

IBM SPSS Modeler 15

Benutzerhandbuch



Hinweis: Lesen Sie zunächst die allgemeinen Informationen unter Hinweise auf S. 280, bevor Sie dieses Informationsmaterial sowie das zugehörige Produkt verwenden.

Diese Ausgabe bezieht sich auf IBM SPSS Modeler 15 und alle nachfolgenden Versionen sowie Anpassungen, sofern dies in neuen Ausgaben nicht anders angegeben ist.

Screenshots von Adobe-Produkten werden mit Genehmigung von Adobe Systems Incorporated abgedruckt.

Screenshots von Microsoft-Produkten werden mit Genehmigung der Microsoft Corporation abgedruckt.

Lizenziertes Material - Eigentum von IBM

© Copyright IBM Corporation 1994, 2012.

Eingeschränkte Rechte für Benutzer der US-Regierung: Verwendung, Vervielfältigung und Veröffentlichung eingeschränkt durch GSA ADP Schedule Contract mit der IBM Corp.

Vorwort

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

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

Über IBM Business Analytics

IBM Business Analytics-Software bietet vollständige, einheitliche und genaue Informationen, auf die Entscheidungsträger vertrauen, um die Unternehmensleistung zu steigern. Ein umfassendes Portfolio von Anwendungen für [Unternehmensinformationen](#), [Vorhersageanalysen](#), [Verwaltung der Finanzleistung und Strategie](#) sowie [Analysen](#) bietet sofort klare und umsetzbare Einblicke in die aktuelle Leistung und ermöglicht die Vorhersage zukünftiger Ergebnisse. In Kombination mit umfassenden Branchenlösungen, bewährten Vorgehensweisen und professionellen Dienstleistungen können Unternehmen jeder Größe optimale Produktivität erreichen, die Entscheidungsfindung zuverlässig automatisieren und bessere Ergebnisse erzielen.

Als Teil dieses Portfolios unterstützt die IBM SPSS Predictive Analytics-Software Unternehmen dabei, zukünftige Ereignisse vorherzusagen und aktiv auf diese Erkenntnisse zu reagieren, um bessere Geschäftsergebnisse zu erzielen. Kunden aus den Bereichen Wirtschaft, Behörden und Bildung aus aller Welt verlassen sich auf die IBM SPSS-Technologie. Sie bringt Ihnen beim Gewinnen, Halten und Ausbauen neuer Kundenbeziehungen einen Wettbewerbsvorteil und verringert gleichzeitig das Betrug- sowie andere Risiken. Durch Integration der IBM SPSS-Software in den täglichen Betrieb können diese Unternehmen qualifizierte Vorhersagen treffen und dadurch die Entscheidungsfindung so ausrichten und automatisieren, dass Geschäftsziele erreicht werden und ein messbarer Wettbewerbsvorteil entsteht. Wenn Sie weitere Informationen wünschen oder einen Mitarbeiter kontaktieren möchten, ist dies unter <http://www.ibm.com/spss> möglich.

Technischer Support

Kunden mit Wartungsvertrag können den technischen Support in Anspruch nehmen. Kunden können sich an den technischen Support wenden, wenn sie Hilfe bei der Arbeit mit IBM Corp.-Produkten oder bei der Installation in einer der unterstützten Hardware-Umgebungen benötigen. Die Kontaktdaten des Technischen Supports finden Sie auf der IBM Corp.-Website

unter <http://www.ibm.com/support>. Sie müssen bei der Kontaktaufnahme Ihren Namen, Ihre Organisation und Ihre Supportvereinbarung angeben.

Inhalt

1 Informationen zu IBM SPSS Modeler 1

IBM SPSS Modeler-Produkte	1
IBM SPSS Modeler	1
IBM SPSS Modeler Server	2
IBM SPSS Modeler Administration Console	2
IBM SPSS Modeler Batch	2
IBM SPSS Modeler Solution Publisher	3
IBM SPSS Modeler Server-Adapter für IBM SPSS Collaboration and Deployment Services	3
IBM SPSS Modeler-Editionen	3
IBM SPSS Modeler-Dokumentation	4
SPSS Modeler Professional-Dokumentation	4
SPSS Modeler Premium-Dokumentation	6
Anwendungsbeispiele	6
Ordner "Demos"	6

2 Neue Funktionen 8

Neue und geänderte Funktionen in IBM SPSS Modeler 15	8
Neue Funktionen in IBM SPSS Modeler Professional	8
Neue Funktionen in IBM SPSS Modeler Premium	12
Neue Knoten in dieser Version	13

3 IBM SPSS Modeler-Überblick 15

Erste Schritte	15
Starten von IBM SPSS Modeler	15
Starten über die Befehlszeile	16
Verbindung mit IBM SPSS Modeler Server	16
Ändern des temporären Verzeichnisses	20
Starten mehrerer IBM SPSS Modeler-Sitzungen	21
IBM SPSS Modeler-Oberfläche auf einen Blick	21
Stream-Zeichenbereich von IBM SPSS Modeler	22
Knotenpalette	22
IBM SPSS Modeler-Manager	23
IBM SPSS Modeler-Projekte	25
IBM SPSS Modeler-Symbolleiste	26
Anpassen der Symbolleiste	27
Anpassen des IBM SPSS Modeler-Fensters	28

Ändern der Symbolgröße für einen Stream	29
Verwenden der Maus in IBM SPSS Modeler	30
Verwenden von Direktzugriffstasten	31
Drucke	32
Automatisieren von IBM SPSS Modeler	32
4 Data Mining	34
Data Mining-Überblick	34
Bewerten der Daten	36
Eine Strategie für Data Mining	37
Das Prozessmodell "CRISP-DM"	38
Typen von Modellen	40
Data Mining-Beispiele	47
5 Erstellen von Streams	48
Überblick zum Erstellen von Streams	48
Erstellen von Daten-Streams	48
Arbeiten mit Knoten	49
Arbeiten mit Streams	61
Stream-Beschreibungen	83
Ausführen von Streams	86
Arbeiten mit Modellen	87
Hinzufügen von Kommentaren und Anmerkungen zu Knoten und Streams	88
Speichern von Daten-Streams	99
Laden von Dateien	102
Zuordnen von Daten-Streams	103
Tipps und Direktzugriffstasten	108
6 Umgang mit fehlenden Werten	112
Überblick über fehlende Werte	112
Umgang mit fehlenden Werten	113
Behandeln von Datensätzen mit fehlenden Werten	114
Behandeln von Feldern mit fehlenden Werten	114
Imputieren oder Auffüllen fehlender Werte	115
CLEM-Funktionen für fehlende Werte	116

7 Erstellen von CLEM-Ausdrücken 118

Informationen zu CLEM	118
CLEM-Beispiele	121
Werte und Datentypen	123
Ausdrücke und Bedingungen	124
Stream-, Sitzungs- und Superknoten-Parameter	125
Arbeiten mit Zeichenketten	126
Umgang mit Leerzeichen und fehlenden Werten	127
Arbeiten mit Zahlen	128
Arbeiten mit Zeit- und Datumsangaben	128
Zusammenfassen mehrerer Felder	129
Arbeiten mit Mehrfachantwortdaten	130
Expression Builder	131
Zugriff auf Expression Builder	133
Erstellen von Ausdrücken	133
Auswählen von Funktionen	134
Auswählen von Feldern, Parametern und globalen Variablen	135
Anzeigen oder Auswählen von Werten	136
Überprüfen von CLEM-Ausdrücken	137
Suchen und ersetzen	138

8 CLEM-Sprachreferenz 141

CLEM-Referenz – Überblick	141
CLEM-Datentypen	141
Ganze Zahlen	142
Reelle Zahlen	142
Zeichen	143
Zeichenketten	143
Listen	143
Felder	143
Datumsangaben	144
Zeit	145
CLEM Operatoren	145
Funktionsreferenz	148
Konventionen bei Funktionsbeschreibungen	149
Informationsfunktionen	149
Konvertierungsfunktionen	150
Vergleichsfunktionen	151

Logische Funktionen	154
Numerische Funktionen	154
Trigonometrische Funktionen	156
Wahrscheinlichkeitsfunktionen	156
Bitweise Operationen mit ganzen Zahlen	157
Zufallsfunktionen	158
String-Funktionen	159
SoundEx-Funktionen	164
Datums- und Uhrzeitfunktionen	165
Sequenzfunktionen	170
Globale Funktionen	175
Funktionen zum Umgang mit Leerstellen und Nullwerten	176
Sonderfelder	177

9 Verwenden von IBM SPSS Modeler mit einem Repository 180

Informationen zu IBM SPSS Collaboration and Deployment Services Repository	180
Speichern und Bereitstellen von Repository-Objekten	182
Verbinden mit dem Repository	183
Eingabe der Anmeldedaten für das Repository	184
Durchsuchen des Repository-Inhalts	185
Speichern von Objekten im Repository	186
Festlegen der Objekteigenschaften	186
Speichern von Streams	192
Temporäres Speichern von Projekten	192
Speichern von Knoten	193
Speichern von Ausgabeobjekten	193
Speichern von Modellen und Modellpaletten	194
Abrufen von Objekten aus dem Repository	194
Wählen eines abzurufenden Objekts	196
Auswählen einer Objektversion	197
Suchen nach Objekten im Repository	197
Ändern von Repository-Objekten	200
Erstellen, Umbenennen und Löschen von Ordnern	200
Sperren und Entsperren von Repository-Objekten	201
Löschen von Repository-Objekten	201
Verwalten der Eigenschaften von Repository-Objekten	202
Anzeigen von Ordnereigenschaften	202
Anzeigen und Bearbeiten von Objekteigenschaften	204
Verwalten von Objektversionsbeschriftungen	206

Bereitstellen von Streams.....	207
Deployment-Optionen für den Stream	208
Scoring-Verzweigung	212
10 Exportieren in externe Anwendungen	219
Öffnen eines Streams in IBM SPSS Modeler Advantage.....	219
Importieren und Exportieren von Modellen als PMML.....	220
Modelltypen, die PMML unterstützen.....	223
11 Projekte und Berichte	225
Einführung in Projekte	225
CRISP-DM-Ansicht	226
Klassenansicht	227
Erstellen von Projekten.....	227
Erstellen eines neuen Projekts.....	228
Hinzufügen zu einem Projekt	228
Übertragen von Projekten an das IBM SPSS Collaboration and Deployment Services Repository.....	230
Festlegen der Projekteigenschaften.....	231
Erstellen von Anmerkungen für ein Projekt.....	232
Objekteigenschaften	234
Schließen von Projekten	235
Generieren von Berichten.....	235
Speichern und Exportieren von generierten Berichten.....	239
12 Anpassen von IBM SPSS Modeler	241
Anpassen von IBM SPSS Modeler-Optionen.....	241
Festlegen von IBM SPSS Modeler-Optionen.....	241
Systemoptionen	241
Festlegen von Standardverzeichnissen	243
Festlegen von Benutzeroptionen	243
Festlegen von Benutzerinformationen	249

Anpassen der Knotenpalette.....	250
Anpassen des Paletten-Managers.....	250
Ändern der Ansicht einer Palettenregisterkarte.....	256
CEMI-Knotenverwaltung.....	256

13 Erläuterungen zur Leistung für Streams und Knoten 257

Knotenreihenfolge	257
Knoten-Caches.....	258
Leistung: Prozessknoten.....	260
Leistung: Modellierungsknoten.....	262
Leistung: CLEM-Ausdrücke	262

Anhänge

A Eingabehilfen in IBM SPSS Modeler 264

Übersicht über die Eingabehilfen in IBM SPSS Modeler	264
Arten der Unterstützung von Eingabehilfen.....	264
Zugriffsmöglichkeiten für Sehbehinderte	264
Zugriffsmöglichkeiten für Blinde	265
Tastatureingabehilfen	266
Verwenden eines Bildschirm-Lesesystems	275
Tipps zur Verwendung	276
Kompatibilität mit anderer Software.....	277
JAWS und Java	278
Verwenden von Diagrammen in IBM SPSS Modeler	278

B Unicode-Unterstützung 279

Unicode-Unterstützung in IBM SPSS Modeler	279
---	-----

C Hinweise	280
Index	283

Informationen zu IBM SPSS Modeler

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

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

SPSS Modeler kann als Standalone-Produkt oder als Client in Verbindung mit SPSS Modeler Server erworben werden. Außerdem ist eine Reihe von Zusatzoptionen verfügbar, die in den folgenden Abschnitten kurz dargelegt werden. Weitere Informationen finden Sie unter <http://www.ibm.com/software/analytics/spss/products/modeler/>.

IBM SPSS Modeler-Produkte

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

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

IBM SPSS Modeler

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

Mit SPSS Modeler können Sie schnell und intuitiv genaue Vorhersagemodelle erstellen, und das ohne Programmierung. Mithilfe der speziellen visuellen Benutzeroberfläche können Sie ganz einfach den Data Mining-Prozess visualisieren. Mit der Unterstützung der in das Produkt

eingebetteten erweiterten Analyseprozesse können Sie zuvor verborgene Muster und Trends in Ihren Daten aufdecken. Sie können Ergebnisse modellieren und Einblick in die Faktoren gewinnen, die Einfluss auf diese Ergebnisse haben, wodurch Sie in die Lage versetzt werden, Geschäftschancen zu nutzen und Risiken abzuschwächen.

SPSS Modeler ist in zwei Editionen erhältlich: SPSS Modeler Professional und SPSS Modeler Premium. Für weitere Informationen siehe Thema IBM SPSS Modeler-Editionen auf S. 3.

IBM SPSS Modeler Server

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

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

IBM SPSS Modeler Administration Console

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

IBM SPSS Modeler Batch

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

IBM SPSS Modeler Solution Publisher

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

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

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

IBM SPSS Modeler-Editionen

SPSS Modeler ist in den folgenden Editionen erhältlich.

SPSS Modeler Professional

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

SPSS Modeler Premium

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

IBM® SPSS® Modeler Entity Analytics fügt eine völlig neue Dimension zu den IBM® SPSS® Modeler-Vorhersageanalysen hinzu. Während bei Vorhersageanalysen versucht wird, zukünftiges Verhalten aus früheren Daten vorherzusagen, liegt der Schwerpunkt bei der Entitätsanalyse auf der Verbesserung von Kohärenz und Konsistenz der aktuellen Daten, indem Identitätskonflikte innerhalb der Datensätze selbst aufgelöst werden. Bei der Identität kann es sich um die Identität einer Person, einer Organisation, eines Objekts oder einer anderen Entität handelt, bei der Unklarheiten bestehen könnten. Die Identitätsauflösung kann in einer Reihe von Bereichen

entscheidend sein, darunter Customer Relationship Management, Betrugserkennung, Bekämpfung der Geldwäsche sowie nationale und internationale Sicherheit.

IBM SPSS Modeler Social Network Analysis transformiert Informationen zu Beziehungen in Felder, die das Sozialverhalten von Einzelpersonen und Gruppen charakterisieren. Durch die Verwendung von Daten, die die Beziehungen beschreiben, die sozialen Netzwerken zugrunde liegen, ermittelt IBM® SPSS® Modeler Social Network Analysis Führer in sozialen Netzwerken, die das Verhalten anderer Personen im Netzwerk beeinflussen. Außerdem können Sie feststellen, welche Personen am meisten durch andere Teilnehmer im Netzwerk beeinflusst werden. Durch die Kombination dieser Ergebnisse mit anderen Maßzahlen können Sie aussagekräftige Profile für Einzelpersonen, die Sie als Grundlage für Ihre Vorhersagemodele verwenden können. Modelle, die diese sozialen Informationen berücksichtigen, sind leistungsstärker als Modelle, die dies nicht tun.

Text Analytics for IBM® SPSS® Modeler verwendet hoch entwickelte linguistische Technologien und die Verarbeitung natürlicher Sprache (Natural Language Processing, NLP), um eine schnelle Verarbeitung einer großen Vielfalt an unstrukturierten Textdaten zu ermöglichen, um die Schlüsselkonzepte zu extrahieren und zu ordnen und um diese Konzepte in Kategorien zusammenzufassen. Extrahierte Konzepte und Kategorien können mit bestehenden strukturierten Daten, beispielsweise demografischen Informationen, kombiniert und mithilfe der vollständigen Suite der Data-Mining-Tools von SPSS Modeler auf die Modellierung angewendet werden, um bessere und fokussiertere Entscheidungen zu ermöglichen.

IBM SPSS Modeler-Dokumentation

Dokumentation im Online-Hilfe-Format finden Sie im Hilfe-Menü von SPSS Modeler. Dazu gehören die Dokumentation für SPSS Modeler, SPSS Modeler Server und SPSS Modeler Solution Publisher sowie das Anwendungshandbuch und weiteres Material zur Unterstützung.

Die vollständige Dokumentation für die einzelnen Produkte (einschließlich Installationsanweisungen) steht im PDF-Format im Ordner *\Documentation* auf der jeweiligen Produkt-DVD zur Verfügung. Installationsdokumente können auch aus dem Internet unter <http://www-01.ibm.com/support/docview.wss?uid=swg27023172> heruntergeladen werden:

Dokumentation in beiden Formaten steht auch im SPSS Modeler Information Center unter <http://publib.boulder.ibm.com/infocenter/spssmodl/v15r0m0/> zur Verfügung.

SPSS Modeler Professional-Dokumentation

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

- **IBM SPSS Modeler-Benutzerhandbuch.** Allgemeine Einführung in die Verwendung von SPSS Modeler, in der u. a. die Erstellung von Daten-Streams, der Umgang mit fehlenden Werten, die Erstellung von CLEM-Ausdrücken, die Arbeit mit Projekten und Berichten sowie das Packen von Streams für das Deployment in IBM SPSS Collaboration and Deployment Services, Predictive Applications (Prognoseanwendungen) oder IBM SPSS Modeler Advantage beschrieben werden.

- **Quellen-, Prozess- und Ausgabeknoten in IBM SPSS Modeler.** Beschreibung aller Knoten, die zum Lesen, zum Verarbeiten und zur Ausgabe von Daten in verschiedenen Formaten verwendet werden. Im Grunde sind sie alle Knoten, mit Ausnahme der Modellierungsknoten.
- **IBM SPSS Modeler Modellierungsknoten.** Beschreibungen sämtlicher für die Erstellung von Data Mining-Modellen verwendeter Knoten. IBM® SPSS® Modeler bietet eine Vielzahl von Modellbildungsmethoden, die aus dem maschinellen Lernen, der künstlichen Intelligenz und der Statistik stammen. Für weitere Informationen siehe Thema [Überblick über Modellierungsknoten in Kapitel 3 in IBM SPSS Modeler 15 Modellierungsknoten](#).
- **IBM SPSS Modeler-Algorithmushandbuch.** Beschreibung der mathematischen Grundlagen der in SPSS Modeler verwendeten Modellierungsmethoden. Dieses Handbuch steht nur im PDF-Format zur Verfügung.
- **IBM SPSS Modeler-Anwendungshandbuch.** Die Beispiele in diesem Handbuch bieten eine kurze, gezielte Einführung in bestimmte Modellierungsmethoden und -verfahren. Eine Online-Version dieses Handbuchs kann auch über das Hilfe-Menü aufgerufen werden. Für weitere Informationen siehe Thema [Anwendungsbeispiele auf S. 6](#).
- **Skripterstellung und Automatisierung in IBM SPSS Modeler.** Informationen zur Automatisierung des Systems über Skripterstellung, einschließlich der Eigenschaften, die zur Bearbeitung von Knoten und Streams verwendet werden können.
- **IBM SPSS Modeler Deployment-Handbuch.** Informationen zum Ausführen von SPSS Modeler-Streams und -Szenarien als Schritte bei der Verarbeitung von Jobs im IBM® SPSS® Collaboration and Deployment Services Deployment Manager.
- **IBM SPSS Modeler CLEF-Entwicklerhandbuch.** CLEF bietet die Möglichkeit, Drittanbieterprogramme, wie Datenverarbeitungsroutinen oder Modellierungsalgorithmen, als Knoten in SPSS Modeler zu integrieren.
- **In-Database Mining-Handbuch für IBM SPSS Modeler.** Informationen darüber, wie Sie Ihre Datenbank dazu einsetzen, die Leistung zu verbessern, und wie Sie die Palette der Analysefunktionen über Drittanbieteralgorithmen erweitern.
- **IBM SPSS Modeler Server-Verwaltungs- und -Leistungshandbuch.** Informationen zur Konfiguration und Verwaltung von IBM® SPSS® Modeler Server.
- **IBM SPSS Modeler Administration Console – Benutzerhandbuch.** Informationen zur Installation und Nutzung der Konsolen-Benutzeroberfläche zur Überwachung und Konfiguration von SPSS Modeler Server. Die Konsole ist als Plugin für die Deployment Manager-Anwendung implementiert.
- **IBM SPSS Modeler Solution Publisher-Handbuch.** SPSS Modeler Solution Publisher ist eine Zusatzkomponente, mit der Unternehmen Streams zur Verwendung außerhalb der SPSS Modeler-Standardumgebung veröffentlichen können.
- **IBM SPSS Modeler-Handbuch zu CRISP-DM.** Schritt-für-Schritt-Anleitung für das Data Mining mit SPSS Modeler unter Verwendung der CRISP-DM-Methode.
- **IBM SPSS Modeler Batch-Benutzerhandbuch.** Vollständiges Handbuch für die Verwendung von IBM SPSS Modeler im Batch-Modus, einschließlich Details zur Ausführung des Batch-Modus und zu Befehlszeilenargumenten. Dieses Handbuch steht nur im PDF-Format zur Verfügung.

SPSS Modeler Premium-Dokumentation

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

- **IBM SPSS Modeler Entity Analytics – Benutzerhandbuch.** Information zur Verwendung von Entitätsanalysen mit SPSS Modeler, unter Behandlung von Repository-Installation und -Konfiguration, Entity Analytics-Knoten und Verwaltungsaufgaben.
- **IBM SPSS Modeler Social Network Analysis – Benutzerhandbuch.** Ein Handbuch zur Durchführung sozialer Netzwerkanalyse mit SPSS Modeler, einschließlich Gruppenanalyse und Diffusionsanalyse.
- **Text Analytics for SPSS Modeler – Benutzerhandbuch.** Informationen zur Verwendung von Textanalysen mit SPSS Modeler, unter Behandlung der Text Mining-Knoten, der interaktiven Workbench sowie von Vorlagen und anderen Ressourcen.
- **Text Analytics for IBM SPSS Modeler Administration Console – Benutzerhandbuch.** Informationen zur Installation und Nutzung der Konsolen-Benutzeroberfläche zur Überwachung und Konfiguration von IBM® SPSS® Modeler Server für die Verwendung mit Text Analytics for SPSS Modeler. Die Konsole ist als Plugin für die Deployment Manager-Anwendung implementiert.

Anwendungsbeispiele

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

Sie können auf die Beispiele zugreifen, indem Sie im Menü “Hilfe” in SPSS Modeler auf die Option Anwendungsbeispiele klicken. Die Datendateien und Beispiel-Streams wurden im Ordner *Demos*, einem Unterordner des Produktinstallationsverzeichnisses, installiert. [Für weitere Informationen siehe Thema Ordner “Demos” auf S. 6.](#)

Beispiele für die Datenbank-Modellierung. Die Beispiele finden Sie im *IBM SPSS Modeler In-Database Mining-Handbuch*.

Skriptbeispiele. Die Beispiele finden Sie im *IBM SPSS Modeler Handbuch für die Skripterstellung und Automatisierung*.

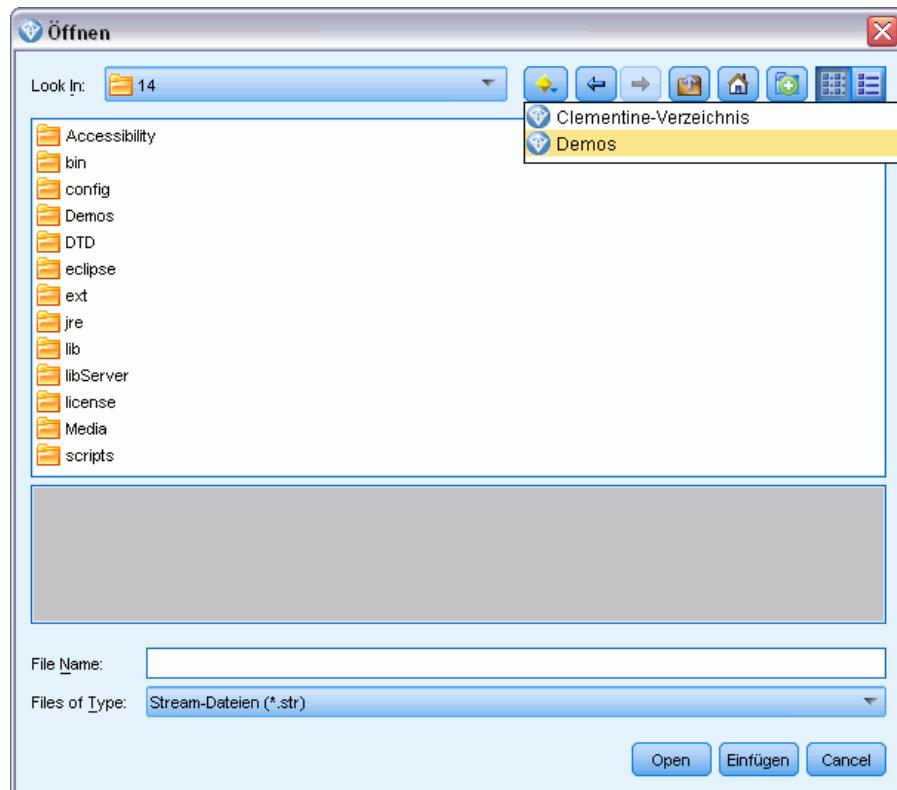
Ordner “Demos”

Die in den Anwendungsbeispielen verwendeten Datendateien und Beispiel-Streams wurden im Ordner *Demos*, einem Unterordner des Produktinstallationsverzeichnisses, installiert. Auf diesen Ordner können Sie auch über die Programmgruppe IBM SPSS Modeler 15 im Windows-Startmenü

oder durch Klicken auf *Demos* in der Liste der zuletzt angezeigten Verzeichnisse im Dialogfeld "Datei öffnen" zugreifen.

Abbildung 1-1

Auswahl des Ordners "Demos" in der Liste der zuletzt angezeigten Verzeichnisse



New Funktionen

New und geänderte Funktionen in IBM SPSS Modeler 15

Ab dieser Version gibt es folgende Editionen von IBM® SPSS® Modeler.

- **IBM® SPSS® Modeler Professional** ist der neue Name für das bestehende SPSS Modeler-Produkt.
- **IBM® SPSS® Modeler Premium** ist ein Produkt, für das eine gesonderte Lizenz erforderlich ist und das zusätzliche Funktionen zu den von SPSS Modeler Professional bereitgestellten bietet.

