni speciali 115
 - globali 113
 - informazione 92
 - logiche 95
 - numerico 96
 - probabilità 97
 - sequence 109
 - stringa 99
 - trigonometriche 97
 - valori mancanti 69
 - vuoti e null 114
- Funzioni CLEM
 - confronto 93
- funzioni data ora
 - datetime_date 105
 - datetime_day 105
 - datetime_day_name 105
 - datetime_day_short_name 105
 - datetime_hour 105
 - datetime_in_seconds 105
 - datetime_minute 105
 - datetime_month 105
 - datetime_month_name 105
 - datetime_month_short_name 105
 - datetime_now datetime_second 105
 - datetime_time 105
 - datetime_timestamp 105
 - datetime_weekday 105
 - datetime_year 105
- funzioni definite dall'utente (UDF) 79
- funzioni di confronto 93
- funzioni di conversione 92
- funzioni di data 87, 88
 - date_before 93, 105
 - date_days_difference 105
 - date_in_days 105
 - date_in_months 105
 - date_in_weeks 105
 - date_in_years 105
 - date_months_difference 105
 - date_weeks_difference 105
 - date_years_difference 105
 - funzione @TODAY 105
- funzioni di data e ora 87, 88
- funzioni di distribuzione 97
- funzioni di informazione 92
- funzioni di ora 87, 88
 - time_before 93, 105
 - time_hours_difference 105
 - time_in_hours 105
 - time_in_mins 105
 - time_in_secs 105
 - time_mins_difference 105
 - time_secs_difference 105
- funzioni di probabilità 97
- funzioni di sequenza 109
- funzioni globali 113

- funzioni logiche 95
- funzioni numeriche 96
- funzioni speciali 115
- funzioni stringa 99
- funzioni trigonometriche 97
- funzione to_real 92

G

- Generatore espressioni 169
- Generazione SQL
 - anteprima 46
 - registrazione 46
- gestione degli spazi vuoti
 - funzioni CLEM 114
- gradi
 - unità di misura 44
- grafici
 - aggiunta a progetti 142
 - salvataggio dell'output 61

H

- hasssubstring, funzione 99

I

- IBM SPSS Collaboration and Deployment Services 118
- IBM SPSS Collaboration and Deployment Services Enterprise View 118
- IBM SPSS Collaboration and Deployment Services Repository 117, 118
 - archiviazione oggetti 120
 - Blocco e sblocco degli oggetti 127
 - cartelle 127, 128
 - connessione a 119
 - eliminazione di oggetti e versioni 127
 - esplorazione 119
 - proprietà oggetto 129
 - recupero di oggetti 124
 - ricerca in 125
 - single sign-on (SSO) 119
 - trasferimento di progetti in 143
- IBM SPSS Modeler 1, 15
 - cenni generali 11, 149
 - documentazione 3
 - esecuzione dalla riga di comando 11
 - funzionalità di accesso facilitato 165
 - guida introduttiva 11
 - opzioni 149
 - suggerimenti e tasti di scelta rapida 65
- IBM SPSS Modeler Advantage 118, 130
- IBM SPSS Modeler Server 1
 - ID utente 12
 - nome dominio (Windows) 12
 - nome host 12, 13
 - numero porta 12, 13
 - password 12
- icone
 - impostazione delle opzioni 22, 46
- ID utente
 - IBM SPSS Modeler Server 12
- if then else, funzioni 95

- importazione
 - PMML 138
- incolla 19
- insieme di regole
 - valutazione 42
- insiemi 42
- insiemi di categorie multiple
 - in espressioni CLEM 78
- insiemi di dicotomie multiple
 - in espressioni CLEM 78
- insiemi di risposte multiple
 - in espressioni CLEM 78, 80
- interi 85
- interrompi esecuzione 19
- introduzione 85
 - IBM SPSS Modeler 11, 149
- ISW (IBM InfoSphere Warehouse)
 - esportazione PMML 138

J

- Java 176
- JAWS 165, 174, 175, 176

K

- knowledge discovery 25

L

- lettori di schermo 167, 169, 174
 - esempio 173
- lingua
 - opzioni 149
- locale
 - opzioni 149