Die neuen Funktionen für diese Editionen werden in den folgenden Abschnitten beschrieben.

New Funktionen in IBM SPSS Modeler Professional

Bei der Edition IBM® SPSS® Modeler Professional kommen in dieser Version folgende Funktionen hinzu:

GLMM-Modellierungsknoten. Verallgemeinerte lineare gemischte Modelle (GLMMs) erweitern das lineare Modell wie folgt: Das Ziel steht über eine angegebene Verknüpfungsfunktion in einer linearen Beziehung zu den Faktoren und Kovariaten, das Ziel kann eine von der Normalverteilung abweichende Verteilung aufweisen und es kann eine Korrelation zwischen den Beobachtungen bestehen. Verallgemeinerte lineare gemischte Modelle decken eine breite Palette verschiedener Modelle ab, von einfacher linearer Regression bis hin zu komplexen Mehrebenenmodellen für nicht normalverteilte Longitudinaldaten. [Für weitere Informationen siehe Thema Neue Knoten in dieser Version auf S. 13.](#)

Unterstützung für Karten im Diagrammtafelknoten Der Diagrammtafelknoten beinhaltet nun Unterstützung für eine große Anzahl an Kartentypen. Dazu gehören Choroplethenkarten (in denen Bereiche unterschiedliche Farben oder Muster erhalten können, um verschiedene Werte anzuzeigen) sowie Punktüberlagerungskarten (bei denen raumbezogene Punkte als Überlagerung auf der Karte angezeigt werden). [Für weitere Informationen siehe Thema Erstellen von Kartenvizualisierungen in Kapitel 5 in IBM SPSS Modeler 15 Quellen-, Prozess- und Ausgabeknoten.](#)

Im Lieferumfang von IBM® SPSS® Modeler sind mehrere Kartendateien enthalten. Sie können jedoch auch das Dienstprogramm zur Konvertierung verwenden, um Ihre vorhandenen Karten-Shapedateien für die Verwendung mit der Vorlagenauswahl für Diagrammtafeln zu konvertieren. [Für weitere Informationen siehe Thema Konvertieren und Verteilen von Shapefiles für Karten in Kapitel 5 in IBM SPSS Modeler 15 Quellen-, Prozess- und Ausgabeknoten.](#)

Knoten "Netezza-Zeitreihe" und "Netezza Allgemeines lineares Modell". Für IBM® Netezza® Analytics In-Database Mining sind zwei neue Knoten verfügbar: "Zeitreihe" und "Allgemeines lineares Modell". [Für weitere Informationen siehe Thema Neue Knoten in dieser Version auf S. 13.](#)

Netezza-Knoten über Hilfsprogramme aktiviert. Die Netezza Analytics-Datenbank-Modellierungsknoten sind nun auf dieselbe Weise aktiviert wie die anderen Datenbankmodellierungsknoten. Für weitere Informationen siehe Thema Aktivieren der IBM Netezza Analytics-Integration in IBM SPSS Modeler in Kapitel 6 in *IBM SPSS Modeler 15 In-Database Mining-Handbuch*.

Vergrößern und Verkleinern der Stream-Ansicht. Es ist nun möglich, die gesamte Stream-Ansicht größer oder kleiner als die Standardgröße zu skalieren. Diese Funktion ist besonders nützlich, um einen allgemeinen Überblick über einen komplexen Stream zu erhalten bzw. um die für das Drucken eines Streams erforderliche Anzahl an Seiten zu minimieren. Für weitere Informationen siehe Thema Ändern der Symbolgröße für einen Stream in Kapitel 3 auf S. 29.

Standardeinstellungen für Datenbankverbindungen. Sie können nun Standardeinstellungen für Verbindungen zu SQL Server- und Oracle-Datenbanken einrichten sowie für alle Datenbanken, die bereits für IBM DB2 InfoSphere Warehouse unterstützt werden. Für weitere Informationen siehe Thema Angeben von voreingestellten Werten für eine Datenbankverbindung in Kapitel 2 in *IBM SPSS Modeler 15 Quellen-, Prozess- und Ausgabeknoten*.

Stream-Eigenschaften und Neugestaltung der Optimierung. Die Registerkarte “Optionen” im Dialogfeld “Stream-Eigenschaften” wurde neu gestaltet, um die Optionen in Kategorien zu gruppieren. Außerdem wurden die Optimierungsoptionen aus den Benutzeroptionen in die Stream-Eigenschaften verschoben. Für weitere Informationen siehe Thema Festlegen von Optionen für Streams in Kapitel 5 auf S. 62.

Knoten-Ausführungszeit. Sie können nun eine Option festlegen, um die einzelnen Ausführungszeiten für die Knoten in einem Stream anzuzeigen. Für weitere Informationen siehe Thema Anzeigen der Knoten-Ausführungszeiten in Kapitel 5 auf S. 75.

Außerdem können Sie eine Option (time_ecode_execution_log) in der Serverkonfigurationsdatei festlegen, um diese Ausführungszeiten im Meldungsprotokoll aufzuzeichnen. Für weitere Informationen siehe Thema Optionen nur in options.cfg sichtbar in Kapitel 4 in *IBM SPSS Modeler Server 15-Verwaltungs- und Leistungshandbuch*.

Stream-Parameter in SQL-Abfragen aus Datenbankquellenknoten. Sie können nun SPSS Modeler-Stream-Parameter in SQL-Abfragen einschließen, die Sie in den Datenbankquellenknoten eingeben. Für weitere Informationen siehe Thema Abfragen der Datenbank in Kapitel 2 in *IBM SPSS Modeler 15 Quellen-, Prozess- und Ausgabeknoten*.

Expression Builder unterstützt datenbankinterne Funktionen. Wenn ein Stream über einen Datenbankknoten mit einer Datenbank verknüpft ist und Sie den Expression Builder mit einem weiter unten liegenden Knoten verwenden, können Sie datenbankinterne Funktionen aus der verbundenen Datenbank direkt in den zu erstellenden Ausdruck aufnehmen. Für weitere Informationen siehe Thema Auswählen von Funktionen in Kapitel 7 auf S. 134.

Verbesserungen am IBM Cognos BI-Knoten Der Cognos BI-Quellenknoten unterstützt nun den Import von Cognos-Listenberichten und Daten und unterstützt die Verwendung von Parametern und Filtern. Für weitere Informationen siehe Thema IBM Cognos BI-Quellenknoten in Kapitel 2 in *IBM SPSS Modeler 15 Quellen-, Prozess- und Ausgabeknoten*.

Bei den Cognos BI-Quellen- und -Exportknoten erkennt SPSS Modeler nun automatisch die verwendete IBM Cognos BI-Version.

Verbesserungen am Aggregatknoten. Der Aggregatknoten unterstützt nun mehrere neue Aggregationsmodi für Aggregatfelder: Median, Anzahl, Varianz sowie erstes und drittes Quartil. Für weitere Informationen siehe Thema [Festlegen der Optionen für den Aggregatknoten in Kapitel 3 in IBM SPSS Modeler 15 Quellen-, Prozess- und Ausgabeknoten](#).

Zusammenführungsknoten unterstützt bedingte Zusammenführung. Sie können nun Zusammenführungen von Eingabedatensätzen durchführen, von der Erfüllung einer bestimmten Bedingung abhängen. Sie können die Bedingung direkt im Knoten angeben, oder die Bedingung mithilfe des Expression Builder erstellen. Für weitere Informationen siehe Thema [Angeben von Bedingungen für das Zusammenführen in Kapitel 3 in IBM SPSS Modeler 15 Quellen-, Prozess- und Ausgabeknoten](#).

Verbesserungen an In-Database Mining-Knoten für IBM DB2 InfoSphere Warehouse. Für In-Database Mining mit IBM DB2 InfoSphere Warehouse unterstützt der ISW-Knoten für Clustering nun zusätzlich zum demografischen Clustering und zum Kohnen-Clustering auch den erweiterten BIRCH-Algorithmus. Für weitere Informationen siehe Thema [ISW Clustering in Kapitel 5 in IBM SPSS Modeler 15 In-Database Mining-Handbuch](#). Zusätzlich bietet der ISW-Assoziationsknoten eine Auswahl an Layoutoptionen für Nicht-Transaktionsdaten (Tabellendaten). Für weitere Informationen siehe Thema [Feldoptionen für ISW-Assoziationen in Kapitel 5 in IBM SPSS Modeler 15 In-Database Mining-Handbuch](#).

Tabellenkomprimierung für Datenbankexport. Beim Exportieren in eine Datenbank können Sie nun Tabellenkomprimierungsoptionen für Verbindungen zu SQL Server- und Oracle-Datenbanken einrichten sowie für alle Datenbanken, die bereits für IBM DB2 InfoSphere Warehouse unterstützt werden. Für weitere Informationen siehe Thema [Schemaoptionen für den Datenbankexport in Kapitel 7 in IBM SPSS Modeler 15 Quellen-, Prozess- und Ausgabeknoten](#).

Massenladen für Datenbankexport. Zusätzliche Hilfeinformationen stehen für das Massenladen von Datenbanken mithilfe eines externen Ladeprogramms zur Verfügung. Für weitere Informationen siehe Thema [Programmierung des Massenladeprogramms in Kapitel 7 in IBM SPSS Modeler 15 Quellen-, Prozess- und Ausgabeknoten](#).

Verbesserungen bei der SQL-Erzeugung. Der Aggregatknoten unterstützt nun die SQL-Erzeugung für die Datentypen Datum, Zeit, Zeitstempel und Zeichenkette (zusätzlich zu ganzen Zahlen und reellen Zahlen). Bei IBM Netezza-Datenbanken unterstützt der Stichprobenknoten die SQL-Erzeugung für einfache und komplexe Stichprobenziehung und der Klasserknoten unterstützt die SQL-Erzeugung für alle Klassiermethoden mit Ausnahme von "n-tile". Für weitere Informationen siehe Thema [Knoten zur Unterstützung der SQL-Erzeugung in Kapitel 6 in IBM SPSS Modeler Server 15-Verwaltungs- und Leistungshandbuch](#).

Modell-Scoring innerhalb von Datenbanken. Bei IBM DB2 unter z/OS-, IBM Netezza- und Teradata-Datenbanken kann für eine große Anzahl der Modell-Nuggets SQL-Pushback aktiviert werden, um Modell-Scoring (statt In-Database Mining) innerhalb der Datenbank durchzuführen. Dazu können Sie einen Scoring-Adapter in der Datenbank installieren. Wenn Sie ein Modell für einen Scoring-Adapter veröffentlichen, können Modelle die benutzerdefinierten Funktionen

(UDF, User-defined Function) der Datenbank nutzen, um das Scoring auszuführen. Für weitere Informationen siehe Thema [Veröffentlichen von Modellen für einen Scoring-Adapter in Kapitel 3 in IBM SPSS Modeler 15 Modellierungsknoten](#).

Eine neue Konfigurationsoption, `db_udf_enabled` in `options.cfg`, bewirkt, dass die Option zur SQL-Erzeugung standardmäßig UDF SQL, also SQL mit benutzerdefinierten Funktionen, generiert. Für weitere Informationen siehe Thema [Optionen nur in options.cfg sichtbar in Kapitel 4 in IBM SPSS Modeler Server 15-Verwaltungs- und Leistungshandbuch](#).

Neues Format für Datenbankverbindung im Batchmodus. Das Format für die Angabe einer Datenbankverbindung im Batchmodus besteht nun nur noch aus einem einzigen Argument, um Konsistenz mit der entsprechenden Angabe bei der Skripterstellung zu gewährleisten. Für weitere Informationen siehe Thema [Parameter-Argumente in Kapitel 7 in IBM SPSS Modeler 15 Handbuch für Skripterstellung und Automatisierung](#).

Verbesserungen bei der SPSS Statistics-Integration. Im Statistikausgabeknoten stehen zusätzliche Verfahren auf der Registerkarte “Syntax” zur Verfügung, und zwar über die Schaltfläche Dialog wählen. Das Untermenü “Regression” unterstützt nun Regression mit partiellen kleinsten Quadraten und es gibt ein neues Untermenü “Vorhersage” mit folgenden Optionen: Spektralanalyse, Sequenzdiagramme, Autokorrelationen und Kreuzkorrelationen. Weitere Informationen finden Sie in der SPSS Statistics-Dokumentation.

Die Registerkarte “Syntax” des Statistics-Ausgabeknotens beinhaltet ebenfalls eine neue Option zum Generieren eines Statistikdatei-Quellenknotens, der zum Importieren der Daten dient, die aus der Ausführung eines Streams resultieren, der den Knoten enthält. Dies ist nützlich, wenn eine Prozedur zusätzlich zur Anzeige der Ausgabe Felder, beispielsweise für Scores, in das aktive Daten-Set schreibt, da diese Felder ansonsten nicht sichtbar wären. Für weitere Informationen siehe Thema [Statistikausgabeknoten - Registerkarte “Syntax” in Kapitel 8 in IBM SPSS Modeler 15 Quellen-, Prozess- und Ausgabeknoten](#).

Nicht-Root-Benutzer auf UNIX-Servern. Wenn bei Ihnen SPSS Modeler Server auf einem UNIX-Server installiert ist, können Sie nun SPSS Modeler Server als Nicht-Root-Benutzer installieren, konfigurieren sowie starten und stoppen, ohne dass Sie eine private Passworddatenbank benötigen. Für weitere Informationen siehe Thema [Einführung in Anhang C in IBM SPSS Modeler Server 15-Verwaltungs- und Leistungshandbuch](#).

Bereitgestellte Streams können nun auf die Modellverwaltungsfunktionen von IBM SPSS Collaboration and Deployment Services zugreifen. Wenn ein Stream in IBM SPSS Collaboration and Deployment Services als Stream bereitgestellt wird, kann er nun dieselben Modellverwaltungsfunktionen verwenden wie bei einer Bereitstellung als Szenario. Zu diesen Funktionen gehören Evaluierung, Aktualisierung, Scoring und Champion/Challenger. Für weitere Informationen siehe Thema [IBM SPSS Modeler-Jobeigenschaften – Allgemein in Kapitel 3 in IBM SPSS Modeler 15-Bereitstellungs- Handbuch](#).

Verbesserte Methode zum Ändern der ODBC-Verbindung für Jobschritte für SPSS Modeler-Streams und -Szenarien. Bei den Jobschritten für Streams und Szenarien in IBM SPSS Collaboration and Deployment Services gelten Änderungen an einer ODBC-Verbindung und den zugehörigen Anmeldedaten für alle zugehörigen Jobschritte. Dies bedeutet, dass Sie die Jobschritte nicht

mehr einzeln nacheinander ändern müssen. Für weitere Informationen siehe Thema [IBM SPSS Modeler-Jobeigenschaften – ODBC-Datenquellen in Kapitel 3 in IBM SPSS Modeler 15-Bereitstellungs- Handbuch](#).

Auswahl der Ausführungsverzweigung in bereitgestellten Streams. Bei Stream-Jobschritte in IBM SPSS Collaboration and Deployment Services können Sie nun, wenn der Stream Verzweigungen enthält, einen oder mehrere Stream-Verzweigungen für die Ausführung auswählen. Für weitere Informationen siehe Thema [IBM SPSS Modeler-Jobeigenschaften – Allgemein in Kapitel 3 in IBM SPSS Modeler 15-Bereitstellungs- Handbuch](#).

Neue Funktionen in IBM SPSS Modeler Premium

IBM® SPSS® Modeler Premium ist ein Produkt, für das eine gesonderte Lizenz erforderlich ist und das zusätzliche Funktionen zu den von IBM® SPSS® Modeler Professional bereitgestellten bietet. Bisher beinhaltete SPSS Modeler Premium lediglich Text Analytics for IBM® SPSS® Modeler. Das vollständige Funktionsspektrum von SPSS Modeler Premium lautet nun wie folgt:

- Text Analytics for SPSS Modeler
- IBM® SPSS® Modeler Entity Analytics
- IBM® SPSS® Modeler Social Network Analysis

Text Analytics for SPSS Modeler verwendet hoch entwickelte linguistische Technologien und die Verarbeitung natürlicher Sprache (Natural Language Processing, NLP), um eine schnelle Verarbeitung einer großen Vielfalt an unstrukturierten Textdaten zu ermöglichen, um die Schlüsselkonzepte zu extrahieren und zu ordnen und um diese Konzepte in Kategorien zusammenzufassen. Extrahierte Konzepte und Kategorien können mit bestehenden strukturierten Daten, beispielsweise demografischen Informationen, kombiniert und mithilfe der vollständigen Suite der Data-Mining-Tools von IBM® SPSS® Modeler auf die Modellierung angewendet werden, um bessere und fokussiertere Entscheidungen zu ermöglichen.

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

IBM SPSS Modeler Social Network Analysis transformiert Informationen zu Beziehungen in Felder, die das Sozialverhalten von Einzelpersonen und Gruppen charakterisieren. Durch die Verwendung von Daten, die die Beziehungen beschreiben, die sozialen Netzwerken zugrunde liegen, ermittelt IBM SPSS Modeler Social Network Analysis Führer in sozialen Netzwerken, die das Verhalten anderer Personen im Netzwerk beeinflussen. Außerdem können Sie feststellen, welche Personen am meisten durch andere Teilnehmer im Netzwerk beeinflusst werden. Durch die Kombination dieser Ergebnisse mit anderen Maßzahlen können Sie aussagekräftige Profile für

Einzelpersonen, die Sie als Grundlage für Ihre Vorhersagemodelle verwenden können. Modelle, die diese sozialen Informationen berücksichtigen, sind leistungsstärker als Modelle, die dies nicht tun.

Hinweis: SPSS Modeler Professional muss installiert sein, bevor Funktionen von SPSS Modeler Premium installiert werden können.

Neue Knoten in dieser Version

IBM SPSS Modeler Professional



Verallgemeinerte lineare gemischte Modelle (Generalized Linear Mixed Models; GLMM) erweitern lineare Modelle so, dass das Ziel nicht normalverteilt zu sein braucht und über eine angegebene Verknüpfungsfunktion in einer linearen Beziehung zu den Faktoren und Kovariaten steht und die Beobachtungen korreliert werden können. Verallgemeinerte lineare gemischte Modelle decken eine breite Palette verschiedener Modelle ab, von einfacher linearer Regression bis hin zu komplexen Mehrebenenmodellen für nicht normalverteilte Longitudinaldaten. Für weitere Informationen siehe Thema [GLMM-Knoten in Kapitel 10 in IBM SPSS Modeler 15 Modellierungsknoten](#).



Der Knoten “Netezza-Zeitreihe” analysiert Zeitreihendaten und kann zukünftiges Verhalten aus vergangenen Ereignissen vorhersagen. Für weitere Informationen siehe Thema [Netezza-Zeitreihe in Kapitel 6 in IBM SPSS Modeler 15 In-Database Mining-Handbuch](#).



Das Modell “Netezza Allgemeines lineares Modell” erweitert das lineare Regressionsmodell so, dass die abhängige Variable über eine angegebene Verknüpfungsfunktion in Beziehung mit den Einflussvariablen gesetzt wird. Außerdem ist es mit diesem Modell möglich, dass die abhängige Variable eine von der Normalverteilung abweichende Verteilung aufweist. Für weitere Informationen siehe Thema [Netezza Allgemeines lineares Modell in Kapitel 6 in IBM SPSS Modeler 15 In-Database Mining-Handbuch](#).

IBM SPSS Modeler Premium



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



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



The Streaming EA node compares new cases against the entity data in the repository.



Der SNA Group Analysis-Knoten erstellt ein Modell eines sozialen Netzwerks basierend auf den Eingabedaten zu den sozialen Gruppen innerhalb des Netzwerks. Diese Methode identifiziert Verknüpfungen zwischen den Gruppenmitgliedern und analysiert die Interaktionen innerhalb der Gruppen, um Leistungskennzahlen (Key Performance Indicators, KPIs) zu erstellen. Die KPIs können zur Abwanderungsprognose, Anomalieerkennung oder Identifizierung von Gruppenführern verwendet werden.



Der SNA Diffusion Analysis-Knoten erstellt ein Modell des Informationsflusses eines Gruppenmitglieds in seinem sozialen Umfeld. Einem Gruppenmitglied wird eine erste Gewichtung zugewiesen, die als allmählich sinkender Wert im Netzwerk verbreitet wird. Dieser Prozess wird fortgesetzt, bis jedes Mitglied des Netzwerks eine Gewichtung in Relation zum ursprünglichen Gruppenmitglied erhalten hat, basierend auf der Menge an Informationen, die das Mitglied erreicht hat. Die Scores der einzelnen Mitglieder werden dann direkt von diesen Gewichtungen abgeleitet. So kann ein Dienstanbieter beispielsweise Kunden identifizieren, die basierend auf ihrer Beziehung zu einem kürzlich abgewanderten Mitglied ein größeres Abwanderungsrisiko aufweisen.

IBM SPSS Modeler-Überblick

Erste Schritte

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

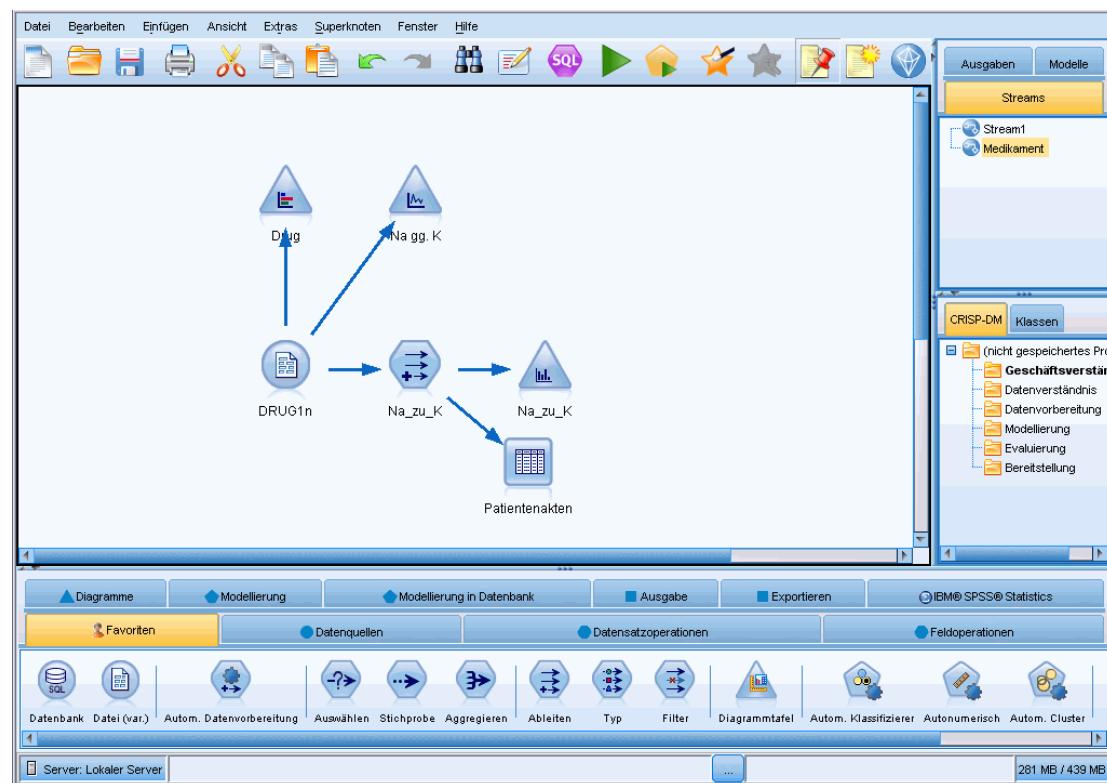
Starten von IBM SPSS Modeler

Klicken Sie zum Starten der Anwendung auf:

Start > [Alle] Programme > IBM SPSS Modeler15 > IBM SPSS Modeler15

Nach einigen Sekunden wird das Hauptfenster angezeigt.

Abbildung 3-1
IBM SPSS Modeler-Hauptfenster



Starten über die Befehlszeile

Sie können die Befehlszeile Ihres Betriebssystems wie folgt verwenden, um IBM® SPSS® Modeler zu starten:

- ▶ Öffnen Sie auf einem Computer, auf dem IBM® SPSS® Modeler installiert ist, ein DOS- oder Befehlszeilenfenster.
- ▶ Um die SPSS Modeler-Schnittstelle im interaktiven Modus zu starten, geben Sie den Befehl `modelerclient` und dann die gewünschten Argumente ein, z. B.:

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

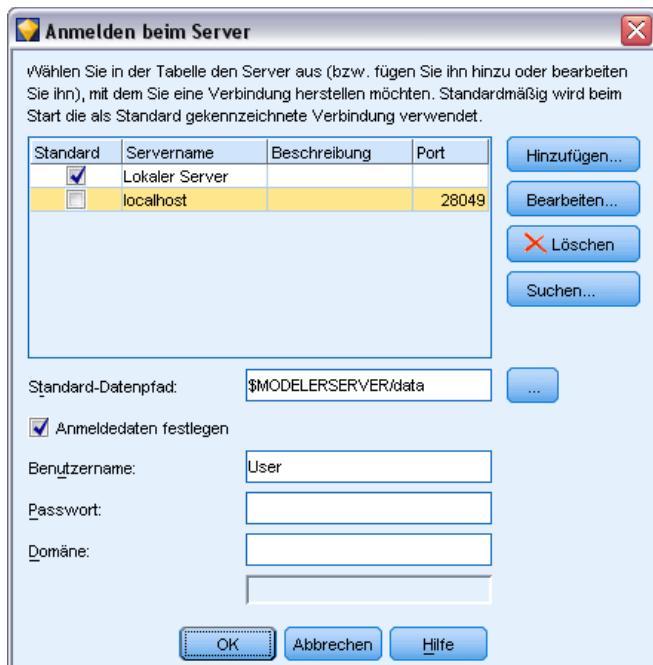
Mithilfe der verfügbaren Argumente (Flags) können Sie eine Verbindung zu einem Server herstellen, Streams laden, Skripts ausführen oder je nach Bedarf weitere Parameter angeben.

Verbindung mit IBM SPSS Modeler Server

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

Wenn Sie eine Verbindung zu einem Server herstellen möchten, können Sie den Namen des Servers, mit dem eine Verbindung hergestellt werden soll, manuell eingeben oder einen zuvor definierten Namen auswählen. Wenn Sie IBM SPSS Collaboration and Deployment Services verwenden, können Sie im Dialogfeld für die Anmeldung beim Server eine Liste von Servern bzw. Server-Clustern durchsuchen. Die Möglichkeit, die auf einem Netzwerk ausgeführten Statistics-Dienste zu durchsuchen, wird über den Coordinator of Processes bereitgestellt. Für weitere Informationen siehe Thema Lastenausgleich mit Server-Clustern in Anhang D in *IBM SPSS Modeler Server 15-Verwaltungs- und Leistungshandbuch*.

Abbildung 3-2
Dialogfeld "Anmelden beim Server"



So stellen Sie eine Verbindung mit einem Server her:

- ▶ Klicken Sie im Menü „Extras“ auf die Option Anmelden beim Server. Das Dialogfeld „Anmelden beim Server“ wird geöffnet. Alternativ können Sie auf den Bereich des Verbindungsstatus im SPSS Modeler-Fenster doppelklicken.
- ▶ Legen Sie in diesem Dialogfeld die Optionen zum Verbinden mit dem lokalen Servercomputer fest oder wählen Sie eine Verbindung in der Tabelle aus.
 - Klicken Sie auf Hinzufügen bzw. Bearbeiten, um eine Verbindung hinzuzufügen bzw. zu bearbeiten. [Für weitere Informationen siehe Thema Hinzufügen und Bearbeiten der IBM SPSS Modeler Server-Verbindung auf S. 18.](#)
 - Klicken Sie auf Suche, um auf einen Server bzw. Server-Cluster in Coordinator of Processes zuzugreifen. [Für weitere Informationen siehe Thema Suchen nach Servern in IBM SPSS Collaboration and Deployment Services auf S. 19.](#)

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

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

Anmeldedaten einstellen. Belassen Sie dieses Kontrollkästchen nicht aktiviert, um die Funktion für die **Einzelanmeldung** zu aktivieren. Diese versucht, Sie mithilfe Ihres lokalen Benutzernamens und Passworts beim Server anzumelden. Falls die Einzelanmeldung nicht möglich ist oder Sie das Kontrollkästchen zur Deaktivierung der Einzelanmeldung aktivieren (z. B. zur Anmeldung in ein Administratorkonto), wird ein weiteres Fenster angezeigt, in dem Sie aufgefordert werden, Ihre Anmeldedaten einzugeben.

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

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

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

- ▶ Klicken Sie auf OK, um die Verbindung herzustellen.

So trennen Sie eine Verbindung mit einem Server:

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

Hinzufügen und Bearbeiten der IBM SPSS Modeler Server-Verbindung

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

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

Abbildung 3-3
Dialogfeld “Anmeldung beim Server: Server hinzufügen/bearbeiten”



So fügen Sie Serververbindungen hinzu:

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

So bearbeiten Sie Serververbindungen:

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

Suchen nach Servern in IBM SPSS Collaboration and Deployment Services

Anstatt eine Serververbindung manuell einzugeben, können Sie einen im Netzwerk verfügbaren Server oder Server-Cluster über Coordinator of Processes auswählen. Diese Funktion ist in IBM® SPSS® Collaboration and Deployment Services verfügbar. Ein Server-Cluster ist eine Gruppe von Servern, aus denen Coordinator of Processes den Server ermittelt, der am besten für die Beantwortung einer Verarbeitungsanforderung geeignet ist. Für weitere Informationen siehe Thema [Lastenausgleich mit Server-Clustern in Anhang D in IBM SPSS Modeler Server 15-Verwaltungs- und Leistungshandbuch](#).

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

Hinweis: Wenn Sie keinen Zugriff auf die Funktion Coordinator of Processes haben, können Sie dennoch den Namen des Servers, mit dem eine Verbindung hergestellt werden soll, manuell eingeben oder einen zuvor definierten Namen auswählen. Für weitere Informationen siehe Thema [Hinzufügen und Bearbeiten der IBM SPSS Modeler Server-Verbindung auf S. 18.](#)

Abbildung 3-4
Dialogfeld "Nach Servern suchen"



So suchen Sie nach Servern und Clustern:

- ▶ Klicken Sie im Menü „Extras“ auf die Option Anmelden beim Server. Das Dialogfeld „Anmelden beim Server“ wird geöffnet.
- ▶ Klicken Sie in diesem Dialogfeld auf Suche, um das Dialogfeld „Nach Servern suchen“ zu öffnen. Wenn Sie versuchen, Coordinator of Processes zu durchsuchen, ohne bei IBM SPSS Collaboration and Deployment Services angemeldet zu sein, werden Sie zur Anmeldung aufgefordert. Für weitere Informationen siehe Thema [Verbinden mit dem Repository in Kapitel 9 auf S. 183.](#)
- ▶ Wählen Sie den Server bzw. Server-Cluster in der Liste aus.
- ▶ Klicken Sie auf OK, um das Dialogfeld zu schließen und diese Verbindung zu der Tabelle im Dialogfeld „Anmelden beim Server“ hinzuzufügen.

Ändern des temporären Verzeichnisses

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

- ▶ Erstellen Sie ein neues Verzeichnis namens *spss* und ein Unterverzeichnis namens *servertemp*.
- ▶ Bearbeiten Sie die Datei *options.cfg*, die sich im Unterverzeichnis */config* Ihres SPSS Modeler-Installationsverzeichnisses befindet. Bearbeiten Sie den Parameter *temp_directory* in diesem Feld, damit er wie folgt lautet: *temp_directory*, "C:/spss/servertemp".
- ▶ Anschließend müssen Sie den SPSS Modeler Server-Dienst neu starten. Dies ist möglich, wenn Sie auf die Registerkarte Dienste Ihrer Windows-Systemsteuerung klicken. Halten Sie den Dienst

einfach an und starten Sie ihn erneut, um die von Ihnen durchgeführten Änderungen zu aktivieren. Durch das Neustarten des Rechners wird auch der Dienst neu gestartet.

Alle temporären Dateien werden nun in dieses neue Verzeichnis geschrieben.

Hinweis: Der häufigste Fehler bei diesem Vorgang besteht darin, dass der falsche Schrägstrich verwendet wird. Aufgrund der UNIX-Vergangenheit von SPSS Modeler werden normale Schrägstriche (/) verwendet.

Starten mehrerer IBM SPSS Modeler-Sitzungen

Wenn Sie mehrere IBM® SPSS® Modeler-Sitzungen gleichzeitig starten möchten, müssen Sie einige Änderungen an den IBM® SPSS® Modeler- und Windows-Einstellungen vornehmen. Dies kann beispielsweise erforderlich sein, wenn Sie zwei separate Serverlizenzen besitzen und zwei Streams vom selben Client-Computer aus mit zwei verschiedenen Servern ausführen möchten.

So aktivieren Sie mehrere SPSS Modeler-Sitzungen:

- ▶ Klicken Sie auf:
Start > [Alle] Programme > IBM SPSS Modeler15
- ▶ Klicken Sie mit der rechten Maustaste auf die Verknüpfung “IBM SPSS Modeler15” (die mit dem Symbol) und wählen Sie die Option Eigenschaften.
- ▶ Fügen Sie im Textfeld Ziel die Option -noshare am Ende der Zeichenfolge hinzu.
- ▶ Wählen Sie im Windows Explorer folgende Optionen:
Werkzeuge > Ordneroptionen...
- ▶ Wählen Sie auf der Registerkarte “Dateitypen” die Option für SPSS Modeler-Streams und klicken Sie auf Erweitert.
- ▶ Wählen Sie im Dialogfeld “Dateityp bearbeiten” die Option Öffnen mit SPSS Modeler und klicken Sie auf Bearbeiten.
- ▶ Fügen Sie im Textfeld Anwendung für diesen Vorgang vor dem Argument -stream das Argument -noshare ein.

IBM SPSS Modeler-Oberfläche auf einen Blick

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

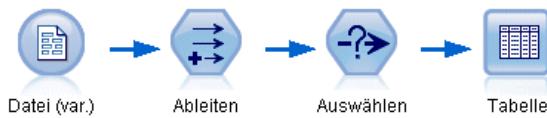
Die Datenarbeit mit SPSS Modeler besteht aus drei Schritten.

- Zunächst lesen Sie Daten in SPSS Modeler ein.

- Anschließend lassen Sie die Daten eine Reihe von Bearbeitungen durchlaufen.
- Schließlich senden Sie die Daten an ein Ziel.

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

Abbildung 3-5
Ein einfacher Stream



Stream-Zeichenbereich von IBM SPSS Modeler

Der Stream-Zeichenbereich ist der größte Bereich des IBM® SPSS® Modeler-Fensters. Hier erstellen und bearbeiten Sie Daten-Streams.

Die Erstellung von Streams erfolgt durch Zeichnen von Diagrammen von Datenoperationen, die für Ihren Geschäftsbetrieb relevant sind, auf den Hauptzeichenbereich in der Benutzeroberfläche. Jede Operation wird durch ein Symbol oder einen **Knoten** dargestellt, und die Knoten sind in einem **Stream** miteinander verbunden, der den Datenfluss durch jede Operation darstellt.

Sie können in SPSS Modeler mit mehreren Streams gleichzeitig arbeiten, entweder im selben Stream-Zeichenbereich oder durch Öffnen eines neuen Stream-Zeichenbereichs. Während einer Sitzung werden die Streams im Stream-Manager, rechts oben im SPSS Modeler-Fenster gespeichert.

Knotenpalette

Die meisten Daten und Modellierungstools in IBM® SPSS® Modeler befinden sich in der **Knotenpalette**, unten im Fenster, unterhalb des Stream-Zeichenbereichs.

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

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

Abbildung 3-6
Die Registerkarte “Datensatzoperationen” in der Knotenpalette



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

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

Je vertrauter Sie im Umgang mit SPSS Modeler werden, desto besser können Sie den Paletteninhalt für Ihre eigene Verwendung anpassen. Für weitere Informationen siehe Thema [Anpassen der Knotenpalette in Kapitel 12 auf S. 250](#).

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

IBM SPSS Modeler-Manager.

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

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

Abbildung 3-7
Registerkarte “Streams”



Die Registerkarte “Ausgaben” enthält eine Reihe von Dateien, wie beispielsweise Diagramme und Tabellen, die von den Stream-Operationen in IBM® SPSS® Modeler produziert wurden. Sie können die auf dieser Registerkarte aufgeführten Tabellen, Diagramme und Berichte anzeigen, speichern, umbenennen und schließen.

Abbildung 3-8
Registerkarte “Ausgaben”



Die Registerkarte “Modelle” ist die leistungsstärkste der Manager-Registerkarten. Diese Registerkarte enthält sämtliche Modell-Nuggets, die die in SPSS Modeler erstellten Modelle enthalten, für die aktuelle Sitzung. Auf der Registerkarte “Modelle” können diese Modelle direkt durchsucht oder dem Stream im Zeichenbereich hinzugefügt werden.

Abbildung 3-9
Registerkarte "Modelle" mit Modell-Nuggets



IBM SPSS Modeler-Projekte

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

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

Abbildung 3-10
CRISP-DM, Ansicht



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

Abbildung 3-11
Ansicht “Klassen”



IBM SPSS Modeler-Symbolleiste

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

	Neuen Stream erstellen		Stream öffnen
	Stream speichern		Aktuellen Stream drucken
	Ausschneiden & in die Zwischenablage verschieben		In Zwischenablage kopieren
	Einfügen der Auswahl		Letzte Aktion rückgängig
	Wiederholen		Knoten suchen
	Stream-Eigenschaften bearbeiten		Vorschau für SQL-Erzeugung

	Aktuellen Stream ausführen		Stream-Auswahl ausführen
	Stream anhalten (wird nur während der Stream-Ausführung aktiv)		Superknoten hinzufügen
	Vergrößern (nur Superknoten)		Verkleinern (nur Superknoten)
	Kein Markup in Stream		Kommentar einfügen
	Stream-Markup ausblenden (falls vorhanden)		Ausgeblendeten Stream-Markup einblenden
	Stream in IBM® SPSS® Modeler Advantage öffnen		

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

Weitere Informationen zu Stream-Kommentaren erhalten Sie unter [Hinzufügen von Kommentaren und Anmerkungen zu Knoten und Streams auf S. 88](#).

Weitere Informationen zur Markierung von Scoring-Verzweigungen erhalten Sie unter [Scoring-Verzweigung auf S. 212](#).

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

Anpassen der Symbolleiste

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

- ob sie angezeigt wird,
- ob für die Symbole eine QuickInfo verfügbar ist,
- ob große oder kleine Symbole angezeigt werden.

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

- ▶ Klicken Sie im Hauptmenü auf Folgendes:

Ansicht > Symbolleiste > Anzeigen

So ändern Sie die Einstellungen für QuickInfo oder Symbolgröße:

- ▶ Klicken Sie im Hauptmenü auf Folgendes:

Ansicht > Symbolleiste > Anpassen

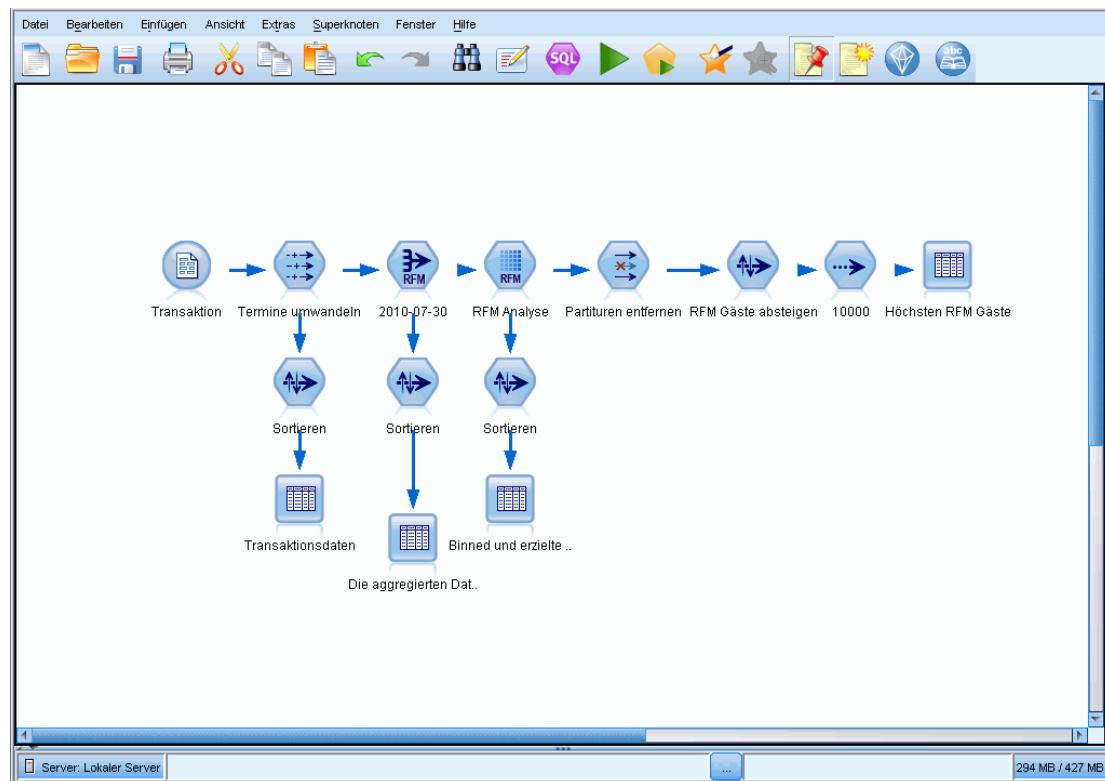
Klicken Sie wie gewünscht auf QuickInfos einblenden oder Große Symbole.

Anpassen des IBM SPSS Modeler-Fensters

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

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

Abbildung 3-12
Maximierter Stream-Zeichenbereich



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

Sie können auch die Anzeige von Bildschirm-Markup steuern, die Stream-Kommentare, Modellverknüpfungen und die Anzeige von Scoring-Verzweigungen umfasst. Um diese Anzeige ein- oder auszublenden, klicken Sie auf:

Ansicht > Stream-Markup

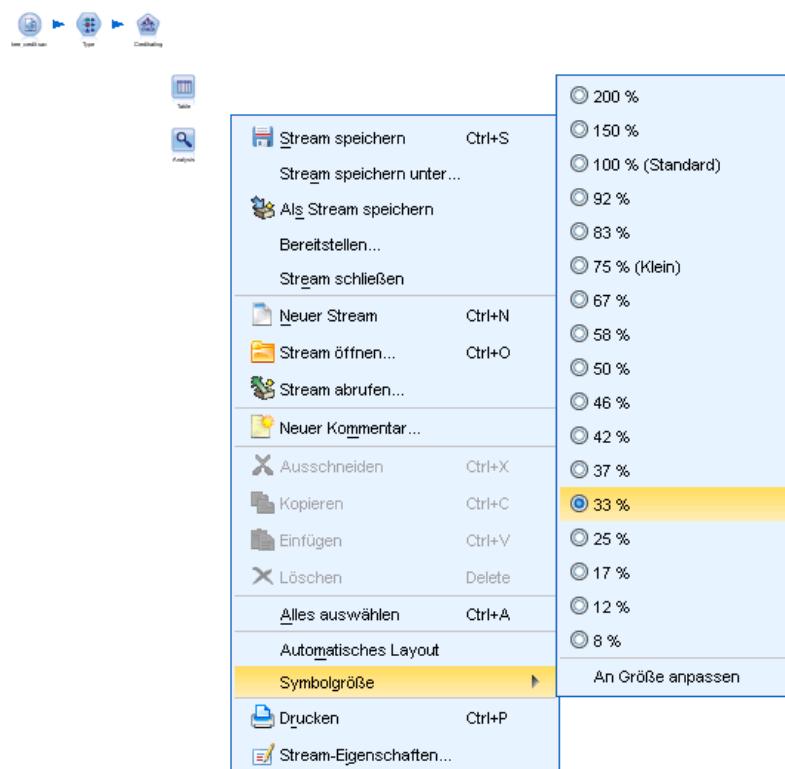
Ändern der Symbolgröße für einen Stream

Sie können die Größe der Stream-Symbole auf folgende Weisen ändern:

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

Sie können die gesamte Stream-Ansicht auf eine Anzahl an Größenwerten zwischen 8 und 200 % der Standardsymbolgröße skalieren.

Abbildung 3-13
Ändern der Symbolgröße



Skalierung des gesamten Streams (Methode mit Stream-Eigenschaften)

- ▶ Wählen Sie im Hauptmenü Folgendes:
Werkzeuge > Stream-Eigenschaften > Optionen > Layout.
- ▶ Wählen Sie im Menü “Symbolgröße” die gewünschte Größe aus.
- ▶ Klicken Sie auf Anwenden, um das Ergebnis anzuzeigen.
- ▶ Klicken Sie auf OK, um die Änderung zu speichern.

Skalierung des gesamten Streams (Methode über Menü)

- ▶ Klicken Sie mit der rechten Maustaste auf den Stream-Hintergrund im Zeichenbereich.
- ▶ Wählen Sie die Option Symbolgröße und wählen Sie die gewünschte Größe aus.

Skalierung des gesamten Streams (Methode über Tastatur)

- ▶ Drücken Sie auf der Haupttastatur die Tastenkombination “Strg + [-]”, um auf die nächstkleinere Größe umzuschalten.
- ▶ Drücken Sie auf der Haupttastatur die Tastenkombination “Strg + Umschalt + [+]”, um auf die nächstgrößere Größe umzuschalten.

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

Verwenden der Maus in IBM SPSS Modeler

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

- **Einfaches Klicken.** Verwenden Sie entweder die rechte oder die linke Maustaste, um Optionen aus den Menüs auszuwählen, Popup-Menüs zu öffnen und verschiedene andere Standardsteuerelemente und Optionen zu verwenden. Klicken Sie und halten Sie die Maustaste gedrückt, um Knoten zu verschieben und zu ziehen.
- **Doppelklicken.** Doppelklicken Sie mit der linken Maustaste, um Knoten auf dem Stream-Zeichenbereich abzulegen und die bereits vorhandenen Knoten zu bearbeiten.
- **Klicken mit der mittleren Maustaste.** Klicken Sie auf die mittlere Maustaste und ziehen Sie den Cursor, um Knoten im Stream-Zeichenbereich miteinander zu verbinden. Doppelklicken Sie auf die mittlere Maustaste, um die Verbindung eines Knotens zu lösen. Wenn Sie nicht über eine Maus mit drei Tasten verfügen, können Sie diese Funktion simulieren, indem Sie auf die ALT-Taste drücken und gleichzeitig auf die Maus klicken und ziehen.

Verwenden von Direktzugriffstasten

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

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

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

Tabelle 3-1
Unterstützte Direktzugriffstasten

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

Tabelle 3-2
Unterstützte Direktzugriffstasten für alte Tastenkombinationen

Tastenkürzel	Funktion
Strg+Alt+D	Knoten duplizieren
Strg+Alt+L	Knoten laden
Strg+Alt+R	Knoten umbenennen
Strg+Alt+U	Benutzereingabeknoten erstellen
Strg+Alt+C	Cache aktivieren/deaktivieren
Strg+Alt+F	Cache löschen
Strg+Alt+X	Superknoten erweitern

Tastenkürzel	Funktion
Strg+Alt+Z	Vergrößern/Verkleinern
Delete	Knoten oder Verbindung löschen

Drucke

Folgende Objekte können in IBM® SPSS® Modeler gedruckt werden:

- Stream-Diagramme
- Grafiken
- Tabellen
- Berichte (aus dem Berichtsknoten und den Projektberichten)
- Skripts (aus den Dialogfeldern “Stream-Eigenschaften”, “Standalone-Skript” oder “Superknoten-Skript”)
- Modelle (Modellbrowser, Registerkarten des Dialogfelds mit aktuellem Fokus, Baumansichten)
- Anmerkungen (unter Verwendung der Registerkarte “Anmerkungen” für die Ausgabe)

So drucken Sie ein Objekt:

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

Automatisieren von IBM SPSS Modeler

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

- **Control Language for Expression Manipulation** (CLEM) ist eine Sprache zur Analyse und Bearbeitung der Daten, die SPSS Modeler-Streams durchlaufen. Data-Mining-Experten verwenden CLEM in hohem Maße für Stream-Operationen, um so einfache Aufgaben durchzuführen, wie das Ableiten von Gewinn aus Kosten- und Einkommensdaten, oder um so komplexe Aufgaben durchzuführen, wie das Umwandeln von Webprotokolldaten in ein

Set von Feldern und Datensätzen mit nützlichen Informationen. [Für weitere Informationen siehe Thema Informationen zu CLEM in Kapitel 7 auf S. 118.](#)

- Die **Skripterstellung** bildet ein leistungsstarkes Tool, mit dem Prozesse in der Benutzeroberfläche automatisiert werden. Mit Skripts können dieselben Aktionen durchgeführt werden, die die Benutzer mithilfe von Maus und Tastatur durchführen. Sie können für Knoten Optionen festlegen und mithilfe einer Untergruppe von CLEM Ableitungen durchführen. Sie können auch die Ausgabe angeben und generierte Modelle bearbeiten. [Für weitere Informationen siehe Thema Skripterstellung – Überblick in Kapitel 2 in IBM SPSS Modeler 15 Handbuch für Skripterstellung und Automatisierung.](#)

Data Mining

Data Mining-Überblick

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

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

Modellierungsverfahren

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

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

Datenmanipulation und Information Discovery

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

- **Datenmanipulation.** Sie können neue Datenelemente erstellen, die aus bestehenden Daten abgeleitet sind, und die Daten in sinnvolle Untergruppen einteilen. Daten aus verschiedenen Quellen können zusammengeführt und gefiltert werden.

- **Durchsuchen und Visualisierung.** Sie können Aspekte der Daten mithilfe des Data Audit-Knotens anzeigen, um ein ursprüngliches Audit einschließlich Grafiken und Statistiken durchzuführen. Zur erweiterten Visualisierung gehören interaktive Grafiken, die zur Aufnahme in Projektberichte exportiert werden können.
- **Statistiken.** Mit Statistiken können Sie vermutete Beziehungen zwischen den Variablen in den Daten überprüfen. Die Statistikfunktionen aus IBM® SPSS® Statistics können auch in SPSS Modeler verwendet werden.
- **Hypothesentests.** Sie können Modelle für das Verhalten der Daten erstellen und überprüfen. Normalerweise verwenden Sie diese Funktionen zur Ermittlung eines erfolgversprechenden Sets von Attributen in den Daten. Diese Attribute können anschließend in die Modellierungsverfahren eingespeist werden, die wiederum versuchen, etwaige zugrunde liegende Regeln und Beziehungen zu ermitteln.

Typische Anwendungen

Typische Anwendungsgebiete für Data Mining-Verfahren sind beispielsweise:

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

Kredit-Scoring. Verwenden Sie die Kreditgeschichte einer Person als Grundlage für Kreditvergabeentscheidungen.

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

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

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

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

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

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

Terminologie

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

Bewerten der Daten

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

Stellen Sie sicher, dass die Daten verfügbar sind.

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

Überprüfen Sie, ob die Daten die relevanten Attribute abdecken.

Ziel des Data Mining ist die Ermittlung relevanter Attribute, sodass diese Überprüfung zunächst merkwürdig erscheinen mag. Es ist jedoch sinnvoll zu untersuchen, welche Daten verfügbar sind, und zu versuchen, die vermutlich relevanten Faktoren zu ermitteln, die nicht aufgezeichnet sind. Beim Versucht, den Absatz an Speiseeis vorherzusagen, liegen Ihnen eventuell umfassende Informationen über Verkaufsstellen und frühere Absatzdaten vor, möglicherweise jedoch keine Informationen zu Wetter und Temperatur, die mit großer Wahrscheinlichkeit eine signifikante Rolle spielen. Fehlende Attribute bedeuten nicht unbedingt dass das Data Mining keine nützlichen Ergebnisse hervorbringt, die Genauigkeit der resultierenden Prognosen kann jedoch eingeschränkt sein.

Eine schnelle Möglichkeit zur Einschätzung der Situation besteht in der Durchführung eines umfassenden Audits der vorliegenden Daten. Bevor Sie fortfahren, sollten Sie der Datenquelle einen Data Audit-Knoten hinzufügen und ihn ausführen, um einen vollständigen Bericht zu generieren. [Für weitere Informationen siehe Thema Data Audit-Knoten in Kapitel 6 in IBM SPSS Modeler 15 Quellen-, Prozess- und Ausgabeknoten.](#)

Vorsicht bei verrauschten Daten.

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

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

Stellen Sie sicher, dass ausreichend Daten vorliegen.

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

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

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

Kontaktieren Sie diejenigen, die sich mit den Daten auskennen.

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

Eine Strategie für Data Mining

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

- Welches inhaltliche Problem soll gelöst werden?
- Welche Datenquellen sind verfügbar und welche Teile der Daten sind für das aktuelle Problem relevant?
- Welche Art der Vorverarbeitung und Datenbereinigung ist erforderlich bevor mit dem Data Mining der Daten begonnen werden kann?
- Welche Data Mining-Verfahren sollen eingesetzt werden?
- Wie sollen die Ergebnisse der Data Mining-Analyse evaluiert werden?
- Wie können Sie die aus dem Data Mining gewonnenen Informationen möglichst optimal nutzen?