M

- maggiore di, operatore 93
- manager 17
- mappatura dei dati 64
- mappatura di campi 62
- memoria
 - gestione 149
- messaggi
 - visualizzazione SQL generato 46
- messaggi di errore 47
- minore di, operatore 93
- modellazione
 - ramo 54
- modelli 54, 63
 - aggiornamento 134
 - aggiunta a progetti 142
 - archiviazione in IBM SPSS
 - Collaboration and Deployment Services Repository 124
 - esportazione 153
 - sostituzione 150
- Modelli PMML
 - regressione lineare 153
 - regressione logistica 153
- modello di elaborazione CRISP-DM 27, 28

- mouse
 - utilizzo in IBM SPSS Modeler 22, 36

N

- nodi 11
 - aggiunta 36, 38
 - aggiunta a progetti 142
 - aggiunta di commenti a 55
 - annotazione 55, 59
 - anteprima dei dati 40
 - archiviazione in IBM SPSS
 - Collaboration and Deployment Services Repository 123
 - attivazione 37
 - blocco 41
 - caricamento 62
 - connessione in un flusso 36
 - creazione di palette
 - personalizzate 155
 - creazione di sottopalette
 - personalizzate 156
 - disattivazione 37, 38
 - disattivazione in un flusso 37
 - uplicazione 38
 - eliminazione 36
 - eliminazione delle connessioni 38
 - ignorare in un flusso 37
 - impostazione delle opzioni 38
 - introduzione 35
 - modifica 38
 - ordine dei 159
 - performance 161, 162
 - personalizzazione delle schede della
 - palette 156
 - ricerca di 51
 - rimozione dalla palette 155
 - salvataggio 60
 - tempi di esecuzione 48
 - visualizzazione nella palette 155
- nodi di processo 35
 - performance 161
- nodi Modelli 35
 - performance 162
 - personalizzazione della scheda della
 - palette Modelli 156
- nodi origine 35
 - aggiornamento 42
 - mappatura dei dati 63
- nodi output 35
- nodo Aggregazione
 - performance 161
- nodo Creazione regola
 - caricamento 62
- nodo Elimina duplicati
 - performance 161
- nodo Enterprise View 131
- nodo Esplora
 - utilizzo nel data mining 26
 - utilizzo nell'esplorazione 25
- nodo File cache
 - caricamento 62
- nodo finale 35
- nodo Kohonen
 - insiemi grandi 42
 - performance 162
- nodo Medie K
 - insiemi grandi 42
 - performance 162
- nodo Ordina
 - performance 161
- nodo Qualità
 - valori mancanti 68
- nodo Raccolta
 - performance 161
- nodo Rete neurale
 - insiemi grandi 42
 - performance 162
- nodo riempimento
 - valori mancanti 69
- nodo Selezione funzioni
 - valori mancanti 68
- nodo Tipo
 - performance 161
- nodo tipologia
 - valori mancanti 69
- nodo Unione
 - performance 161
- nodo Valutazione
 - performance 161
- nome dominio (Windows)
 - IBM SPSS Modeler Server 12
- nome host
 - IBM SPSS Modeler Server 12, 13
- nomi di nodo 59
- nomi di stream 59
- not, operatore 95
- notazione scientifica
 - formato di visualizzazione 44
- notifiche
 - impostazione delle opzioni 150
- nugget 54
 - definizione 17
- nugget del modello 54
- null 76
- numeri 76, 85, 86
- numeri reali 85, 86
- numero porta
 - IBM SPSS Modeler Server 12, 13
- nuove funzionalità 7

O

- oggetti
 - proprietà 145
- oggetti di output
 - archiviazione in IBM SPSS
 - Collaboration and Deployment Services Repository 123
- Operatori
 - in espressioni CLEM 79
 - unione di stringhe 92
- opzioni 149
 - per IBM SPSS Modeler 149
 - PMML 153
 - proprietà flusso 41, 42, 43, 44, 46, 47, 48
 - sintassi 153
 - utente 150
 - visualizza 152
- opzioni di distribuzione
 - scenari 131
- opzioni utente 150