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

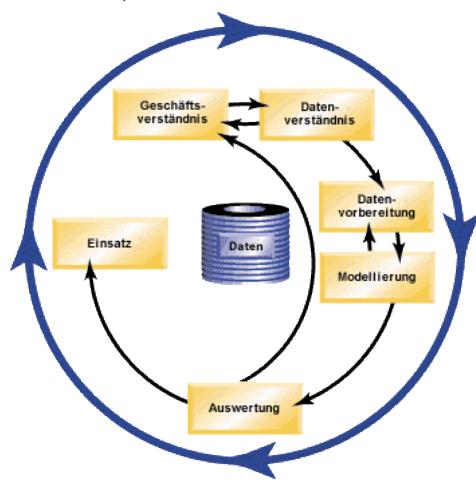
Um nicht den Überblick zu verlieren, sollten Sie über ein explizit definiertes Prozessmodell für Data Mining verfügen. Das Prozessmodell hilft Ihnen dabei, die in diesem Abschnitt aufgeführten Fragen zu beantworten, und stellt sicher, dass die wichtigen Punkte beachtet werden. Es dient als grundlegender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

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

Das Prozessmodell “CRISP-DM”

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

Abbildung 4-1
CRISP-DM, Prozessmodell



Bei den sechs Phasen handelt es sich um folgende:

- **Geschäftsverständnis.** Dies ist wahrscheinlich die wichtigste Phase des Data Mining. Das Geschäftsverständnis umfasst die Ermittlung von Geschäftszielen, die Einschätzung der Situation, die Bestimmung der Data Mining-Ziele und die Erstellung eines Projektplans.
- **Dataverständnis.** Daten sind das “Rohmaterial” beim Data Mining. Diese Phase widmet sich der Notwendigkeit, einen Einblick in die Datenressourcen und die Eigenschaften dieser Ressourcen zu gewinnen. Dazu gehört das Sammeln der ursprünglichen Daten, das Beschreiben von Daten, das Untersuchen von Daten und das Überprüfen der Datenqualität. Der Data Audit-Knoten, der in der Palette “Ausgabeknoten” verfügbar ist, ist ein unverzichtbares Tool für das Verständnis der Daten.
- **Datenvorbereitung.** Nach der Katalogisierung der Datenressourcen müssen Sie die Daten für das Data Mining vorbereiten. Zu den Vorbereitungen gehören die Auswahl, Bereinigung, Konstruktion, Integration und Formatierung der Daten.

- **Modellierung.** Dies ist natürlich der Hauptteil des Data Mining, in dem mit hoch entwickelten Analysemethoden Informationen aus den Daten extrahiert werden. Zu dieser Phase gehört die Auswahl von Modellierungsverfahren, das Generieren von Test-Designs und der Aufbau und die Auswertung von Modellen.
- **Evaluierung.** Nach der Auswahl der Modelle können Sie evaluieren, wie die Data Mining-Ergebnisse zum Erreichen Ihrer Geschäftsziele beitragen können. Zu den Elementen in dieser Phase gehören die Evaluation von Ergebnissen, die Überprüfung des Data Mining-Prozesses und die Bestimmung der nächsten Schritte.
- **Bereitstellung.** Dies ist die Phase, in der Sie erstmals praktischen Nutzen aus ihren Bemühungen ziehen können. Diese Phase konzentriert sich auf die Integration des neu erworbenen Wissens in die alltäglichen Geschäftsabläufe zur Lösung des ursprünglichen Geschäftsproblems. Zu dieser Phase gehören die Planbereitstellung, Überwachung und Anpassung, das Erstellen eines endgültigen Berichts und die Überprüfung des Projekts.

Einige Punkte in diesem Prozessmodell sind besonders wichtig. Zum einen besteht zwar die allgemeine Tendenz, dass der Prozess die Schritte in der in den vorherigen Abschnitten angegebenen Reihenfolge durchläuft, aber es gibt auch eine Reihe von Stellen, an denen die Phasen sich gegenseitig auf nichtlineare Weise beeinflussen. So geht die Datenvorbereitung in der Regel der Modellierung voraus. Die während der Modellierungsphase getroffenen Entscheidungen und gesammelten Informationen können den Benutzer häufig veranlassen, Teile der Datenvorbereitungsphase zu überdenken, was wiederum neue Modellierungsfragen aufwerfen kann. Bei den beiden Phasen findet eine gegenseitige Rückkopplung statt, bis beide Phasen adäquat gelöst wurden. Ebenso kann die Evaluationsphase zu einer Neubewertung des ursprünglichen Geschäftsverständnisses führen und Sie könnten zu dem Schluss kommen, dass Sie sich bisher mit der falschen Frage auseinandergesetzt haben. An diesem Punkt können Sie Ihr Geschäftsverständnis überarbeiten und den Rest des Prozesses mit einem besseren Ziel im Auge erneut durchführen.

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

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

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

Typen von Modellen

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

Im *SPSS Modeler-Anwendungshandbuch* finden Sie Beispiele für viele dieser Methoden sowie eine allgemeine Einführung in den Modellierungsprozess. Dieses Handbuch ist als Online-Lernprogramm und im PDF-Format verfügbar. [Für weitere Informationen siehe Thema Anwendungsbeispiele in Kapitel 1 auf S. 6.](#)

Modellierungsmethoden werden in drei Kategorien unterteilt:

- Klassifikation
- Assoziation
- Segmentierung

Klassifizierungsmodelle

Bei *Klassifizierungsmodellen* werden die Werte von einem oder mehreren Feldern **Eingabe** verwendet, um den Wert von einem oder mehreren Feldern “Ausgabe” oder **Ziel** vorherzusagen. Zu diesen Techniken zählen zum Beispiel: Entscheidungsbäume (C&R-Baum-, QUEST-, CHAID- und C5.0-Algorithmen), Regression (lineare, logistische, verallgemeinerte lineare und Cox-Regressionsalgorithmen), neuronale Netze, Support Vector Machines und Bayes-Netzwerke.

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

Klassifikationsknoten



Mit dem Knoten “Autom. Klassifizierer” können Sie eine Reihe verschiedener Modelle für binäre Ergebnisse (“Ja” oder “Nein”, “Abwanderung” oder “Keine Abwanderung” usw.) erstellen und vergleichen, um den besten Ansatz für die jeweilige Analyse auszuwählen. Es wird eine Reihe von Modellierungsalgorithmen unterstützt, sodass Sie die gewünschten Methoden, die spezifischen Optionen für die jeweilige Methode und die Kriterien zum Vergleich der Ergebnisse auswählen können. Der Knoten generiert eine Gruppe von Modellen, die auf den angegebenen Optionen beruhen, und erstellt anhand der von Ihnen angegebenen Kriterien eine Rangordnung der besten Kandidaten. [Für weitere Informationen siehe Thema Knoten “Automatischer Klassifizierer” in Kapitel 5 in IBM SPSS Modeler 15 Modellierungsknoten.](#)



Der Knoten "Auto-Numerisch" schätzt und vergleicht mit einer Reihe verschiedener Methoden Modelle für die Ergebnisse stetiger numerischer Bereiche. Der Knoten arbeitet auf dieselbe Weise wie der Knoten "Automatischer Klassifizierer": Sie können die zu verwendenden Algorithmen auswählen und in einem Modellierungsdurchlauf mit mehreren Optionskombinationen experimentieren. Folgende Algorithmen werden unterstützt: C&RT-Baum, CHAID, lineare Regression, verallgemeinerte lineare Regression und Support Vector Machines (SVM). Modelle können anhand von Korrelation, relativem Fehler bzw. Anzahl der verwendeten Variablen verglichen werden. Für weitere Informationen siehe Thema Knoten "Auto-Numerisch" in Kapitel 5 in *IBM SPSS Modeler 15 Modellierungsknoten*.



Der Knoten für Klassifizierungs- und Regressions-Bäume (C&RT-Bäume) erstellt einen Entscheidungsbau, mit dem Sie zukünftige Beobachtungen vorhersagen oder klassifizieren können. Bei dieser Methode wird eine rekursive Partitionierung verwendet, um die Trainingsdatensätze in Segmente aufzuteilen. Dabei wird bei jedem Schritt die Unreinheit verringert und ein Knoten im Baum wird als "rein" betrachtet, wenn 100 % der Fälle in eine bestimmte Kategorie des Zielfelds fallen. Ziel- und Eingabefelder können numerische Bereiche oder kategorial (nominal, ordinal oder Flags) sein. Alle Aufteilungen sind binär (nur zwei Untergruppen). Für weitere Informationen siehe Thema *C&R-Baumknoten* in Kapitel 6 in *IBM SPSS Modeler 15 Modellierungsknoten*.



Der QUEST-Knoten bietet eine binäre Klassifizierungsmethode zum Erstellen von Entscheidungsbäumen, die dazu dient, die für große C&R-Baum-Analysen erforderliche Verarbeitungszeit zu verkürzen. Gleichzeitig soll die in den Klassifizierungsbäummodellen festgestellte Tendenz verringert werden, die darin besteht, dass Eingaben bevorzugt werden, die mehr Aufteilungen erlauben. Eingabefelder können stetig (numerische Bereiche) sein, das Zielfeld muss aber kategorial sein. Alle Aufteilungen sind binär. Für weitere Informationen siehe Thema *QUEST-Knoten* in Kapitel 6 in *IBM SPSS Modeler 15 Modellierungsknoten*.



Der CHAID-Knoten erzeugt Entscheidungsbäume unter Verwendung von Chi-Quadrat-Statistiken zur Ermittlung optimaler Aufteilungen. Im Gegensatz zu den Knoten vom Typ "C&RT-Baum" und "QUEST" kann CHAID nichtbinäre Bäume generieren, d. h. Bäume mit Aufteilungen mit mehr als zwei Verzweigungen. Ziel- und Eingabefelder können in einem numerischen Bereich (stetig) oder kategorial sein. Exhaustive CHAID ist eine Änderung von CHAID, die noch gründlicher vorgeht, indem sie alle möglichen Aufteilungen untersucht, allerdings mehr Rechenzeit beansprucht. Für weitere Informationen siehe Thema *CHAID-Knoten* in Kapitel 6 in *IBM SPSS Modeler 15 Modellierungsknoten*.



Der C5.0-Knoten erstellt entweder einen Entscheidungsbau oder ein Regel-Set. Das Modell teilt die Stichprobe auf der Basis des Felds auf, das auf der jeweiligen Ebene den maximalen Informationsgewinn liefert. Das Zielfeld muss kategorial sein. Es sind mehrere Aufteilungen in mehr als zwei Untergruppen zulässig. Für weitere Informationen siehe Thema *C5.0-Knoten* in Kapitel 6 in *IBM SPSS Modeler 15 Modellierungsknoten*.



Der Knoten "Entscheidungsliste" kennzeichnet Untergruppen bzw. Segmente, die eine höhere oder geringere Wahrscheinlichkeit für ein bestimmtes binäres Ergebnis aufweisen als die Gesamtpopulation. Sie könnten beispielsweise nach Kunden suchen, deren Abwanderung unwahrscheinlich ist oder die mit großer Wahrscheinlichkeit positiv auf eine Kampagne reagieren. Sie können Ihr Geschäftswissen in das Modell integrieren, indem Sie eigene, benutzerdefinierte Segmente hinzufügen und eine Vorschau anzeigen, in der alternative Modelle nebeneinander angezeigt werden, um die Ergebnisse zu vergleichen. Entscheidungslistenmodelle bestehen aus einer Liste von Regeln, bei denen jede Regel eine Bedingung und ein Ergebnis aufweist. Regeln werden in der vorgegebenen Reihenfolge angewendet und die erste Regel, die zutrifft, bestimmt das Ergebnis. Für weitere Informationen siehe [Thema Entscheidungsliste in Kapitel 9 in IBM SPSS Modeler 15 Modellierungsknoten](#).



Bei linearen Regressionsmodellen wird ein stetiges Ziel auf der Basis linearer Beziehungen zwischen dem Ziel und einem oder mehreren Prädiktoren vorhergesagt. Für weitere Informationen siehe [Thema Lineare Modelle in Kapitel 10 in IBM SPSS Modeler 15 Modellierungsknoten](#).



Der Faktor/PCA-Knoten bietet leistungsstarke Datenreduktionsverfahren zur Verringerung der Komplexität der Daten. Die Hauptkomponentenanalyse (PCA) findet lineare Kombinationen der Eingabefelder, die die Varianz im gesamten Set der Felder am besten erfassen, wenn die Komponenten orthogonal (senkrecht) zueinander sind. Mit der Faktorenanalyse wird versucht, die zugrunde liegenden Faktoren zu bestimmen, die die Korrelationsmuster innerhalb eines Sets beobachteter Felder erklären. Bei beiden Ansätzen besteht das Ziel darin, eine kleinere Zahl abgeleiteter Felder zu finden, mit denen die Informationen in der ursprünglichen Menge der Felder effektiv zusammengefasst werden können. Für weitere Informationen siehe [Thema Faktor/PCA-Knoten in Kapitel 10 in IBM SPSS Modeler 15 Modellierungsknoten](#).



Der Merkmalsauswahlknoten sichtet die Eingabefelder, um auf der Grundlage einer Reihe von Kriterien (z. B. dem Prozentsatz der fehlenden Werte) zu entscheiden, ob diese entfernt werden sollen. Anschließend erstellt er eine Wichtigkeitsrangfolge der verbleibenden Eingaben in Bezug auf ein angegebenes Ziel. Beispiel: Angenommen, Sie haben ein Daten-Set mit Hunderten potenzieller Eingaben. Welche davon sind voraussichtlich für die Modellierung von medizinischen Behandlungsergebnissen von Bedeutung? Für weitere Informationen siehe [Thema Merkmalsauswahlknoten in Kapitel 4 in IBM SPSS Modeler 15 Modellierungsknoten](#).



Bei der Diskriminanzanalyse werden strengere Annahmen als bei der logistischen Regression verwendet, sie kann jedoch eine wertvolle Alternative oder Ergänzung zu einer logistischen Regressionsanalyse sein, wenn diese Annahmen erfüllt sind. Für weitere Informationen siehe [Thema Diskriminanzknoten in Kapitel 10 in IBM SPSS Modeler 15 Modellierungsknoten](#).



Die logistische Regression ist ein statistisches Verfahren zur Klassifizierung von Datensätzen auf der Grundlage der Werte von Eingabefeldern. Sie ist analog zur linearen Regression, außer dass statt eines numerischen Bereichs ein kategoriales Zielfeld verwendet wird. Für weitere Informationen siehe [Thema Logistikknoten in Kapitel 10 in IBM SPSS Modeler 15 Modellierungsknoten](#).



Das verallgemeinerte lineare Modell erweitert das allgemeine lineare Modell so, dass die abhängige Variable über eine angegebene Verknüpfungsfunktion in linearem Zusammenhang zu den Faktoren und Kovariaten steht. Außerdem ist es mit diesem Modell möglich, dass die abhängige Variable eine von der Normalverteilung abweichende Verteilung aufweist. Es deckt die Funktionen einer großen Bandbreite an Statistikmodellen ab, darunter lineare Regression, logistische Regression, loglineare Modelle für Häufigkeitsdaten und Überlebensmodelle mit Intervallzensierung. Für weitere Informationen siehe Thema GenLin-Knoten in Kapitel 10 in *IBM SPSS Modeler 15 Modellierungsknoten*.



Verallgemeinerte lineare gemischte Modelle (Generalized Linear Mixed Models; GLMM) erweitern lineare Modelle so, dass das Ziel nicht normalverteilt zu sein braucht und über eine angegebene Verknüpfungsfunktion in einer linearen Beziehung zu den Faktoren und Kovariaten steht und die Beobachtungen korreliert werden können. Verallgemeinerte lineare gemischte Modelle decken eine breite Palette verschiedener Modelle ab, von einfacher linearer Regression bis hin zu komplexen Mehrebenenmodellen für nicht normalverteilte Longitudinaldaten. Für weitere Informationen siehe Thema GLMM-Knoten in Kapitel 10 in *IBM SPSS Modeler 15 Modellierungsknoten*.



Der Knoten vom Typ "Cox-Regression" ermöglicht Ihnen auch bei zensierten Datensätzen die Erstellung eines Überlebensmodells für Daten über die Zeit bis zum Eintreten des Ereignisses. Das Modell erstellt eine Überlebensfunktion, die die Wahrscheinlichkeit vorhersagt, dass das untersuchte Ereignis für bestimmte Werte der Eingabevervariablen zu einem bestimmten Zeitpunkt (t) eingetreten ist. Für weitere Informationen siehe Thema Cox-Knoten in Kapitel 10 in *IBM SPSS Modeler 15 Modellierungsknoten*.



Der Knoten "Support Vector Machine" (SVM) ermöglicht die Klassifizierung von Daten in eine von zwei Gruppen ohne Überanpassung. SVM eignet sich gut für umfangreiche Daten-Sets, beispielsweise solche mit einer großen Anzahl an Eingabefeldern. Für weitere Informationen siehe Thema SVM-Knoten in Kapitel 15 in *IBM SPSS Modeler 15 Modellierungsknoten*.



Mithilfe des Bayes-Netzwerk-Knotens können Sie ein Wahrscheinlichkeitsmodell erstellen, indem Sie beobachtete und aufgezeichnete Hinweise mit Weltwissen kombinieren, um die Wahrscheinlichkeit ihres Vorkommens zu ermitteln. Der Knoten ist speziell für Netzwerke vom Typ "Tree Augmented Naïve Bayes" (TAN) und "Markov-Decke" gedacht, die in erster Linie zur Klassifizierung verwendet werden. Für weitere Informationen siehe Thema Bayes-Netzwerk-Knoten in Kapitel 7 in *IBM SPSS Modeler 15 Modellierungsknoten*.



Mithilfe des Knotens für das Selbstlern-Antwortmodell (Self-Learning Response Model, SLRM) können Sie ein Modell erstellen, in dem das Modell anhand eines einzelnen neuen Falls oder einer kleinen Anzahl neuer Fälle neu eingeschätzt werden kann, ohne dass das Modell mit allen Daten neu trainiert werden muss. Für weitere Informationen siehe Thema SLM-Knoten in Kapitel 14 in *IBM SPSS Modeler 15 Modellierungsknoten*.



Der Zeitreihenknoten berechnet Schätzungen für die exponielle Glättung sowie univariate und multivariate ARIMA-Modelle (ARIMA steht für Autoregressive Integrated Moving Average (autoregressiver integrierter gleitender Durchschnitt)) für Zeitreihendaten und erstellt Vorhersagen über die zukünftige Leistung. Einem Zeitreihenknoten muss stets ein Zeitintervallknoten vorangehen. Für weitere Informationen siehe Thema [Zeitreihen – Modellierungsknoten](#) in Kapitel 13 in IBM SPSS Modeler 15 Modellierungsknoten.



Der Knoten “ k -Nächste Nachbarn” (KNN) verknüpft einen neuen Fall mit der Kategorie oder dem Wert der k Objekte, die ihm im Prädiktorraum am nächsten liegen, wobei k eine Ganzzahl ist. Ähnliche Fälle liegen nah beieinander und Fälle mit geringer Ähnlichkeit sind weit voneinander entfernt. Für weitere Informationen siehe Thema [KNN-Knoten](#) in Kapitel 16 in IBM SPSS Modeler 15 Modellierungsknoten.

Assoziationsmodelle

Assoziationsmodelle finden Muster in Ihren Daten, wenn ein oder mehrere Elemente (wie zum Beispiel Ereignisse, Einkäufe oder Attribute) mit einem oder mehreren Elementen verknüpft werden. Die Modelle erstellen Regel-Sets, die diese Beziehungen definieren. Dabei können die Felder in den Daten sowohl als Eingaben als auch als Ziele dienen. Sie können solche Assoziationen natürlich auch manuell finden, doch erledigen Algorithmen für Assoziationsregeln das sehr viel schneller. Außerdem können sie komplexere Muster untersuchen. A Priori- und Carma-Modelle sind Beispiele für den Einsatz solcher Algorithmen. Ein weiterer Typ von Assoziationsmodell ist ein Sequenzerkennungsmodell, das sequenzielle Muster in zeitlich strukturierten Daten erkennt.

Assoziationsmodelle sind am nützlichsten bei der Vorhersage mehrerer Ergebnisse – wenn beispielsweise Kunden, die Produkt X gekauft haben, auch Y und Z gekauft haben. Assoziationsmodelle verknüpfen eine bestimmte Schlussfolgerung (wie die Entscheidung, etwas zu kaufen) mit einem Set von Bedingungen. Der Vorteil von Algorithmen für Assoziationsregeln im Vergleich zu Algorithmen für Standardentscheidungsbäume (C5.0 und C&RT) besteht darin, dass zwischen beliebigen Attributen Verbindungen bestehen können. Ein Entscheidungsbäumalgoritmus erstellt Regeln mit nur einer Schlussfolgerung, während Assoziationsalgorithmen viele Regeln zu finden versuchen, von denen jede zu einer anderen Schlussfolgerung kommen kann.

Assoziationsknoten



Der A-Priori-Knoten extrahiert eine Regelmenge aus den Daten und daraus die Regeln mit dem höchsten Informationsgehalt. A Priori bietet fünf verschiedene Methoden zur Auswahl von Regeln und verwendet ein ausgereiftes Indizierungsschema zur effizienten Verarbeitung großer Daten-Sets. Bei großen Problemen ist A Priori in der Regel schneller zu trainieren, es gibt keine willkürliche Begrenzung für die Anzahl der Regeln, die beibehalten werden können, und es können Regeln mit bis zu 32 Vorbedingungen verarbeitet werden. Bei A Priori müssen alle Ein- und Ausgabefelder kategorial sein; dafür bietet es jedoch eine bessere Leistung, da es für diesen Datentyp optimiert ist. Für weitere Informationen siehe Thema [A Priori-Knoten](#) in Kapitel 12 in IBM SPSS Modeler 15 Modellierungsknoten.



Beim CARMA-Modell wird eine Regelmenge aus den Daten extrahiert, ohne dass Sie Eingabe- oder Ziel-Felder angeben müssen. Im Gegensatz zu A Priori bietet der CARMA-Knoten Erstellungseinstellungen für die Regelunterstützung (Unterstützung für Antezedens und Sukzedens) und nicht nur für die Antezedens-Unterstützung. Die erstellten Regeln können somit für eine größere Palette an Anwendungen verwendet werden, beispielsweise um eine Liste mit Produkten und Dienstleistungen (Antezedenzen) zu finden, deren Nachfolger (Sukzedens) das Element darstellt, das Sie in der Ferienzeit desselben Jahres bewerben möchten. Für weitere Informationen siehe Thema [CARMA-Knoten in Kapitel 12 in IBM SPSS Modeler 15 Modellierungsknoten](#).



Der Sequenzknoten erkennt Assoziationsregeln in sequenziellen oder zeitorientierten Daten. Eine Sequenz ist eine Liste mit Element-Sets, die in einer vorhersagbaren Reihenfolge auftreten. Beispiel: Ein Kunde, der einen Rasierer und After-Shave-Lotion kauft, kauft möglicherweise beim nächsten Einkauf Rasiercreme. Der Sequenzknoten basiert auf dem CARMA-Assoziationsregelalgorithmus, der eine effiziente bidirektionale Methode zum Suchen von Sequenzen verwendet. Für weitere Informationen siehe Thema [Sequenzknoten in Kapitel 12 in IBM SPSS Modeler 15 Modellierungsknoten](#).

Segmentierungsmodelle

Bei *Segmentierungsmodellen* werden die Daten in Segmente oder Cluster von Datensätzen unterteilt, die ähnliche Muster der Eingabefelder aufweisen. Da sie ausschließlich auf die Eingabefelder konzentriert sind, gibt es bei Segmentierungsmodellen kein Konzept für Ausgabe- oder Zielfelder. Beispiele für Segmentierungsmodelle sind Kohonen-Netzwerke, K-Means-Clustering, Two-Step-Clustering und die Anomalieerkennung.

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

Segmentierungsknoten



Mit dem Knoten “Autom. Cluster” können Sie Clustering-Modelle, die Gruppen und Datensätze mit ähnlichen Merkmalen identifizieren, schätzen und vergleichen. Die Funktionsweise des Knotens gleicht der von anderen Knoten für automatisierte Modellierung, und Sie können in einem einzigen Modellierungsdurchgang mit mehreren Optionskombinationen experimentieren. Modelle können mithilfe grundlegender Messwerte für Filterung und Rangfolge der Nützlichkeit von Cluster-Modellen verglichen werden, um ein Maß auf der Basis der Wichtigkeit von bestimmten Feldern zu liefern. Für weitere Informationen siehe [Thema Knoten “Autom. Cluster” in Kapitel 5 in IBM SPSS Modeler 15 Modellierungsknoten](#).



Der K-Means-Knoten teilt das Daten-Set in unterschiedliche Gruppen (oder Cluster) auf. Bei diesem Verfahren wird eine festgelegte Anzahl von Clustern definiert, den Clustern werden iterativ Datensätze zugewiesen und die Cluster-Zentren werden angepasst, bis eine weitere Verfeinerung keine wesentliche Verbesserung des Modells mehr darstellen würde. Statt zu versuchen, ein Ergebnis vorherzusagen, versucht K-Means mithilfe eines als “nicht überwachtes Lernen” bezeichneten Verfahrens Muster im Set der Eingabefelder zu entdecken. Für weitere Informationen siehe [Thema K-Means-Knoten in Kapitel 11 in IBM SPSS Modeler 15 Modellierungsknoten](#).



Der Kohonen-Knoten erstellt eine Art von neuronalem Netzwerk, das verwendet werden kann, um ein Clustering der Datenmenge in einzelne Gruppen vorzunehmen. Wenn das Netz voll trainiert ist, sollten ähnliche Datensätze auf der Ausgabekarte eng nebeneinander stehen, während Datensätze, die sich unterscheiden, weit voneinander entfernt sein sollten. Die Zahl der von jeder Einheit im Modell-Nugget erfassten Beobachtungen gibt Aufschluss über die starken Einheiten. Dadurch wird ein Eindruck von der ungefähren Zahl der Cluster vermittelt. Für weitere Informationen siehe [Thema Kohonen-Knoten in Kapitel 11 in IBM SPSS Modeler 15 Modellierungsknoten](#).



Der TwoStep-Knoten verwendet eine aus zwei Schritten bestehende Clusterbildungsmethode. Im ersten Schritt wird ein einzelner Durchlauf durch die Daten vorgenommen, bei dem die Eingangsrohdaten zu einem verwaltbaren Set von Unterclustern komprimiert werden. Im zweiten Schritt werden die Untercluster mithilfe einer hierarchischen Methode zur Clusterbildung nach und nach in immer größere Cluster zusammengeführt. TwoStep hat den Vorteil, dass die optimale Anzahl an Clustern für die Trainingsdaten automatisch geschätzt wird. Mit dem Verfahren können gemischte Feldtypen und große Daten-Sets effizient verarbeitet werden. Für weitere Informationen siehe [Thema TwoStep-Cluster-Knoten in Kapitel 11 in IBM SPSS Modeler 15 Modellierungsknoten](#).



Der Anomalieerkennungsknoten ermittelt ungewöhnliche Fälle bzw. “Ausreißer”, die nicht den Mustern der “normalen” Daten entsprechen. Mit diesem Knoten können Ausreißer ermittelt werden, selbst wenn sie keinem bereits bekannten Muster entsprechen und selbst wenn Sie nicht genau wissen, wonach Sie suchen. Für weitere Informationen siehe [Thema Anomalieerkennungsknoten in Kapitel 4 in IBM SPSS Modeler 15 Modellierungsknoten](#).

Modelle für In-Database Mining

SPSS Modeler unterstützt die Integration mit Data Mining-Tools und Daten-Modellierungstools von Datenbankherstellern wie Oracle Data Miner, IBM DB2 InfoSphere Warehouse und Microsoft Analysis Services. Sie können Modelle in der Datenbank erstellen, bewerten und speichern – ohne

dazu die SPSS Modeler-Anwendung verlassen zu müssen. Vollständige Informationen finden Sie im *SPSS Modeler In-Database Mining Handbuch*, das sich auf dem Produkt-DVD befindet.

IBM SPSS Statistics Modelle

Wenn auf Ihrem Computer eine Kopie von IBM® SPSS® Statistics installiert und lizenziert ist, können Sie auf bestimmte SPSS Statistics-Routinen in SPSS Modeler zugreifen und diese ausführen, um Modelle zu erstellen und zu scoren. [Für weitere Informationen siehe Thema IBM SPSS Statistics-Knoten – Überblick in Kapitel 8 in IBM SPSS Modeler 15 Quellen-, Prozess- und Ausgabeknoten.](#)

Weitere Informationen

Es steht auch eine detaillierte Dokumentation zu den Modellierungsalgorithmen zur Verfügung. Weitere Informationen finden Sie im *SPSS Modeler-Algorithmushandbuch*, das sich auf dem Produkt-DVD befindet.

Data Mining-Beispiele

Um einen Einblick in die praktische Verwendung von Data Mining zu gewinnen, beginnen Sie am besten mit einem Beispiel. Eine Reihe von Anwendungsbeispielen steht im *IBM® SPSS® ModelerClementine-Anwendungshandbuch* zur Verfügung. Dieses Handbuch bietet eine kurze, gezielte Einführung in bestimmte Modellierungsmethoden und -verfahren. [Für weitere Informationen siehe Thema Anwendungsbeispiele in Kapitel 1 auf S. 6.](#)

Erstellen von Streams

Überblick zum Erstellen von Streams

Data Mining mit IBM® SPSS® Modeler konzentriert sich darauf, Daten durch eine Reihe von Knoten, die als **Stream** bezeichnet werden, laufen zu lassen. Die Knotenreihe stellt Operationen dar, die an den Daten ausgeführt werden sollen, während die Verknüpfungen zwischen den Knoten die Richtung des Datenflusses angeben. In der Regel verwenden Sie einen Daten-Stream zum Einlesen von Daten in SPSS Modeler und dazu, diese durch eine Reihe von Bearbeitungsoperationen laufen zu lassen und sie dann an ein Ziel zu senden, z. B. an eine Tabelle oder einen Viewer.

Angenommen, Sie möchten beispielsweise eine Datenquelle öffnen, ein neues Feld hinzufügen, Datensätze basierend auf den Werten im neuen Feld auswählen und die Ergebnisse dann in einer Tabelle anzeigen. In diesem Fall besteht Ihr Daten-Stream aus vier Knoten:



Einem Knoten vom Typ “Datei (var.)”, den Sie zum Lesen der Daten aus der Datenquelle einrichten.



Ein Ableitungsknoten, mit dem Sie das neue berechnete Feld zum Daten-Set hinzufügen.



Ein Auswahlknoten, mit dem Sie Auswahlkriterien festlegen, um Datensätze aus dem Daten-Stream auszuschließen.



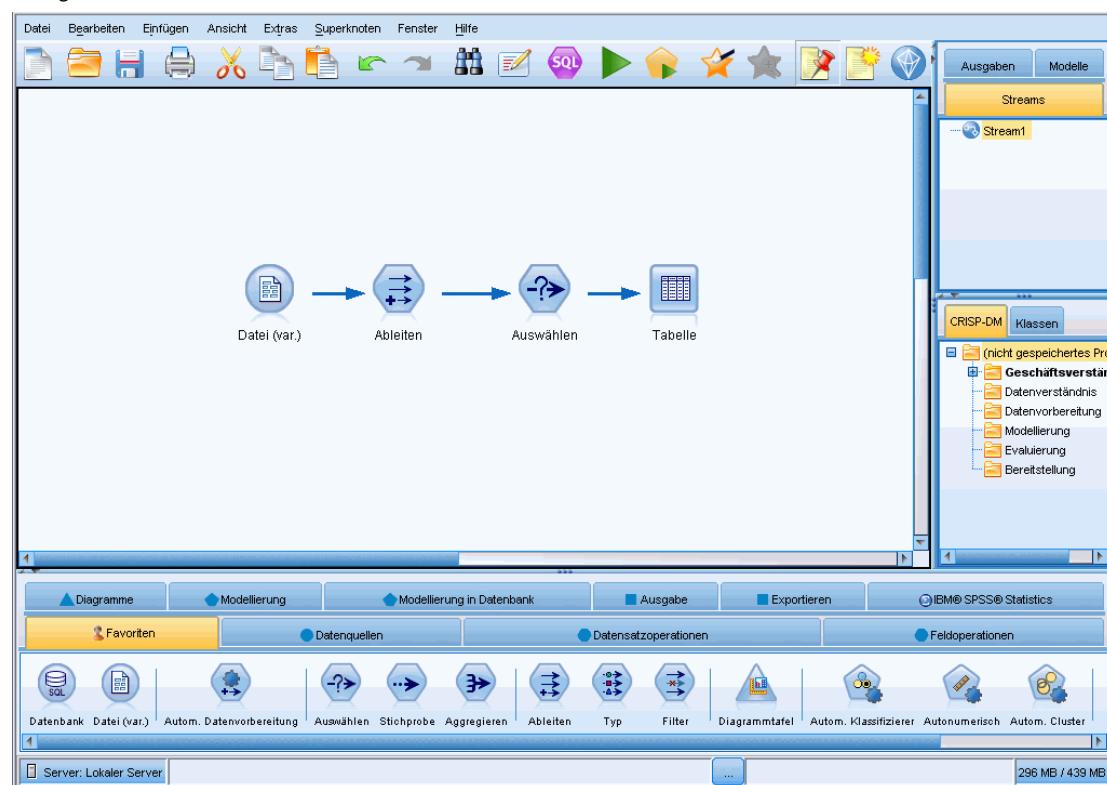
Ein Tabellenknoten, den Sie zum Anzeigen der Ergebnisse Ihrer Bearbeitungen auf dem Bildschirm verwenden.

Erstellen von Daten-Streams

Mit der einmaligen Benutzeroberfläche von IBM® SPSS® Modeler können Sie einen visuellen Mining-Vorgang für Ihre Daten ausführen, indem Sie mit Diagrammen von Daten-Streams arbeiten. Am einfachsten können Sie einen Daten-Stream wie folgt erstellen:

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

Abbildung 5-1
Fertiger Stream im Stream-Zeichenbereich



Dieser Abschnitt enthält detaillierte Informationen zum Arbeiten mit Knoten, um komplexere Daten-Streams zu erstellen. Es werden außerdem Optionen und Einstellungen für Knoten und Streams beschrieben. Beispiele mit Einzelschrittanleitungen zum Erstellen von Streams anhand der mit SPSS Modeler mitgelieferten Daten (im Ordner *Demos* Ihrer Programminstallation), finden Sie hier: [Anwendungsbeispiele auf S. 6.](#)

Arbeiten mit Knoten

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

Es gibt mehrere Knotentypen. **Quellenknoten** bringen Daten in den Stream und befinden sich in der Registerkarte "Quellen" der Knotenpalette. **Prozessknoten** führen Operationen an einzelnen Datensätzen und Feldern aus und sind in den Registerkarten "Datensatzoperationen" und "Feldoperationen" der Palette zu finden. **Ausgabeknoten** erzeugen eine Vielzahl von Ausgaben für Daten-, Diagramm- und Modellergebnisse und befinden sich in den Registerkarten "Diagramme", "Ausgabe" und "Export" der Knotenpalette. **Modellierungsknoten** verwenden statistische Algorithmen, um Modell-Nuggets zu erstellen, und befinden sich in der Registerkarte "Modellierung" sowie der Registerkarte "Datenbankmodellierung" (falls aktiviert) der Knotenpalette. Für weitere Informationen siehe Thema [Knotenpalette in Kapitel 3](#) auf S. 22.

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

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

Hinweis: Sie können die Knotenpalette anpassen. Für weitere Informationen siehe Thema [Anpassen der Knotenpalette in Kapitel 12 auf S. 250](#).

Hinzufügen von Knoten zu einem Stream

Es gibt mehrere Möglichkeiten zum Hinzufügen von Knoten zu einem Stream aus der Knotenpalette:

- Doppelklicken Sie auf einen Knoten in der Palette. *Hinweis:* Durch das Doppelklicken auf einen Knoten wird dieser automatisch mit dem aktuellen Stream verbunden. Für weitere Informationen siehe Thema [Verbinden von Knoten in einem Stream auf S. 50](#).
- Ziehen Sie einen Knoten per Drag-and-Drop aus der Palette in den Stream-Zeichenbereich.
- Klicken Sie auf einen Knoten in der Palette und dann in den Stream-Zeichenbereich.
- Wählen Sie eine geeignete Option aus dem Menü “Einfügen” von IBM® SPSS® Modeler.

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

Entfernen von Knoten

Um einen Knoten aus dem Daten-Stream zu entfernen, klicken Sie darauf und drücken Sie auf Entf oder klicken Sie mit der rechten Maustaste und wählen Sie im Menü die Option Löschen.

Verbinden von Knoten in einem Stream

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

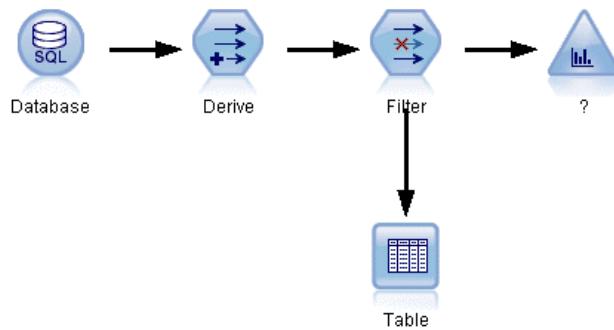
So fügen Sie Knoten durch Doppelklicken hinzu und verbinden Knoten miteinander:

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

Ableitungsknoten automatisch mit dem bereits vorhandenen Datenbankknoten verbunden. Sie können diesen Prozess so lange wiederholen, bis Sie einen Endknoten, z. B. einen Histogramm- oder Tabellenknoten, erreicht haben. Zu diesem Zeitpunkt werden sämtliche neuen Knoten mit dem letzten Knoten, der kein Endknoten ist, aufwärts verbunden.

Abbildung 5-2

Stream, der durch Doppelklicken auf die Knoten in den Paletten erstellt wurde

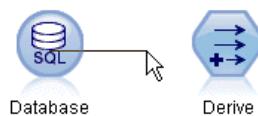


So verbinden Sie Knoten mit der mittleren Maustaste:

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

Abbildung 5-3

Verwenden der mittleren Maustaste zum Verbinden von Knoten



So verbinden Sie Knoten manuell:

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

- ▶ Klicken Sie mit der rechten Maustaste auf den Knoten, der die Verbindung starten soll. Dadurch wird das Knoten-Menü geöffnet.
- ▶ Klicken Sie im Menü auf Verbinden.
- ▶ Sowohl am Anfangsknoten als auch am Cursor wird ein Verbindungssymbol angezeigt. Klicken Sie auf einen zweiten Knoten im Zeichenbereich, um die beiden Knoten zu verbinden.

Abbildung 5-4
Verbinden von Knoten mithilfe der Option "Verbinden" aus dem Popup-Menü

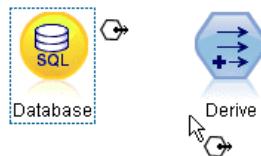


Abbildung 5-5
Verbundene Knoten



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

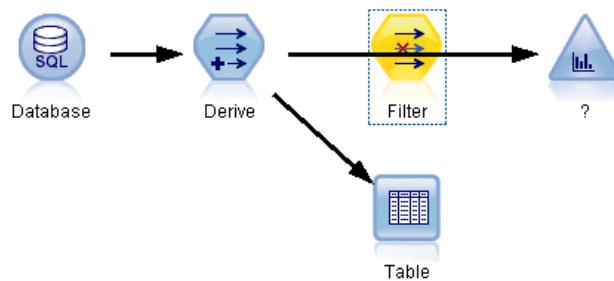
- Eine Verbindung, die zu einem Quellenknoten führt
- Eine Verbindung, die von einem Endknoten kommt
- Ein Knoten mit mehr als der maximalen Anzahl Eingabeverbindungen
- Verbinden von zwei Knoten, die bereits miteinander verbunden sind
- Umlauf (Daten kehren zu einem Knoten zurück, von dem sie bereits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

Umgehen von Knoten in einem Stream

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

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

Abbildung 5-6
Umgehen eines zuvor verbundenen Filterknotens



So umgehen Sie einen Knoten:

- ▶ Verwenden Sie im Stream-Zeichenbereich die mittlere Maustaste, um auf den Knoten doppelzuklicken, den Sie umgehen möchten. Alternativ können Sie auch die Kombination Alt+Doppelklick verwenden.

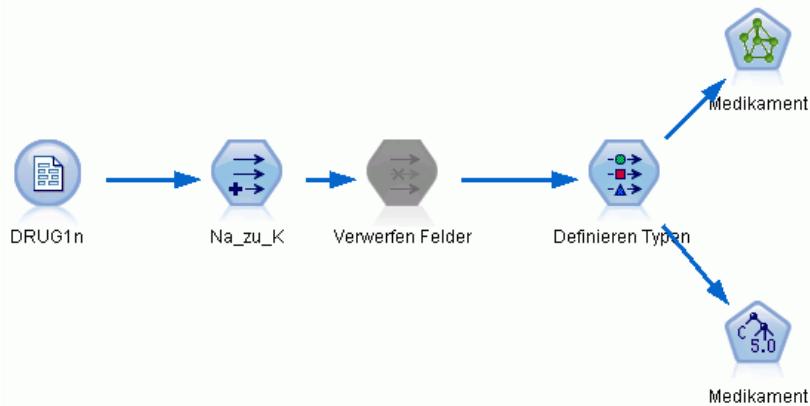
Hinweis: Sie können diese Aktion unter Verwendung der Option Rückgängig im Menü “Bearbeiten” oder mit der Tastenkombination Strg+Z rückgängig machen.

Deaktivieren von Knoten in einem Stream

Prozessknoten mit einer einzelnen Eingabe innerhalb von Streams können deaktiviert werden, mit dem Ergebnis, dass der Knoten bei der Ausführung des Streams ignoriert wird. Auf diese Weise müssen Sie den Knoten nicht löschen oder umgehen und können seine Verbindung mit den restlichen Knoten bestehen lassen. Die Knoteneinstellungen können nach wie vor angezeigt und bearbeitet werden. Änderungen werden jedoch erst dann wirksam, wenn Sie den Knoten wieder aktivieren.

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

Abbildung 5-7
Deaktivierter Filterknoten in einem Stream



So deaktivieren Sie einen Knoten:

- ▶ Klicken Sie im Stream-Zeichenbereich mit der rechten Maustaste auf den Knoten, den Sie deaktivieren möchten.
- ▶ Klicken Sie im Popup-Menü auf Knoten deaktivieren.

Alternativ können Sie auch im Menü “Bearbeiten” auf Knoten > Knoten deaktivieren klicken. Wenn Sie den Knoten wieder in den Stream integrieren möchten, wählen Sie auf dieselbe Weise wie oben beschrieben Knoten aktivieren.

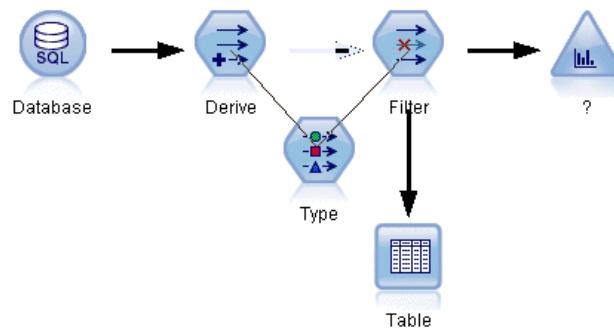
Hinweis: Sie können diese Aktion unter Verwendung der Option Rückgängig im Menü “Bearbeiten” oder mit der Tastenkombination Strg+Z rückgängig machen.

Sie können diese Aktion unter Verwendung der Option Rückgängig im Menü “Bearbeiten” oder mit der Tastenkombination Strg+Z rückgängig machen.

Hinzufügen von Knoten in bereits vorhandenen Verbindungen

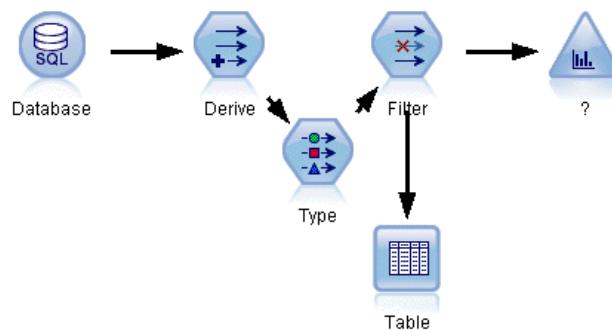
Zwischen zwei verbundenen Knoten können Sie einen neuen Knoten einfügen, indem Sie den Pfeil ziehen, der die beiden Knoten verbindet.

Abbildung 5-8
Verbinden eines neuen Knotens zwischen zwei verbundenen Knoten



- Klicken Sie mit der mittleren Maustaste auf den Verbindungspeil, in den Sie den Knoten einfügen möchten, und ziehen Sie diesen. Alternativ können Sie die ALT-Taste drücken und dann mit der Maus ziehen. Damit simulieren Sie die mittlere Maustaste.

Abbildung 5-9
Neuer Stream



- Ziehen Sie die Verbindung an den Knoten, den Sie einfügen möchten, und lassen Sie die Maustaste los.

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

Löschen von Verbindungen zwischen Knoten

So löschen Sie die Verbindung zwischen zwei Knoten:

- Klicken Sie mit der rechten Maustaste auf den Verbindungspeil.
- Klicken Sie im Menü auf Verbindung löschen.

Abbildung 5-10
Löschen der Verbindung zwischen Knoten in einem Stream



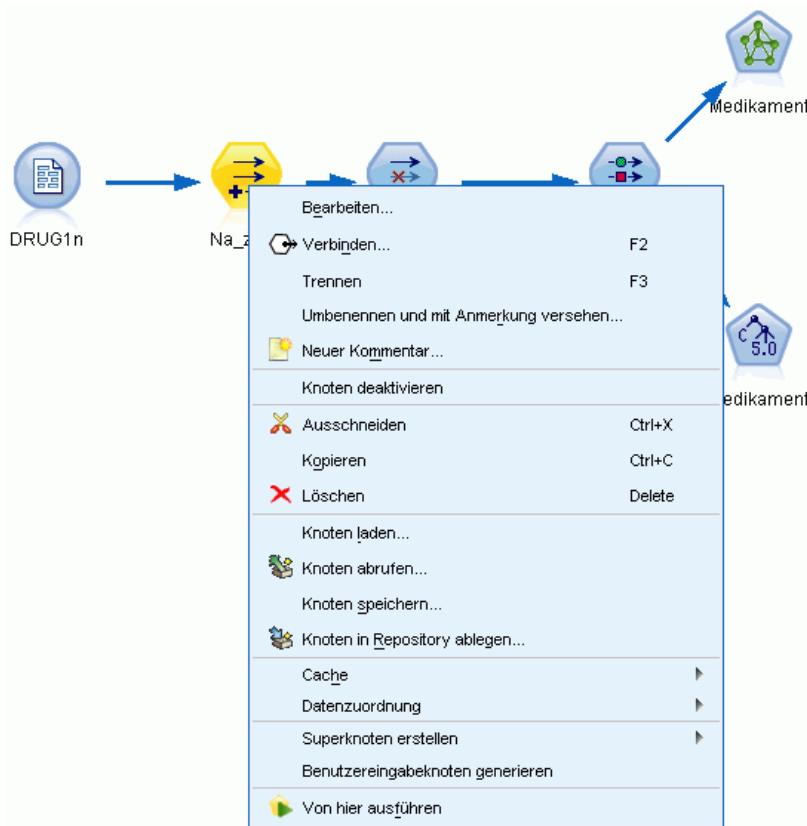
Führen Sie zum Löschen aller Verbindungen zu und von einem Knoten eine der folgenden Aktionen aus:

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

Festlegen von Optionen für Knoten

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

Abbildung 5-11
Optionen im Popup-Menü für Knoten



- Klicken Sie auf Bearbeiten, um das Dialogfeld für den ausgewählten Knoten zu öffnen.
- Klicken Sie auf Verbinden, um einen Knoten manuell mit einem anderen zu verbinden.
- Klicken Sie auf Trennen, um alle Verknüpfungen zum und vom Knoten zu löschen.
- Klicken Sie auf Umbenennen und mit Anmerkung versehen, um die Registerkarte “Anmerkungen” im Bearbeitungsdialegfeld aufzurufen.
- Klicken Sie auf Neuer Kommentar, um einem Knoten einen Kommentar hinzuzufügen. Für weitere Informationen siehe Thema [Hinzufügen von Kommentaren und Anmerkungen zu Knoten und Streams](#) auf S. 88.
- Klicken Sie auf Knoten deaktivieren, um den Knoten während der Verarbeitung “auszublenden”. Um den Knoten wieder “einzublenden”, Klicken Sie auf Knoten aktivieren. Für weitere Informationen siehe Thema [Deaktivieren von Knoten in einem Stream](#) auf S. 53.

- Klicken Sie auf Ausschneiden oder Löschen, um die ausgewählten Knoten aus dem Stream-Zeichenbereich zu entfernen. *Hinweis:* Mit der Option Ausschneiden können Sie Knoten einfügen, mit Löschen nicht.
- Klicken Sie auf Knoten kopieren, um eine Kopie des Knotens ohne Verbindungen zu erstellen. Dieser kann einem neuen oder bereits vorhandenen Stream hinzugefügt werden.
- Klicken Sie auf Knoten laden, um einen zuvor gespeicherten Knoten zu öffnen und seine Optionen in den aktuell ausgewählten Knoten zu laden. *Hinweis:* Die Knoten müssen vom gleichen Typ sein.
- Klicken Sie auf Knoten abrufen, um einen Knoten von einem verbundenen IBM® SPSS® Collaboration and Deployment Services Repository abzurufen. Für weitere Informationen siehe Thema Zugriff auf Objekte in IBM SPSS Collaboration and Deployment Services Repository in Kapitel 5 in *IBM SPSS Modeler 15 Handbuch für Skripterstellung und Automatisierung*.
- Klicken Sie auf Knoten speichern, um die Knotendetails in einer Datei zu speichern Sie können Knotendetails nur in einen anderen Knoten desselben Typs laden.
- Klicken Sie auf Knoten speichern, um den ausgewählten Knoten in einem verbundenen IBM SPSS Collaboration and Deployment Services Repository zu speichern. Für weitere Informationen siehe Thema Zugriff auf Objekte in IBM SPSS Collaboration and Deployment Services Repository in Kapitel 5 in *IBM SPSS Modeler 15 Handbuch für Skripterstellung und Automatisierung*.
- Klicken Sie auf Cache, um das Menü zu erweitern, das Optionen zum Cachen des ausgewählten Knotens enthält.
- Klicken Sie auf Datenzuordnung, um das Menü zu erweitern, das Optionen zum Zuordnen von Daten zu einer neuen Quelle oder zum Festlegen von erforderlichen Feldern enthält.
- Klicken Sie auf Superknoten erstellen, um das Menü zu erweitern, das Optionen zum Erstellen eines Superknotens im aktuellen Stream enthält. Für weitere Informationen siehe Thema Erstellen von Superknoten in Kapitel 9 in *IBM SPSS Modeler 15 Quellen-, Prozess- und Ausgabeknoten*.
- Klicken Sie auf Benutzereingabeknoten generieren, um den ausgewählten Knoten zu ersetzen. Die von diesem Knoten erstellten Beispiele enthalten dieselben Felder wie der aktuelle Knoten. Für weitere Informationen siehe Thema Benutzereingabeknoten in Kapitel 2 in *IBM SPSS Modeler 15 Quellen-, Prozess- und Ausgabeknoten*.
- Klicken Sie auf Ab hier ausführen, um alle Endknoten abwärts vom ausgewählten Knoten auszuführen.

Cachen von Optionen für Knoten

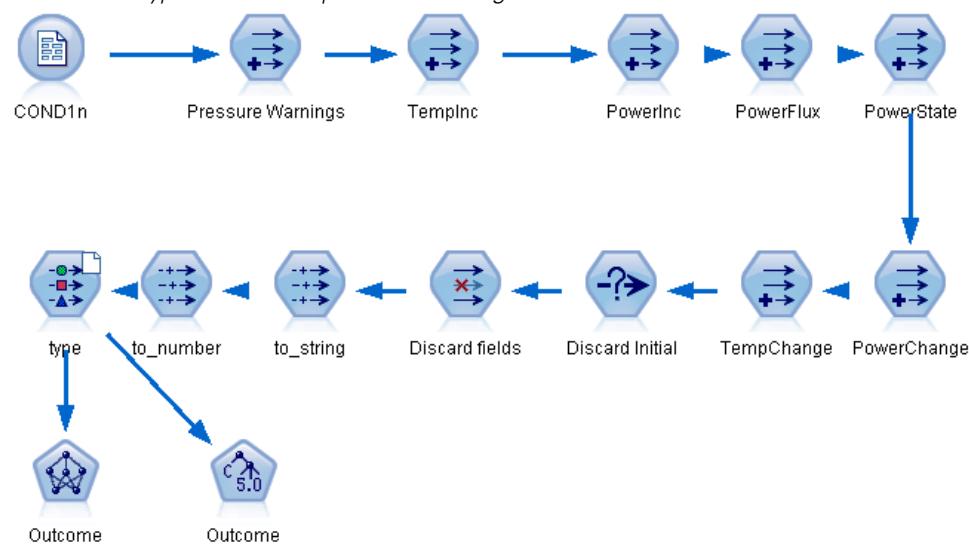
Um die Stream-Ausführung zu optimieren, können Sie einen **Cache** für Knoten einrichten, die keine Endknoten sind. Wenn Sie einen Cache für einen Knoten einrichten, wird der Cache mit den Daten gefüllt, die bei der nächsten Ausführung des Daten-Streams durch den Knoten laufen. Ab diesem Zeitpunkt werden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

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

Hinweis: Ein Caching bei Quellenknoten, bei denen einfach eine Kopie der ursprünglichen Dateien, die in IBM® SPSS® Modeler eingelesen wurden, gespeichert wird, führt unter den meisten Umständen nicht zu einer besseren Leistung.