or, operatore 95
output 17
output HTML
lettore di schermo 174

P

palette 15
personalizzazione 154
palette dei modelli generati 17
palette Modelli 124
parametri
calcolo del punteggio 133
creazione del modello 133
flusso 49
in espressioni CLEM 80
prompt di runtime 49
sessione 49
tipo 49
utilizzo in scenari 133
parametri di sessione 49
parametri flusso 49
parole chiave
annotazione di nodi 59
password
IBM SPSS Modeler Server 12
performance
dei nodi di elaborazione 161
dei nodi Modelli 162
espressioni CLEM 163
personalizzazione delle schede della palette 156
PMML
esportazione di modelli 138
importazione di modelli 138
opzioni di esportazione 153
precedenza 88
precedenza degli operatori 88
progetti 19, 141
aggiunta di oggetti 142
annotazione 144
archiviazione in IBM SPSS
Collaboration and Deployment
Services Repository 122
chiusura 145
creazione 142
creazione di nuovi 142
generazione di report 146
impostazione di proprietà 144
impostazione di una cartella di default 141
in IBM SPSS Collaboration and
Deployment Services
Repository 143
proprietà cartella 145
proprietà oggetto 145
Vista Classi 142
visualizzazione CRISP-DM 141
programmazione visuale 15
prompt, di runtime 49
proprietà
cartella progetto 145
fasi dei report 146
per stream di dati 41
proprietà flusso
Analytic Server 47

proprietà oggetto, IBM SPSS
Collaboration and Deployment Services
Repository 129
pulsante centrale del mouse
simulazione 22, 36
punto 42

R

radianti
unità di misura 44
rami, creazione di modelli e calcolo del
punteggio 54, 133, 135
record 25
valori mancanti 68
recupero di oggetti da IBM SPSS
Collaboration and Deployment Services
Repository 124
regressione 174
regressione lineare
esportazione come PMML 153
regressione logistica 174
esportazione come PMML 153
report
aggiunta a progetti 142
generazione 146
impostazione di proprietà 146
salvataggio dell'output 61
ricerca
di nodi in un flusso 51
ricerca di connessioni in COP 13
ricerca di oggetti in IBM SPSS
Collaboration and Deployment Services
Repository 125
ridenominazione
flussi 52
nodi 59
ridimensionamento 21
ridimensionamento dei flussi 22
riduzione a icona 21
riga di comando
avvio di IBM SPSS Modeler 11
rumore nei dati 26

S

salvataggio
flussi 60
nodi 60
oggetti di output 61
più oggetti 61
stati 61
sblocco oggetti IBM SPSS Collaboration
and Deployment Services
Repository 127
scelte rapide
tastiera 22, 166, 167, 169, 170, 171
utilizzo generale 65
scelte rapide da tastiera 166, 167, 169,
170, 171
scenari 130
definizione 118
opzioni di distribuzione 131
scorrimento
impostazione delle opzioni 46
script 24, 71

selezione della palette dei nodi 155
server
accesso a 12
aggiunta di connessioni 13
directory di default 150
ricerca di server in COP 13
sessioni IBM SPSS Modeler multiple 14
simbolo decimale
formati di visualizzazione numeri 42
simbolo di raggruppamento
formati di visualizzazione numeri 42
single sign-on, IBM SPSS Collaboration
and Deployment Services
Repository 117, 119
single sign-on (SSO) 12
sistema
opzioni 149
sostituzione di modelli 150
sottopalette
creazione 156
rimozione dalla scheda della
palette 155
visualizzazione nella scheda della
palette 155
sovrapposte 25
spazi
rimozione da stringhe 75, 99
spazio vuoto
rimozione da stringhe 75, 99
stampa 24
flussi 22, 38
stati
caricamento 62
salvataggio 61
stream di dati
creazione 35
stringhe 85, 86
corrispondenza 75
manipolazione in espressioni
CLEM 75
sostituzione 75
strumento per la mappatura dei dati 62,
63
strutture ad albero delle decisioni
accesso facilitato 174
suggerimenti
per l'accesso facilitato 175
utilizzo generale 65
Supporto Unicode 177

T

tabelle 169
aggiunta a progetti 142
salvataggio dell'output 61
taglia 19
tasti di scelta rapida 22
tempi di esecuzione, visualizzazione 48
tipi di dati 73
nei parametri 49
tipo di distribuzione 131

U

uguale a, operatore 93

V

- valori 73
 - aggiunta a espressioni CLEM 80
 - visualizzazione da un report del nodo Esplora 80
- valori data/ora 76
- valori globali
 - in espressioni CLEM 80
- valori mancanti 68, 76
 - espressioni CLEM 69
 - in record 68
- verifica delle espressioni CLEM 81
- virgola 42
- Vista IBM SPSS Collaboration and Deployment Services Enterprise 131
- vuoti 68, 76

Z

- Zoom indietro da un Supernodo 19



Printed in Ireland