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

Abbildung 5-12
Cachen am Typknoten zum Speichern neu abgeleiteter Felder



So aktivieren Sie einen Cache:

- ▶ Klicken Sie im Stream-Zeichenbereich mit der rechten Maustaste auf den Knoten und klicken Sie im Menü auf die Option Cache.
- ▶ Klicken Sie im Untermenü auf die Option Aktivieren.
- ▶ Sie können den Cache deaktivieren, indem Sie mit der rechten Maustaste auf den Knoten klicken und im Untermenü auf Deaktivieren klicken.

Cachen von Knoten in einer Datenbank

Bei in einer Datenbank ausgeführten Streams muss nicht das Dateisystem als Cache verwendet werden, sondern die Daten können mitten im Stream in einer temporären Tabelle in der Datenbank zwischengespeichert werden. Bei einer Kombination mit SQL-Optimierung kann dies zu einer erheblichen Leistungssteigerung führen. So kann beispielsweise die Ausgabe aus einem Stream,

der mehrere Tabellen zusammenfasst, um eine Data-Mining-Ansicht zu erstellen, im Cache gespeichert und bei Bedarf wieder verwendet werden. Durch die automatische Erzeugung von SQL für alle abwärtsgelegenen Knoten lässt sich die Leistung weiter erhöhen.

Bei Verwendung von Datenbank-Caching mit Zeichenketten, die mehr als 255 Zeichen umfassen, müssen Sie entweder sicherstellen, dass sich ein Typknoten im Stream oberhalb des Caching-Knotens befindet und dass die Feldwerte gelesen werden, oder Sie müssen die Zeichenkettenlänge mithilfe des Parameters `default_sql_string_length` in der Datei `options.cfg` festlegen. Dies stellt sicher, dass die entsprechende Spalte in der temporären Tabelle auf die korrekte Breite eingestellt wird, um die Strings anzuzeigen. Für weitere Informationen siehe Thema Optionen nur in `options.cfg` sichtbar in Kapitel 4 in *IBM SPSS Modeler Server 15-Verwaltungs- und Leistungshandbuch*.

Um das Caching von Datenbanken nutzen zu können, müssen sowohl SQL-Optimierung als auch Datenbank-Caching aktiviert sein. Die Optimierungseinstellungen auf dem Server haben Vorrang vor jenen auf dem Client. Für weitere Informationen siehe Thema Festlegen von Optimierungsoptionen für Streams auf S. 68.

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

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

So leeren Sie einen Cache:

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

- ▶ Klicken Sie im Stream-Zeichenbereich mit der rechten Maustaste auf den Knoten und klicken Sie im Menü auf die Option Cache.
- ▶ Wählen Sie im Untermenü die Option Leeren.

So speichern Sie einen Cache:

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

- ▶ Klicken Sie im Stream-Zeichenbereich mit der rechten Maustaste auf den Knoten und klicken Sie im Menü auf die Option Cache.
- ▶ Wählen Sie im Untermenü die Option Cache speichern.

- ▶ Blättern Sie im Dialogfeld “Cache speichern” an den Speicherort, an dem die Cache-Datei gespeichert werden soll.
- ▶ Geben Sie im Textfeld “Dateiname” einen Namen ein.
- ▶ Stellen Sie sicher, dass in der Liste “Dateityp” die Option *.sav ausgewählt ist, und klicken Sie auf Speichern.

So laden Sie einen Cache:

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

- ▶ Klicken Sie im Stream-Zeichenbereich mit der rechten Maustaste auf den Knoten und klicken Sie im Menü auf die Option Cache.
- ▶ Wählen Sie im Untermenü die Option Cache laden.
- ▶ Blättern Sie im Dialogfeld “Cache laden” an den Speicherort der Cache-Datei, wählen Sie diese aus und klicken Sie auf Laden.

Vorschau von Daten in Knoten

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

Für Knoten oberhalb eines Modell-Nuggets zeigt die Vorschau die Eingabefelder an. Für ein Modell-Nugget oder Knoten unterhalb des Nuggets (ausgenommen Endknoten) zeigt die Vorschau die Eingabe- und die generierten Felder an.

Die Standardanzahl der angezeigten Zeilen ist 10. Sie können diese Einstellung jedoch in den Stream-Eigenschaften ändern. Für weitere Informationen siehe Thema [Festlegen von allgemeinen Optionen für Streams auf S. 63](#).

Abbildung 5-13
Datenvorschau von einem Modell-Nugget

The screenshot shows a dialog box titled "Vorschau von Knoten taxable_value (10 Felder, 10 Datensätze)". The interface includes a menu bar with "Datei", "Bearbeiten", and "Generieren". Below the menu are two tabs: "Tabelle" (selected) and "Anmerkungen". The main area displays a table with 10 rows of data. The columns are labeled: year_built, volume_interior, volume_other, lot_size, taxable_value, \$XR-taxable_value, and \$XRE-taxable_value. The data rows are as follows:

	year_built	volume_interior	volume_other	lot_size	taxable_value	\$XR-taxable_value	\$XRE-taxable_value
1	979	166	11	100	90500	105523.184	1526.604
2	988	603	73	497	420000	355497.802	14929.477
3	987	303	75	91	152500	147713.631	1733.849
4	926	228	12	55	92000	69313.952	8116.212
5	988	666	145	441	390000	367551.491	16353.591
6	970	563	60	800	435000	395173.708	4759.398
7	902	355	41	225	130000	109865.220	3694.157
8	972	468	78	625	380500	334714.268	6367.557
9	986	315	14	107	158000	145149.106	1263.839
10	975	282	27	141	134000	138747.302	1892.730

At the bottom right of the dialog is an "OK" button.

Mithilfe des Menüs “Generieren” können Sie mehrere Arten von Knoten erzeugen. Für weitere Informationen siehe Thema Tabellen-Browser in Kapitel 6 in *IBM SPSS Modeler 15 Quellen-, Prozess- und Ausgabeknoten*.

Sperren von Knoten

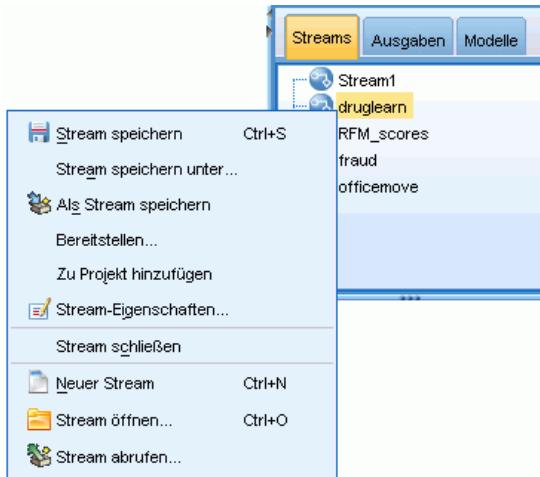
Um andere Benutzer daran zu hindern, die Einstellungen eines oder mehrerer Knoten in einem Stream zu ändern, können Sie den oder die Knoten im besonderen Knotentyp namens Superknoten verkapseln und dann den Superknoten durch Zuweisen eines Passwortschutzes sperren. Für weitere Informationen siehe Thema Sperren von Superknoten in Kapitel 9 in *IBM SPSS Modeler 15 Quellen-, Prozess- und Ausgabeknoten*.

Arbeiten mit Streams

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

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

Abbildung 5-14
Registerkarte "Streams" im Manager-Bereich mit den Popup-Menü-Optionen



Auf dieser Registerkarte können Sie:

- Auf Streams zugreifen.
- Streams speichern.
- Streams im aktuellen Projekt speichern.
- Streams schließen.
- Neue Streams öffnen.
- Streams aus einem IBM SPSS Collaboration and Deployment Services-Repository (falls an Ihrem Standort verfügbar) speichern und abrufen. Für weitere Informationen siehe Thema [Informationen zu IBM SPSS Collaboration and Deployment Services Repository in Kapitel 9 auf S. 180](#).

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

Festlegen von Optionen für Streams

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

- **Allgemein.** Verschiedene Optionen, wie beispielsweise die im Stream zu verwendenden Symbole und die zu verwendende Textkodierung. Für weitere Informationen siehe Thema [Festlegen von allgemeinen Optionen für Streams auf S. 63](#).
- **Datum/Uhrzeit.** Optionen, die sich auf das Format von Datums- und Zeitausdrücken beziehen. Für weitere Informationen siehe Thema [Festlegen von Datums- und Zeitoptionen für Streams auf S. 66](#).
- **Zahlenformate.** Optionen, die das Format numerischer Ausdrücke festlegen. Für weitere Informationen siehe Thema [Festlegen von Zahlenformatoptionen für Streams auf S. 67](#).

- **Optimierung.** Optionen zur Optimierung der Stream-Leistung. Für weitere Informationen siehe Thema Festlegen von Optimierungsoptionen für Streams auf S. 68.
- **Protokollierung und Status.** Optionen zur Festlegung von SQL-Protokollierung und Datensatzstatus. Für weitere Informationen siehe Thema Festlegen von Optionen für SQL-Protokollierung und Datensatzstatus für Streams auf S. 71.
- **Layout.** Optionen, die sich auf das Layout des Streams im Zeichenbereich beziehen. Für weitere Informationen siehe Thema Festlegen von Layoutoptionen für Streams auf S. 72.

So legen Sie Stream-Optionen fest:

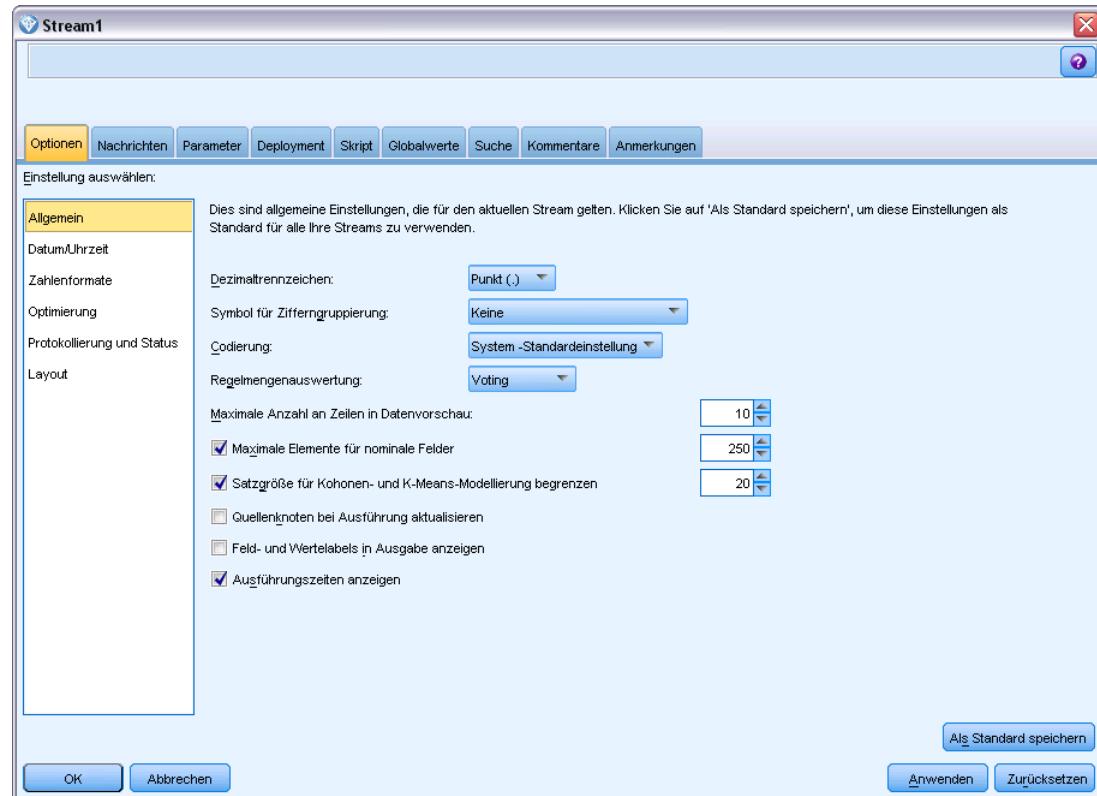
- ▶ Klicken Sie im Menü “Datei” auf die Option Stream-Eigenschaften (oder wählen Sie den Stream in der Registerkarte “Streams” im Manager-Bereich aus, klicken Sie mit der rechten Maustaste und wählen Sie dann im Popup-Menü die Option Stream-Eigenschaften).
- ▶ Klicken Sie auf die Registerkarte Optionen.

Klicken Sie als Alternative im Menü “Extras” auf:
Stream-Eigenschaften > Optionen

Festlegen von allgemeinen Optionen für Streams

Bei den allgemeinen Optionen handelt es sich um eine Reihe unterschiedlicher Optionen, die sich auf verschiedene Aspekte des aktuellen Streams beziehen.

Abbildung 5-15
Festlegen von allgemeinen Optionen für einen Stream



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

Symbol für Zifferngruppierung. Wählen Sie bei Zahlenanzeigeformaten aus, welches Symbol zur Gruppierung der Werte verwendet werden soll (z. B. der Punkt in 3.000,00). Folgende Optionen stehen zur Auswahl: "Keine", "Punkt", "Komma", "Leerzeichen" und "Durch Ländereinstellung definiert" (in diesem Fall wird der Standardwert für die aktuelle Ländereinstellung verwendet).

Kodierung. Legen Sie die Stream-Standardmethode zur Textkodierung fest. (*Hinweis:* Gilt nur für Quellenknoten "Datei (var.)" und Exportknoten "Textdatei". Alle anderen Knoten verwenden diese Einstellung nicht. Die meisten Datendateien verfügen über eingebettete Kodierungsinformationen.) Sie können entweder die System-Standardeinstellung oder UTF-8 wählen. Die System-Standardeinstellung wird in der Windows-Systemsteuerung oder – bei der Ausführung im verteilten Modus – auf dem Server-Computer angegeben. Für weitere Informationen siehe [Thema Unicode-Unterstützung in IBM SPSS Modeler in Anhang B auf S. 279](#).

Regelmengenauswertung. Legt fest, wie Regelmengenmodelle evaluiert werden. Standardmäßig verwenden Regelmengen Voting, um Vorhersagen von einzelnen Regeln zu kombinieren und die endgültige Prognose zu ermitteln. Um sicherzustellen, dass Regelmengen die erste gefundene Regel standardmäßig verwenden, wählen Sie Erster Treffer. Für weitere Informationen siehe [Thema Regelmengen-Modell-Nuggets in Kapitel 6 in IBM SPSS Modeler 15 Modellierungsknoten](#).

Beachten Sie, dass diese Option nicht für Entscheidungslistenmodelle gilt, die stets den ersten Treffer gemäß der Definition durch den Algorithmus verwenden.

Maximale Anzahl an Zeilen, die in der Datenvorschau gezeigt werden sollen. Geben Sie die Anzahl der Zeilen an, die in einer angeforderten Vorschau der Daten für einen Knoten angezeigt werden sollen. [Für weitere Informationen siehe Thema Vorschau von Daten in Knoten auf S. 60.](#)

Maximale Elemente für nominale Felder. Wählen Sie diese Option aus, um eine maximale Anzahl Mitglieder für nominale (Set-) Felder anzugeben, nach denen der Datentyp des Felds zu **ohne Typ** wird. Diese Option ist beim Arbeiten mit großen nominalen Feldern hilfreich. *Hinweis:* Wenn das Messniveau eines Felds auf **Ohne Typ** eingestellt ist, wird dessen Rolle automatisch auf **Keine** festgelegt. Dies bedeutet, dass die Felder nicht zur Modellierung zur Verfügung stehen.

Dimension der Variablen für neuronale Netze, Kohonen netze und K-Means-Modellierung begrenzen. Wählen Sie diese Option, um die maximale Anzahl Mitglieder für nominale Felder anzugeben, die in neuronalen Netzen, Kohonen netzen und in der K-Means-Modellierung verwendet werden. Die Standard-Set-Größe beträgt 20. Danach wird das Feld ignoriert und eine Warnung ausgegeben, die Informationen zum fraglichen Feld enthält.

Quellenknoten bei Ausführung aktualisieren. Wählen Sie diese Option, um alle Quellenknoten automatisch zu aktualisieren, wenn der aktuelle Stream ausgeführt wird. Diese Aktion entspricht dem Klicken auf die Schaltfläche Aktualisieren an einem Quellenknoten, außer dass diese Option automatisch alle Quellenknoten für den aktuellen Stream aktualisiert (mit Ausnahme von Benutzereingabeknoten).

Hinweis: Durch die Auswahl dieser Option werden die Caches von abwärts liegenden Knoten auch dann geleert, wenn sich die Daten nicht geändert haben. Die Leerung findet jedoch nur einmal pro Ausführung des Streams statt, was bedeutet, dass Sie abwärts liegende Caches auch weiterhin als temporären Speicher für eine Einzelausführung verwenden können. Angenommen, Sie haben z. B. einen Cache nach einer komplexen Ableitungsoperation in der Stream-Mitte gesetzt und mehrere Diagramme und Berichte sind abwärts dieses Ableitungsknotens miteinander verbunden. Bei der Ausführung des Streams wird der Cache am Ableitungsknoten geleert und erneut gefüllt, jedoch nur für das erste Diagramm oder den ersten Bericht. Nachfolgende Endknoten lesen die Daten aus dem Cache des Ableitungsknotens.

Feld- und Wertelabels in Ausgabe anzeigen. Zeigt Feld- und Wertelabels in Tabellen, Diagrammen und anderen Ausgaben an. Wenn keine Beschriftungen vorhanden sind, werden stattdessen die Feldnamen und Datenwerte angezeigt. Die Beschriftungen werden standardmäßig deaktiviert; Sie können die Beschriftungen jedoch an anderer Stelle in IBM® SPSS® Modeler auf Einzelbasis ändern. Sie können auch festlegen, dass die Beschriftungen im Ausgabefenster mithilfe einer Umschalttaste auf der Symbolleiste angezeigt werden

Abbildung 5-16

Symbol der Symbolleiste zum Wechseln von Feld- und Wertelabels



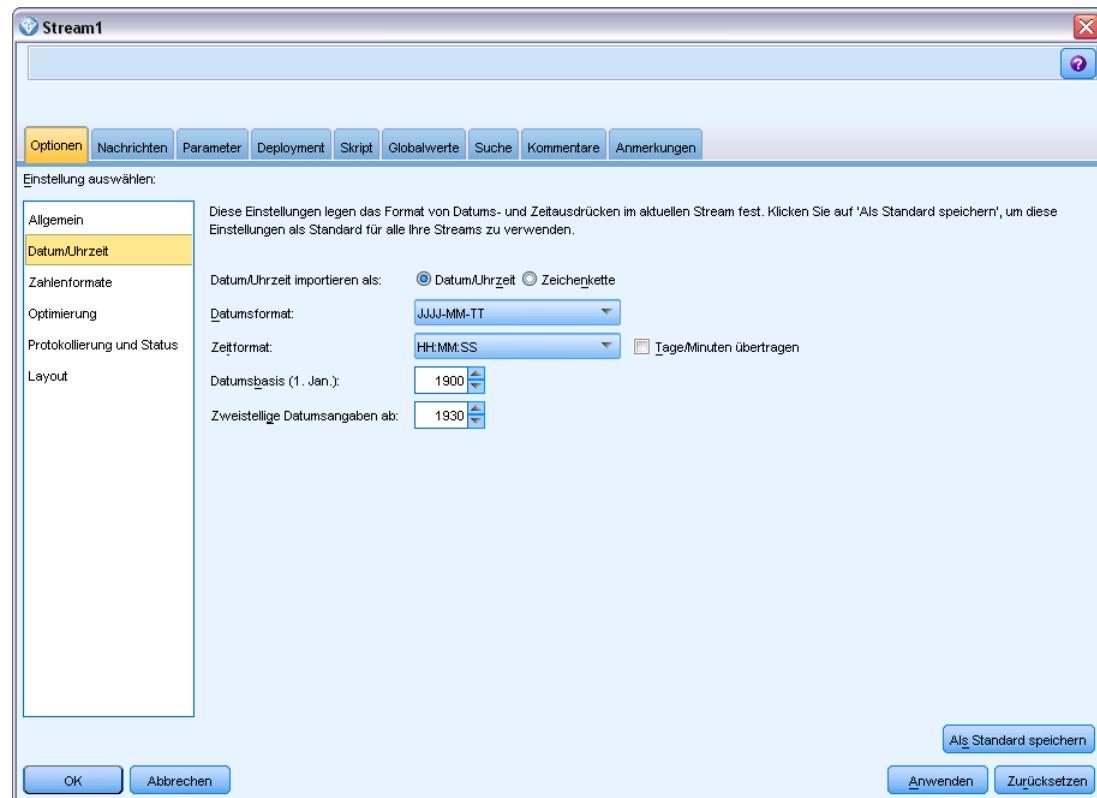
Ausführungszeiten anzeigen. Zeigt die einzelnen Ausführungszeiten für Stream-Knoten auf der Registerkarte "Ausführungszeiten" an, nachdem der Stream ausgeführt wurde. [Für weitere Informationen siehe Thema Anzeigen der Knoten-Ausführungszeiten auf S. 75.](#)

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

Festlegen von Datums- und Zeitoptionen für Streams

Diese Optionen geben das für verschiedene Datums- und Zeitausdrücke im aktuellen Stream zu verwendende Format an.

Abbildung 5-17
Festlegen von Datums- und Zeitoptionen für einen Stream



Datum/Uhrzeit importieren als. Wählen Sie, ob der Speichertyp “Datum/Uhrzeit” für Datums-/Uhrzeitfelder verwendet werden soll oder ob sie als String-Variablen importiert werden sollen.

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

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

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

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

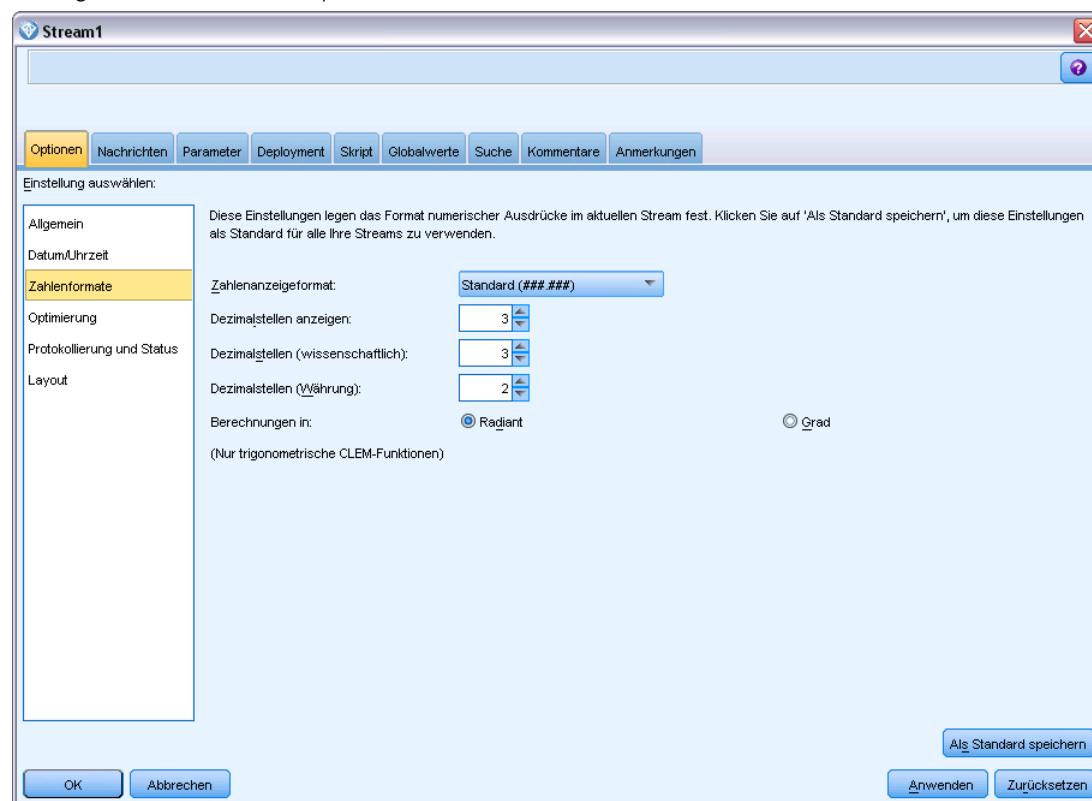
Zweistellige Datumsangaben ab. Geben Sie das Schwellenjahr an, ab dem Jahrhundertziffern für Jahre hinzugefügt werden, die mit lediglich zwei Ziffern angegeben sind. Wenn Sie beispielsweise 1930 als Schwellenjahr festlegen, wird angenommen, dass das Datum 05/11/02 im Jahr 2002 liegt. Für Daten nach 30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

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

Festlegen von Zahlenformatoptionen für Streams

Diese Optionen geben das für verschiedene numerische Ausdrücke im aktuellen Stream zu verwendende Format an.

Abbildung 5-18
Festlegen von Zahlenformatoptionen für einen Stream



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

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

Berechnungen in. Wählen Sie Radiant oder Grad als Maßeinheit für trigonometrische CLEM-Ausdrücke. [Für weitere Informationen siehe Thema Trigonometrische Funktionen in Kapitel 8 auf S. 156.](#)

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

Festlegen von Optimierungsoptionen für Streams

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

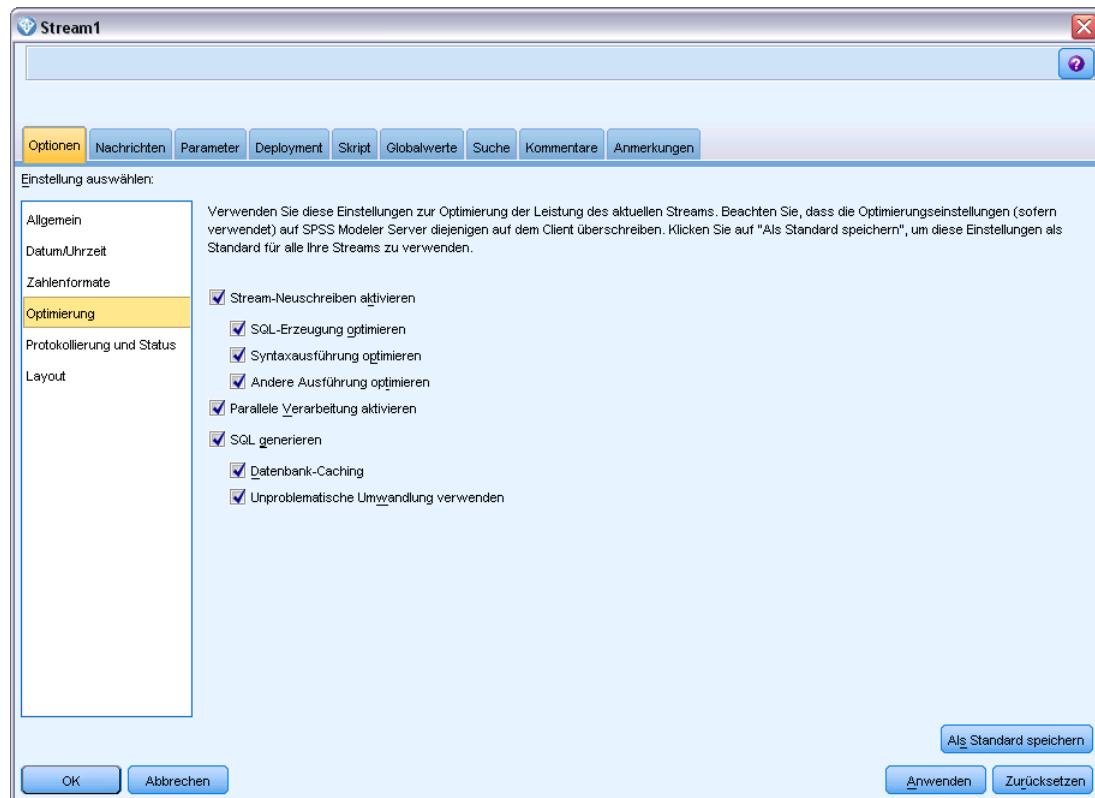
Hinweis: Für Datenbankmodellierung und SQL-Optimierung muss auf dem IBM® SPSS® Modeler-Computer die SPSS Modeler Server-Konnektivität aktiviert sein. Wenn diese Einstellung aktiviert ist, können Sie auf Datenbankalgorithmen zugreifen, SQL direkt aus SPSS Modeler per Pushback übertragen und auf SPSS Modeler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

Hilfe > Info > Zusätzliche Details

Wenn Konnektivität aktiviert ist, wird auf der Registerkarte “Lizenzstatus” die OptionServeraktivierung angezeigt.

[Für weitere Informationen siehe Thema Verbindung mit IBM SPSS Modeler Server in Kapitel 3 auf S. 16.](#)

Abbildung 5-19
Festlegen von Optionen zur Stream-Optimierung



Anmerkung: Die Unterstützung für SQL-Pushback ist von der verwendeten Datenbanktyp abhängig. Neueste Informationen zu Datenbanken und ODBC-Treibern, die für die Verwendung mit IBM® SPSS® Modeler 15 unterstützt werden und getestet wurden, finden Sie auf der unternehmenseigenen Support-Site unter <http://www.ibm.com/support>.

Stream-Neuschreiben aktivieren. Wählen Sie diese Option, um die Stream-Neuerstellung in SPSS Modeler zu aktivieren. Es stehen zwei Neuerstellungsarten zur Verfügung. Sie können eine oder beide auswählen. Die Stream-Neuerstellung ordnet die Knoten in einem Stream für eine effizientere Operation im Hintergrund neu an, ohne die Stream-Semantik zu ändern.

- **SQL-Erzeugung optimieren.** Mit dieser Option können die Knoten im Stream so neu angeordnet werden, dass ein Pushback weiterer Operationen mithilfe der SQL-Generierung für die Ausführung in der Datenbank möglich ist. Wenn die Optimierungsfunktion einen Knoten findet, der nicht in SQL wiedergegeben werden kann, sucht sie weiter unten im Stream nach Knoten, die in SQL wiedergegeben und ohne Beeinträchtigung der Stream-Semantik vor den problematischen Knoten verschoben werden können. Nicht nur, dass die Datenbank die Operationen effizienter als SPSS Modeler durchführen kann, solche Pushbacks dienen auch dazu, die Größe des Daten-Sets, das an SPSS Modeler zur Verarbeitung zurückgegeben wird, zu reduzieren. So wird wiederum der Netzverkehr reduziert und die Geschwindigkeit der Stream-Operationen erhöht. Das Kontrollkästchen SQL generieren muss aktiviert sein, damit die SQL-Optimierung funktionieren kann. Für weitere Informationen siehe

Thema SQL-Optimierung in Kapitel 6 in *IBM SPSS Modeler Server 15-Verwaltungs- und Leistungshandbuch*.

- **Syntaxausführung optimieren.** Diese Methode der Stream-Neuerstellung erhöht die Effizienz von Operationen, die mehr als einen Knoten mit IBM® SPSS® Statistics-Syntax umfassen. Die Optimierung wird durch die Zusammenfassung der Syntaxbefehle in einer einzigen Operation erreicht, anstatt jeden einzelnen Syntaxbefehl als separate Operation auszuführen.
- **Andere Ausführung optimieren.** Diese Methode der Stream-Neuerstellung erhöht die Effizienz von Operationen, die nicht an die Datenbank delegiert werden können. Eine Optimierung wird dadurch erreicht, dass die Datenmenge im Stream so früh wie möglich reduziert wird. Während die Datenintegrität gewahrt bleibt, wird der Stream neu erstellt, um die Operationen näher zur Datenquelle zu bringen. So werden die Daten abwärts für kostenintensive Operationen, z. B. Joins, reduziert.
- Parallele Verarbeitung aktivieren.** Bei Ausführung auf einem Computer mit mehreren Prozessoren ermöglicht diese Option dem System, die Last auf diese Prozessoren zu verteilen, was zu einer schnelleren Leistung führen kann. Bei der Verwendung mehrerer Knoten oder der folgenden Einzelknoten können Sie von der parallelen Verarbeitung profitieren: C5.0, Verbinden (Zusammenführungsknoten, durch Schlüssel), Sortieren, Intervall (Klasserknoten, Rangordnung und Perzentile) sowie Aggregieren (unter Verwendung eines oder mehrerer Schlüsselfelder).
- SQL generieren.** Wählen Sie diese Option zur Aktivierung der SQL-Erzeugung. Dabei wird zugelassen, dass Stream-Operationen per Pushback an die Datenbank zurückübertragen werden, indem SQL-Code zur Generierung der Ausführungsprozesse verwendet wird, was zu einer Leistungssteigerung führen kann. Zur weiteren Steigerung der Leistung kann außerdem SQL-Erzeugung optimieren ausgewählt werden. Dadurch wird die Anzahl der Operationen, die per Pushback an die Datenbank zurückübertragen werden, maximiert. Wenn Operationen für einen Knoten per Pushback an die Datenbank zurückübertragen wurden, wird der Knoten bei der Ausführung des Streams violett gekennzeichnet.
- **Datenbank-Caching.** Bei Streams, die in der Datenbank auszuführende SQL generieren, muss nicht das Dateisystem als Cache verwendet werden, sondern die Daten können mitten im Stream in einer temporären Tabelle in der Datenbank zwischengespeichert werden. Bei einer Kombination mit SQL-Optimierung kann dies zu einer erheblichen Leistungssteigerung führen. So kann beispielsweise die Ausgabe aus einem Stream, der mehrere Tabellen zusammenfasst, um eine Data-Mining-Ansicht zu erstellen, im Cache gespeichert und bei Bedarf wieder verwendet werden. Klicken Sie bei aktiviertem Datenbank-Caching einfach mit der rechten Maustaste auf einen Knoten (keinen Endknoten), um die Daten an dieser Stelle im Cache zu speichern, und der Cache wird bei der nächsten Ausführung des Streams automatisch unmittelbar in der Datenbank erstellt. Dadurch kann SQL für Knoten weiter unten im Stream erzeugt werden, was zu weiteren Leistungsverbesserungen führt. (Alternativ kann diese Option bei Bedarf deaktiviert werden, beispielsweise wenn es aufgrund von Richtlinien oder Berechtigungen nicht möglich ist, Daten in die Datenbank zu schreiben. Wenn Datenbank-Caching oder SQL-Optimierung nicht aktiviert ist, wird der Cache stattdessen in das Dateisystem geschrieben. Für weitere Informationen siehe Thema [Cachen von Optionen für Knoten auf S. 57](#).
- **Unproblematische Umwandlung verwenden.** Mit dieser Option können Daten von Zeichenketten in Zahlen oder von Zahlen in Zeichenketten konvertiert werden, sofern sie in einem geeigneten Format gespeichert sind. Wenn die Daten beispielsweise als Zeichenkette in

der Datenbank gespeichert sind, tatsächlich jedoch eine sinnvolle Zahl enthalten, können die Daten beim Eintreten des Pushbacks zur Verwendung konvertiert werden.

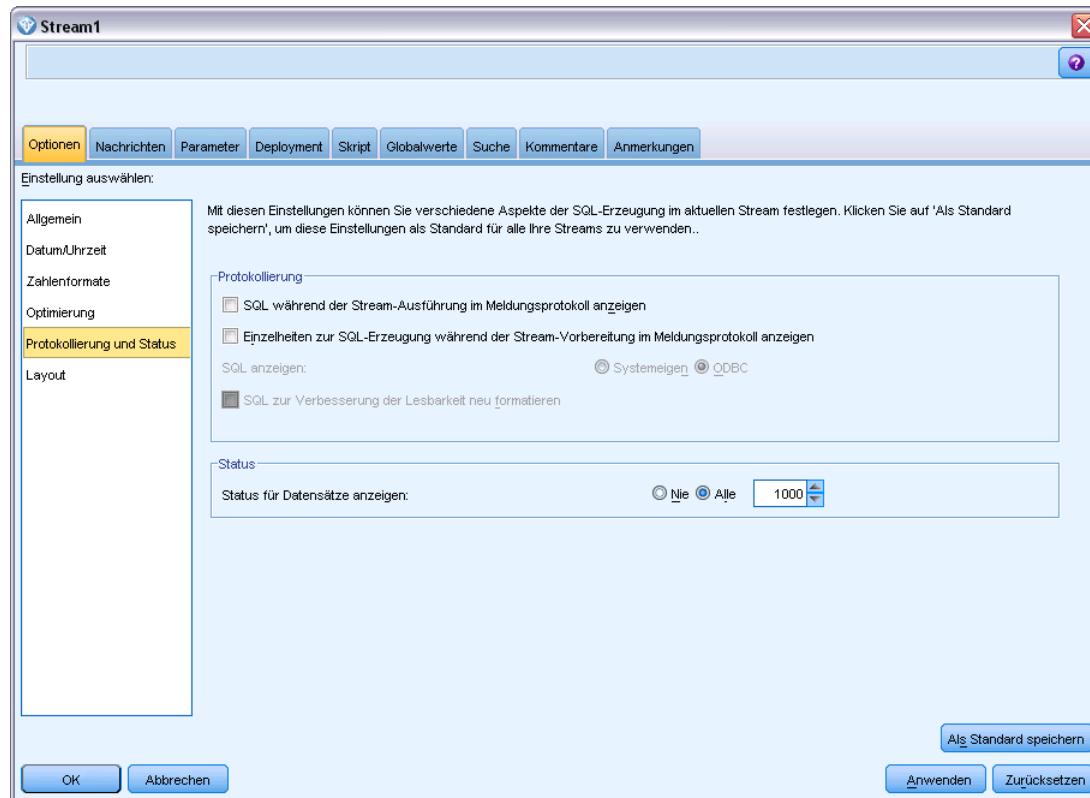
Hinweis: Aufgrund kleinerer Unterschiede in der SQL-Implementierung weichen die Ergebnisse der in einer Datenbank ausgeführten Streams möglicherweise ein wenig von den bei der Ausführung in SPSS Modeler gelieferten Ergebnissen ab. Aus ähnlichen Gründen können diese Unterschiede auch je nach Datenbankanbieter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

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

Festlegen von Optionen für SQL-Protokollierung und Datensatzstatus für Streams

Zu diesen Einstellungen gehören verschiedene Optionen zur Festlegung der Anzeige der vom Stream generierten SQL-Anweisungen sowie der Anzeige für die Anzahl der vom Stream verarbeiteten Datensätze.

Abbildung 5-20
Festlegen von Optionen für SQL-Protokollierung und Datensatzstatus für einen Stream



SQL wird während der Stream-Ausführung im Meldungsprotokoll angezeigt. Gibt an, ob die während der Ausführung des Streams erzeugte SQL an das Meldungsprotokoll übergeben wird.

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

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

SQL zur Verbesserung der Lesbarkeit neu formatieren. Legt fest, ob im Protokoll angezeigte SQL zur Verbesserung der Lesbarkeit formatiert werden soll.

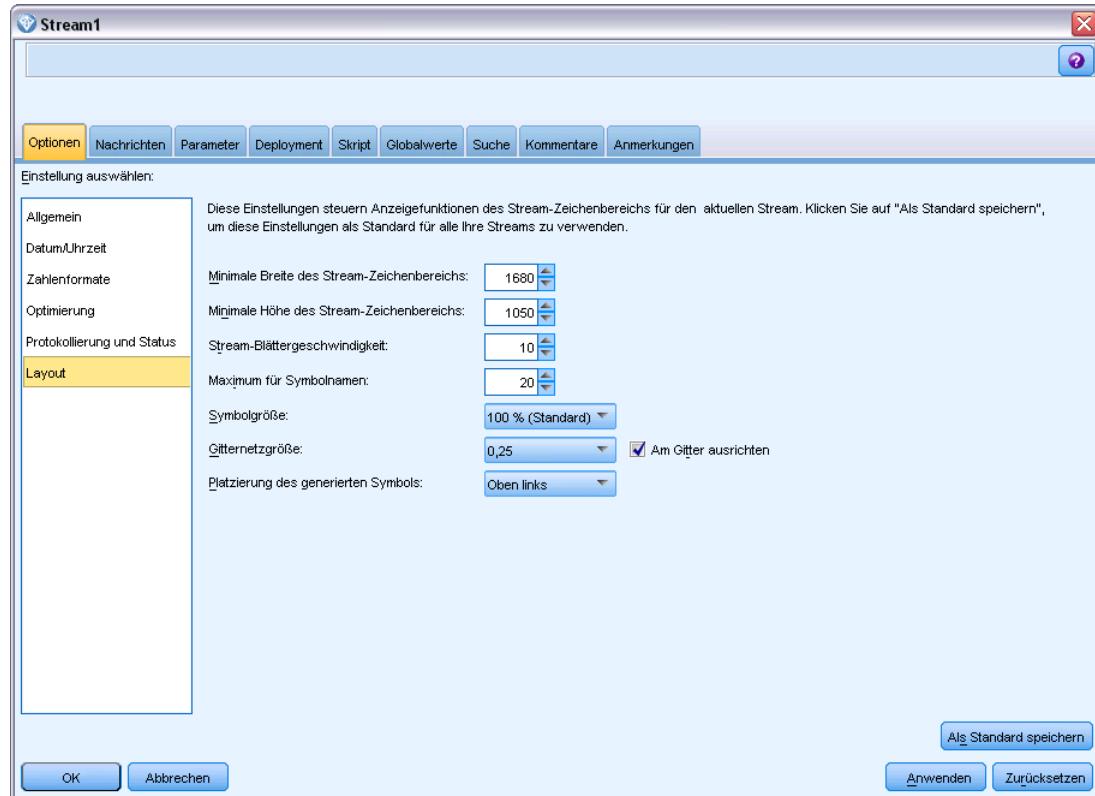
Status für Datensätze anzeigen. Legt fest, ob Datensätze gemeldet werden sollen, wenn sie an Endknoten ankommen. Geben Sie eine Zahl an, um den Status alle N Datensätze zu aktualisieren.

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

Festlegen von Layoutoptionen für Streams

Diese Einstellungen bieten eine Reihe von Optionen, die sich auf Anzeige und Verwendung des Stream-Zeichenbereichs beziehen.

Abbildung 5-21
Festlegen von Layoutoptionen für einen Stream



Minimale Breite des Stream-Zeichenbereichs. Geben Sie die minimale Breite des Stream-Zeichenbereichs in Pixeln an.

Minimale Höhe des Stream-Zeichenbereichs. Geben Sie die minimale Höhe des Stream-Zeichenbereichs in Pixeln an.

Stream-Blättergeschwindigkeit. Geben Sie die Blättergeschwindigkeit für den Stream-Zeichenbereich an, um die Blättergeschwindigkeit des Stream-Zeichenbereichs zu steuern, wenn ein Knoten von einer Position im Zeichenbereich an eine andere gezogen wird. Je höher die Zahl, desto höher die Blättergeschwindigkeit.

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

Symbolgröße. Wählen Sie eine Option aus, um die gesamte Stream-Ansicht auf eine Anzahl an Größenwerten zwischen 8 und 200 % der Standardsymbolgröße zu skalieren.

Gitternetzgröße. Wählen Sie eine Gitternetzgröße in der Liste aus. Diese Zahl wird für die Ausrichtung von Knoten im Stream-Zeichenbereich mithilfe eines unsichtbaren Gitters verwendet. Die Standardgitternetzgröße lautet 0,25.

Am Gitter ausrichten. Wählen Sie diese Option, um Symbole am unsichtbaren Gitternetz auszurichten (Standardauswahl).

Generierte Symbolplatzierung. Wählen Sie im Zeichenbereich die gewünschte Position für Symbole der Knoten, die aus Modell-Nuggets erzeugt wurden. Standard ist oben links.

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

Anzeigen von Stream-Operationsmeldungen

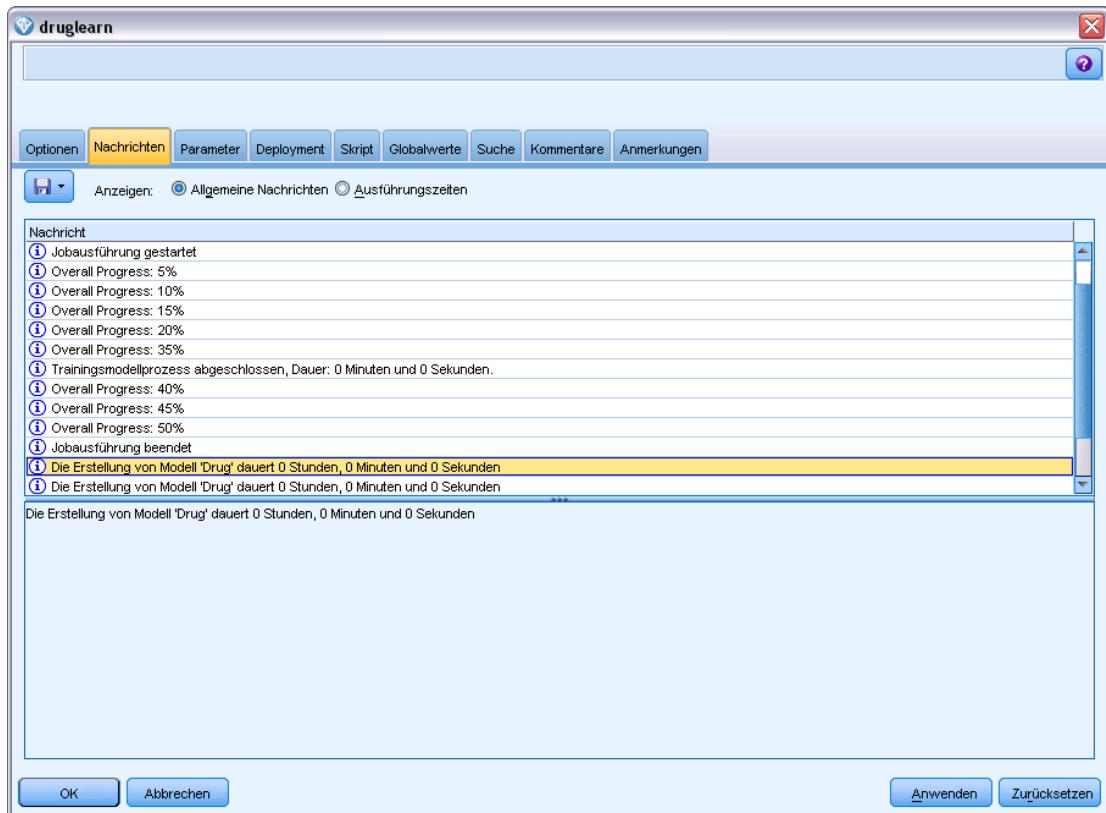
Nachrichten zu Stream-Operationen wie Ausführung, Optimierung und vergangene Zeit für die Modelleinstellung und -evaluierung können problemlos mithilfe der Registerkarte "Nachrichten" im Dialogfeld "Stream-Eigenschaften" angezeigt werden. In dieser Tabelle werden auch Fehlermeldungen gespeichert.

So zeigen Sie Stream-Nachrichten an:

- ▶ Klicken Sie im Menü "Datei" auf die Option Stream-Eigenschaften (oder wählen Sie den Stream in der Registerkarte "Streams" im Manager-Bereich aus, klicken Sie mit der rechten Maustaste und wählen Sie dann im Popup-Menü die Option Stream-Eigenschaften).
- ▶ Klicken Sie auf die Registerkarte Nachrichten.

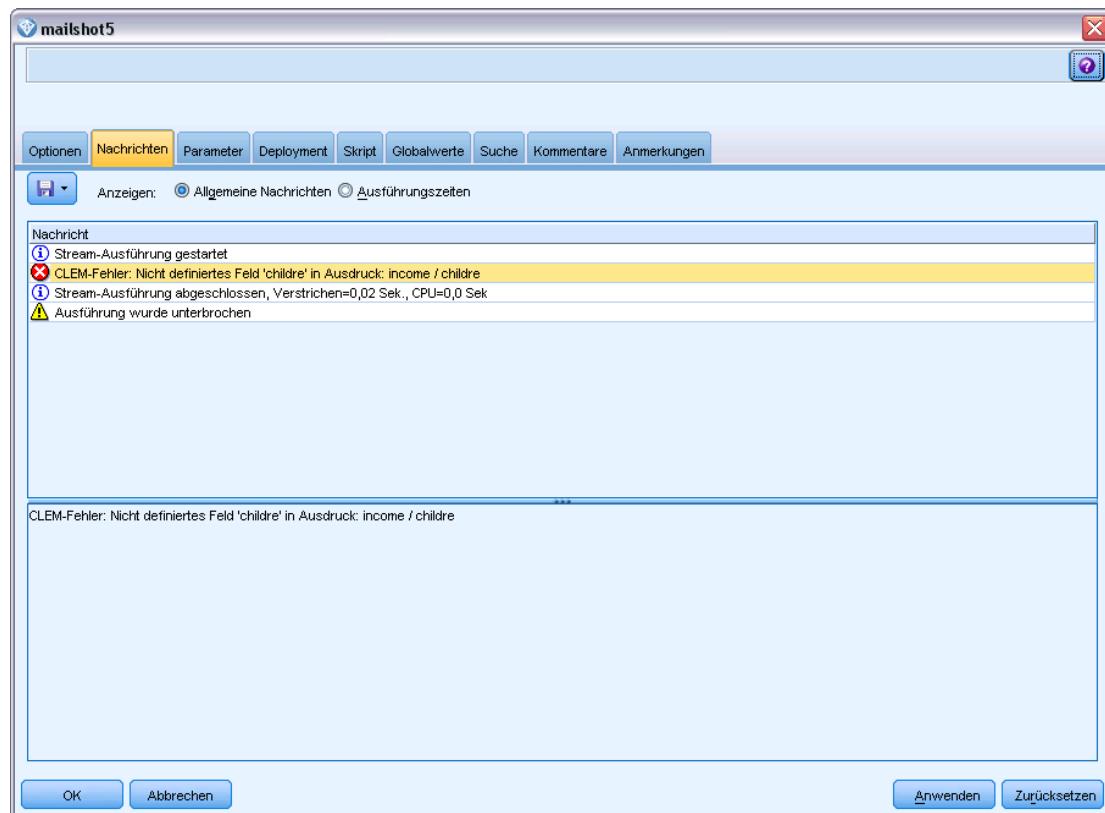
Klicken Sie als Alternative im Menü "Extras" auf:
Stream-Eigenschaften > Nachrichten

Abbildung 5-22
Registerkarte "Nachrichten" im Dialogfeld "Stream-Eigenschaften"



Zusätzlich zu den Nachrichten zu Stream-Operationen werden hier Fehlermeldungen gespeichert. Wenn die Stream-Ausführung aufgrund eines Fehlers beendet wird, wird in diesem Dialogfeld die Registerkarte "Nachrichten" mit der sichtbaren Fehlermeldung aufgerufen. Außerdem wird der Knoten mit den Fehlern im Stream-Zeichenbereich rot markiert.

Abbildung 5-23
Stream-Ausführung mit Fehlerbericht



Falls die SQL-Optimierungs- und -Protokollierungsoptionen im Dialogfeld “Benutzeroptionen” aktiviert sind, werden auch Informationen über die generierte SQL angezeigt. [Für weitere Informationen siehe Thema Festlegen von Optimierungsoptionen für Streams auf S. 68.](#)

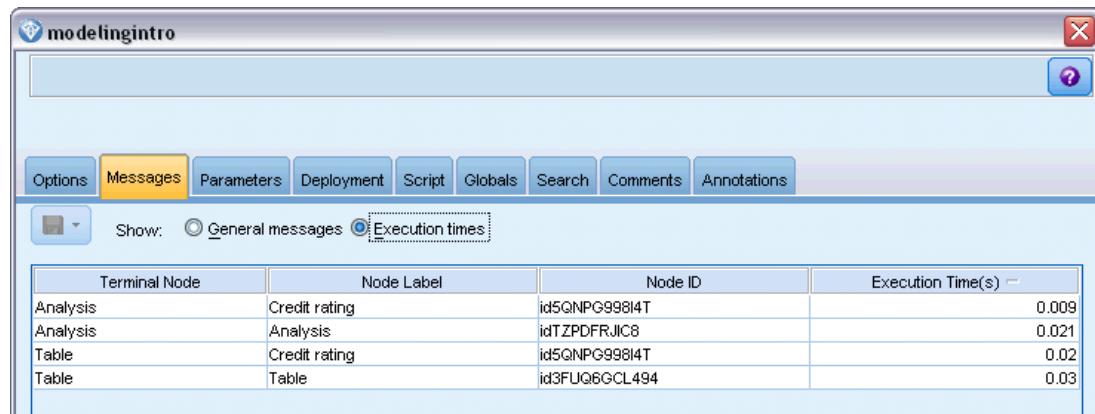
Sie können hier für einen Stream gemeldete Nachrichten speichern, indem Sie in der Dropdown-Liste der Schaltfläche “Speichern” (links, direkt unter dem Register “Nachrichten”) auf Nachrichten speichern klicken. Sie können auch alle Nachrichten für einen bestimmten Stream löschen, indem Sie in dieser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

Anzeigen der Knoten-Ausführungszeiten

Auf der Registerkarte “Meldungen” können Sie außerdem auswählen, dass die Registerkarte “Ausführungszeiten” angezeigt werden soll, auf der die einzelnen Ausführungszeiten für alle Knoten im Stream zu sehen sind.

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

Abbildung 5-24
Anzeigen der Ausführungszeiten für Knoten im Stream



The screenshot shows a software interface titled "modelingintro". At the top, there is a menu bar with tabs: Options, Messages (which is highlighted in yellow), Parameters, Deployment, Script, Globals, Search, Comments, and Annotations. Below the menu bar, there is a toolbar with a dropdown arrow and two radio buttons: "General messages" and "Execution times" (which is selected). A large table below the toolbar contains four rows of data. The columns are labeled: Terminal Node, Node Label, Node ID, and Execution Time(s). The data is as follows:

Terminal Node	Node Label	Node ID	Execution Time(s)
Analysis	Credit rating	id5QNP998I4T	0.009
Analysis	Analysis	idTZPDFRJC8	0.021
Table	Credit rating	id5QNP998I4T	0.02
Table	Table	id3FUQ6GCL494	0.03

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

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

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

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

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

Festlegen der Stream- und Sitzungsparameter

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

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

- Stream-Parameter können in einem Stream-Skript oder im Dialogfeld “Stream-Eigenschaften” festgelegt werden und sie stehen für alle Knoten im Stream zur Verfügung. Sie werden in der Parameterliste im Expression Builder angezeigt.
- Sitzungsparameter können in einem Standalone-Skript oder im Dialogfeld “Sitzungsparameter” festgelegt werden. Sie stehen allen Streams zur Verfügung, die in der aktuellen Sitzung verwendet werden (alle Streams auf der Registerkarte “Streams” im Manager-Bereich).

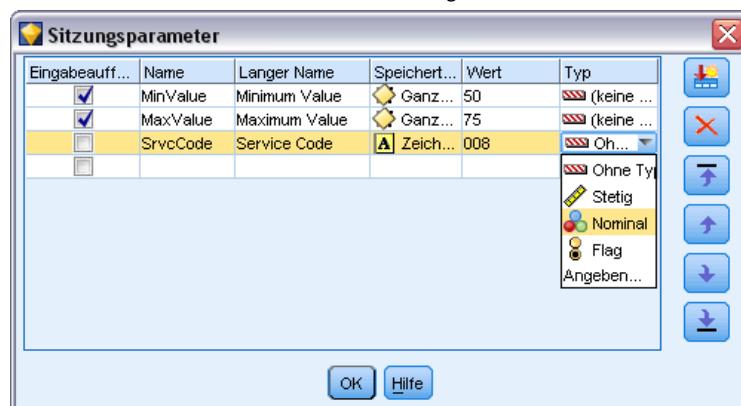
Parameter können auch für Superknoten festgelegt werden. In diesem Fall sind sie nur für Knoten sichtbar, die in dem betreffenden Superknoten ver kapselt sind. [Für weitere Informationen siehe Thema Festlegen von Superknoten-Parametern in Kapitel 9 in IBM SPSS Modeler 15 Quellen-, Prozess- und Ausgabeknoten.](#)

Informationen zum Festlegen und Verwenden von Parametern in Skripts finden Sie hier: Stream-, Sitzungs- und Superknoten-Parameter in Kapitel 3.

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

- ▶ Zum Festlegen der Stream-Parameter klicken Sie im Hauptmenüs auf:
Werkzeuge > Stream-Eigenschaften > Parameter
- ▶ Zum Festlegen der Sitzungsparameter klicken Sie im Menü “Extras” auf die Option Sitzungsparameter festlegen.

Abbildung 5-25
Definieren von Parametern für die Sitzung



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

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

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

Speichertyp. Wählen Sie einen Speichertyp aus der Liste aus. Der Speichertyp gibt an, wie die Datenwerte im Parameter gespeichert werden. Wenn Sie z. B. mit Werten arbeiten, die führende Nullen enthalten und die Sie beibehalten möchten (wie 008), sollten Sie Zeichenkette als Speichertyp wählen. Andernfalls werden die Nullen vom Wert abgezogen. Verfügbare Speichertypen sind „Zeichenkette“, „Ganze Zahl“, „Reelle Zahl“, „Uhrzeit“, „Datum“ und „Zeitstempel“. Beachten Sie, dass bei Datumsparametern die Werte gemäß der im nächsten Absatz erläuterten ISO-Standardnotierung eingegeben werden müssen.

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

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

Die Optionen „Langer Name“, „Speichertyp“ und „Typ“ können nur über die Benutzeroberfläche für Parameter festgelegt werden. Die Festlegung dieser Optionen mithilfe von Skripts ist nicht möglich.

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

Festlegen von Eingabeaufforderungen für Parameterwerte während der Laufzeit

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

Abbildung 5-26
Eingabeaufforderungen für Parameterwerte während der Laufzeit



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

Diese Eingabeaufforderungen deaktivieren. Aktivieren Sie dieses Kontrollkästchen, wenn diese Eingabeaufforderungen beim Ausführen des Streams nicht angezeigt werden sollen. Sie können sie wieder anzeigen, indem Sie das Kontrollkästchen Eingabeaufforderung? im Dialogfeld für Stream- oder Sitzungseigenschaften, in dem die Parameter definiert wurden, aktivieren. Für weitere Informationen siehe Thema Festlegen der Stream- und Sitzungsparameter auf S. 76.

Festlegen von Wertebeschränkungen für einen Parametertyp

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

Abbildung 5-27
Festlegen von verfügbaren Werten für einen Parameter



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

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

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

Stetige Messniveaus

Minimum. Geben Sie eine Untergrenze für die Parameterwerte ein.

Maximum. Geben Sie eine Obergrenze für die Parameterwerte ein.

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

Nomiale Messniveaus

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

Flag-Messniveaus

Wahr. Geben Sie einen Flag-Wert für den Parameter an, wenn die Bedingung erfüllt ist.

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

Beschriftungen. Sie können Beschriftungen für die Werte eines Flag-Felds angeben.

Deployment-Optionen für den Stream

Auf der Registerkarte “Deployment” im Dialogfeld für die Stream-Eigenschaften können Sie Optionen zum Bereitstellen des Streams als Szenario in IBM® SPSS® Collaboration and Deployment Services zur Modellaktualisierung, für automatische Aufgabenplanung oder weitere Verwendung durch IBM® Analytical Decision Management oder Predictive Applications 5.x festlegen. Alle Streams benötigen eine spezielle Scoring-Verzweigung, bevor sie bereitgestellt werden können; weitere Voraussetzungen und Optionen hängen vom Deployment-Typ ab. Für weitere Informationen siehe Thema Speichern und Bereitstellen von Repository-Objekten in Kapitel 9 auf S. 182.

Anzeigen von globalen Werten für Streams

Mit der Registerkarte “Globalwerte” im Dialogfeld “Stream-Eigenschaften” können Sie die globalen, für den aktuellen Stream festgelegten Werte anzeigen. Globalwerte werden mit einem Knoten vom Typ “Globalwerte” erstellt, um Statistiken zu ermitteln, wie Mittelwert, Summe oder Standardabweichung für ausgewählte Felder. Für weitere Informationen siehe Thema [Globalwerteknoten in Kapitel 6 in IBM SPSS Modeler 15 Quellen-, Prozess- und Ausgabeknoten](#).

Sobald der Knoten “Globalwerte” ausgeführt wird, stehen diese Werte für eine Vielzahl von Anwendungen in Stream-Operationen zur Verfügung. Für weitere Informationen siehe Thema [Globale Funktionen in Kapitel 8 auf S. 175](#).

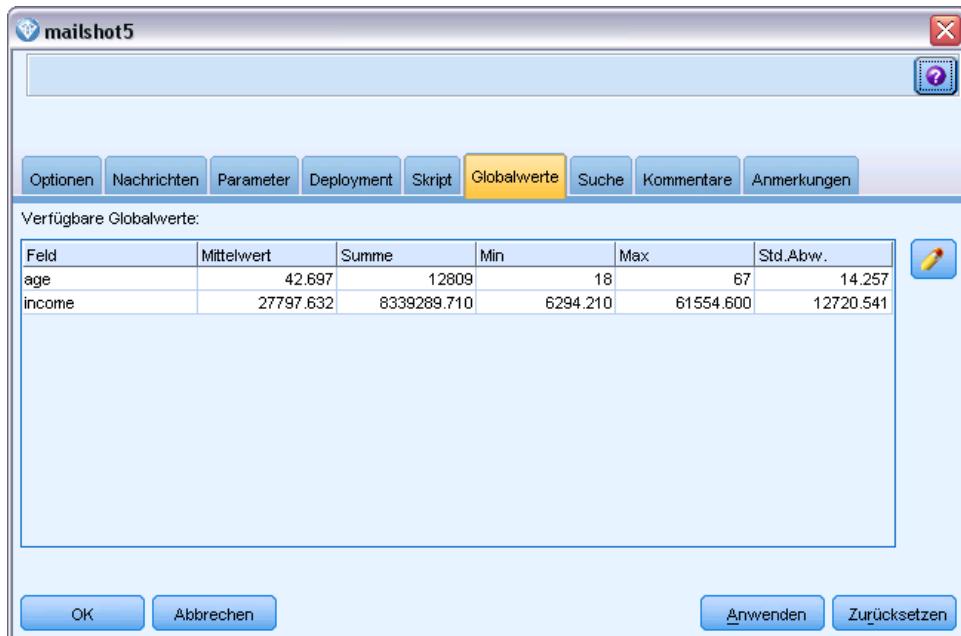
So zeigen Sie Globalwerte für einen Stream an:

- ▶ Klicken Sie im Menü “Datei” auf die Option Stream-Eigenschaften (oder wählen Sie den Stream in der Registerkarte “Streams” im Manager-Bereich aus, klicken Sie mit der rechten Maustaste und wählen Sie dann im Popup-Menü die Option Stream-Eigenschaften).
- ▶ Klicken Sie auf die Registerkarte Globalwerte.

Klicken Sie als Alternative im Menü “Extras” auf:

Stream-Eigenschaften > Globalwerte

Abbildung 5-28
Anzeigen von globalen, für den Stream verfügbaren Werten



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

Suchen nach Knoten in einem Stream

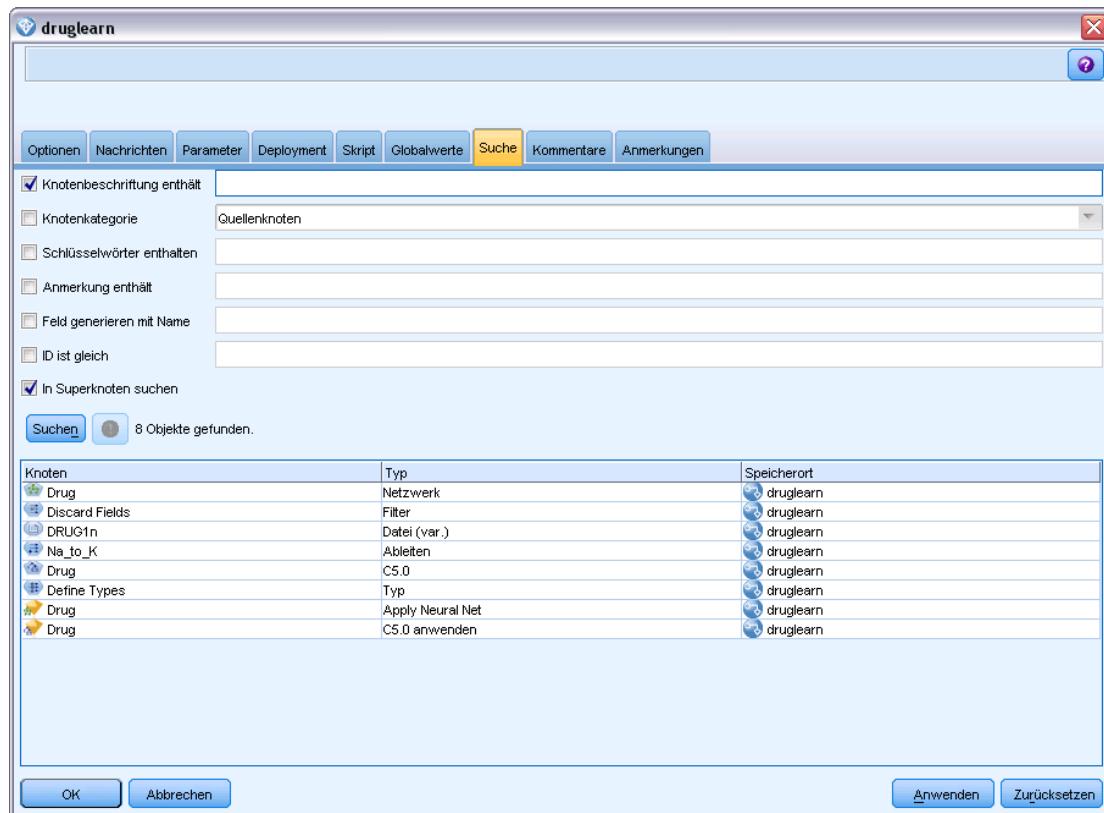
Sie können Knoten in einem Stream suchen, indem Sie eine Reihe von Suchkriterien wie Knotenname, Kategorie und Bezeichner angeben. Diese Funktion ist besonders nützlich für komplexe Streams, die eine große Anzahl an Knoten enthalten.

So suchen Sie Knoten in einem Stream

- ▶ Klicken Sie im Menü “Datei” auf die Option Stream-Eigenschaften (oder wählen Sie den Stream in der Registerkarte “Streams” im Manager-Bereich aus, klicken Sie mit der rechten Maustaste und wählen Sie dann im Popup-Menü die Option Stream-Eigenschaften).
- ▶ Klicken Sie auf die Registerkarte Suche.

Klicken Sie als Alternative im Menü “Extras” auf:
Stream-Eigenschaften > Suchen

Abbildung 5-29
Suchen nach Knoten in einem Stream



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

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

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

Schlüsselwörter enthalten. Aktivieren Sie dieses Kontrollkästchen und geben Sie eines oder mehrere vollständige Schlüsselwörter ein, um Knoten zu suchen, für die diese Schlüsselwörter in der Registerkarte "Anmerkungen" im Knotendialogfeld eingegeben wurden. Der eingegebene Schlüsselworttext muss eine exakte Übereinstimmung sein. Trennen Sie mehrere Schlüsselwörter durch Semikolons, um nach Alternativen zu suchen (z. B. werden mit proton;neutron alle Knoten mit einem dieser Schlüsselwörter gefunden). Für weitere Informationen siehe Thema [Anmerkungen auf S. 97](#).

Anmerkung enthält. Aktivieren Sie dieses Kontrollkästchen und geben Sie eines oder mehrere Wörter ein, um Knoten zu suchen, die diesen Text im Haupttextbereich der Registerkarte “Anmerkungen” im Knotendialogfeld enthalten. Suchvorgänge unterscheiden Groß-/Kleinschreibung nicht und mehrere Wörter werden als ein einziges Textstück betrachtet. Für weitere Informationen siehe Thema Anmerkungen auf S. 97.

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

ID ist gleich.. Markieren Sie dieses Kontrollkästchen und geben Sie eine Knoten-ID ein, um einen bestimmten Knoten mit diesem Bezeichner zu finden. (Durch Auswählen dieser Option werden alle vorherigen Optionen deaktiviert.) Knoten-IDs werden beim Erstellen eines Knotens vom System zugewiesen. Sie können als Verweis auf den Knoten zur Skripterstellung oder Automatisierung benutzt werden. Geben Sie nur eine Knoten-ID ein, die eine exakte Übereinstimmung sein muss. Für weitere Informationen siehe Thema Anmerkungen auf S. 97.

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

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

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

Umbenennen von Streams

Auf der Registerkarte “Anmerkungen” im Dialogfeld “Stream-Eigenschaften” können Sie beschreibende Anmerkungen für einen Stream hinzufügen und einen benutzerdefinierten Namen für den Stream erstellen. Diese Optionen sind besonders hilfreich, wenn Sie Berichte für Streams generieren, die dem Projektbereich hinzugefügt werden. Für weitere Informationen siehe Thema Anmerkungen auf S. 97.

Stream-Beschreibungen

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

Abbildung 5-30
Einführung der Stream-Beschreibung

Stream-Dokument	newschancart_commented
Erstellt am:	27. Januar 2006 01:17
Erstellt von:	llanders
Zuletzt gespeichert am:	28. Mai 2010 20:51
Zuletzt gespeichert von:	djones

Inhalt

[Beschreibung & Kommentar](#)
[Scoring-Informationen](#)
[Modellierungsinformationen](#)

Beschreibung & Kommentar

Kommentar
This stream predicts whether customers are likely to subscribe to our interactive news service, based on characteristics of customers who have already subscribed
This is the source data about our existing customers
This is where we tell the stream what we are trying to predict

Die Stream-Beschreibung wird als HTML-Dokument mit verschiedenen Abschnitten angezeigt.

Allgemeine Stream-Informationen

Dieser Abschnitt enthält den Stream-Namen sowie Details über seine Erstellung und die letzte Speicherung.

Beschreibung und Kommentare

Dieser Abschnitt enthält Folgendes:

- Stream-Anmerkungen (siehe [Anmerkungen auf S. 97](#))
- Nicht mit bestimmten Knoten verknüpfte Kommentare
- Kommentare, die sowohl in der Modellerstellungs- als auch der Scoring-Verzweigung des Streams mit Knoten verknüpft sind

Scoring-Informationen

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

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

Modellinformationen.

Enthält Informationen zur Modellierungsverzweigung des Streams.

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

Vorschau der Stream-Beschreibungen

Sie können den Inhalt einer Stream-Beschreibung in einem Web-Browser anzeigen, indem Sie im Dialogfeld für die Stream-Eigenschaften auf eine Option klicken. Der Inhalt der Beschreibung hängt von den auf der Registerkarte „Deployment“ des Dialogfelds angegebenen Optionen ab. [Für weitere Informationen siehe Thema Deployment-Optionen für den Stream in Kapitel 9 auf S. 208.](#)

So zeigen Sie eine Stream-Beschreibung an:

- ▶ Klicken Sie im IBM® SPSS® Modeler-Hauptmenü auf:
Werkzeuge > Stream-Eigenschaften > Bereitstellung
- ▶ Geben Sie Deployment-Typ, den zugehörigen Scoring-Knoten sowie eventuelle Scoring-Parameter an.
- ▶ Für den Deployment-Typ “Modellaktualisierung” können Sie optional Folgendes auswählen:
 - Modellierungsknoten und eventuelle Modellerstellungsparameter
 - Modell-Nugget in der Scoring-Verzweigung des Streams
- ▶ Klicken Sie auf die Schaltfläche Vorschau der Stream-Beschreibung anzeigen.

Exportieren von Stream-Beschreibungen

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

So exportieren Sie eine Stream-Beschreibung:

- ▶ Klicken Sie im Hauptmenü auf Folgendes:
Datei > Stream-Beschreibung exportieren
- ▶ Geben Sie einen Namen für die HTML-Datei ein und klicken Sie auf Speichern.

Ausführen von Streams

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

- Klicken Sie im Menü “Extras” auf Ausführen.
- Klicken Sie auf eine der Ausführen...-Schaltflächen in der Symbolleiste. Mit diesen Schaltflächen können Sie den gesamten Stream ausführen oder lediglich den ausgewählten Endknoten. [Für weitere Informationen siehe Thema IBM SPSS Modeler-Symbolleiste in Kapitel 3 auf S. 26.](#)
- Um einen einzelnen Daten-Stream auszuführen, klicken Sie mit der rechten Maustaste auf einen Endknoten und dann im Popup-Menü auf Ausführen.
- Um einen Teil eines Daten-Streams auszuführen, klicken Sie mit der rechten Maustaste auf einen Knoten, der kein Endknoten ist, und klicken Sie dann im Popup-Menü auf Ab hier ausführen. Dadurch werden nur die Operationen nach dem ausgewählten Knoten ausgeführt.

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

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

Abbildung 5-31
Dialogfeld "Ausführungs-Feedback"



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

Arbeiten mit Modellen

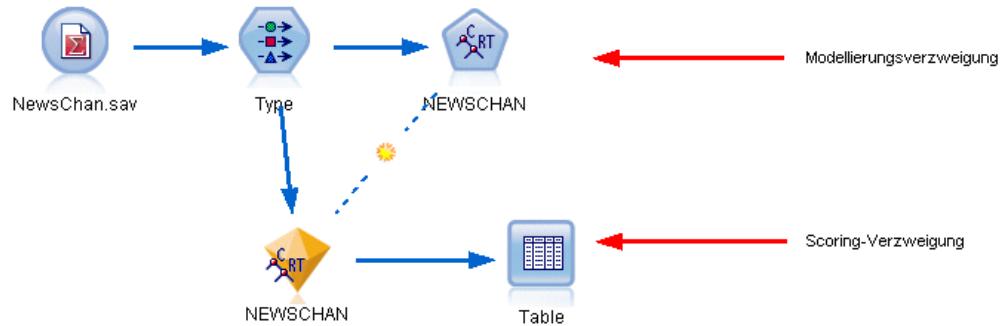
Wenn ein Stream einen Modellierungsknoten enthält (d. h. einen aus der Registerkarte "Modellierung" oder "Datenbankmodellierung" der Knotenpalette), wird bei Ausführung des Streams ein **Modell-Nugget** erstellt. Ein Modell-Nugget ist ein Behälter für ein **Modell**, d. h. das Set von Regeln, Formeln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

Abbildung 5-32
Modell-Nugget



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

Abbildung 5-33
Modellierung und Scoring von Verzweigungen in einem Stream



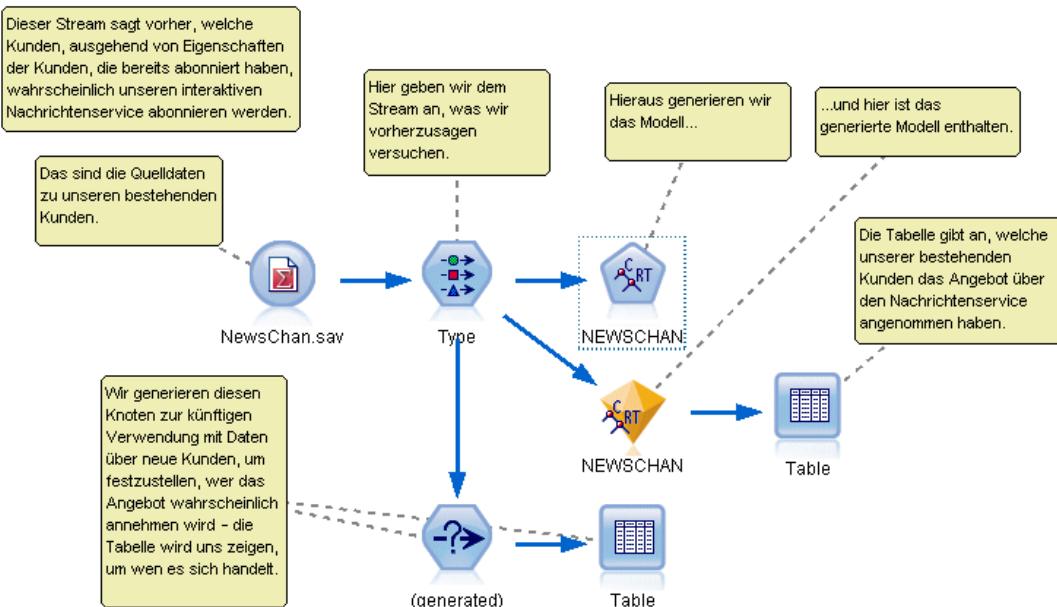
Ein typischer Modellierungs-Stream besteht aus zwei Verzweigungen. Die **Modellierungsverzweigung** enthält den Modellierungsknoten zusammen mit der Quelle und vorangehenden Verarbeitungsknoten. Die **Modellierungsverzweigung** wird beim Ausführen des Modellierungsknotens erstellt und enthält das Modell-Nugget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

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

Hinzufügen von Kommentaren und Anmerkungen zu Knoten und Streams

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

Abbildung 5-34
Stream mit Kommentaren



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

Sie können alle Kommentare für einen Stream oder Superknoten auflisten, die Reihenfolge der Kommentare in der Liste ändern, den Kommentartext bearbeiten und die Vorder- oder Hintergrundfarbe eines Kommentars ändern. [Für weitere Informationen siehe Thema Auflisten von Stream-Kommentaren auf S. 95.](#)

Sie können auch Knoten in der Form von Textanmerkungen zu Streams, Knoten und Nuggets mithilfe der Registerkarte "Anmerkungen" eines Dialogfelds mit Streeameigenschaften, eines Knotendialogfelds oder eines Modell-Nugget-Fensters hinzufügen. Diese Anmerkungen sind nur sichtbar, wenn die Registerkarte "Anmerkungen" geöffnet ist, aber die Stream-Anmerkungen lassen sich auch als Bildschirmkommentare anzeigen. [Für weitere Informationen siehe Thema Anmerkungen auf S. 97.](#)

Kommentare

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

Hinweis: Sie können Stream-Anmerkungen auch als Bildschirmkommentare anzeigen, allerdings lassen sich diese nicht an Knoten oder Nuggets anhängen. Für weitere Informationen siehe Thema [Umwandeln von Anmerkungen in Kommentare auf S. 96](#).

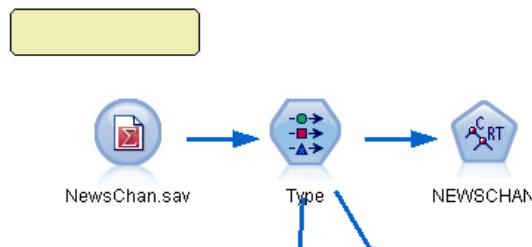
Das Aussehen des Textfelds ändert sich entsprechend dem aktuellen Modus des Kommentars (oder der als Kommentar angezeigten Anmerkung). Siehe dazu die folgende Tabelle.

Tabelle 5-1
Modi des Kommentar- und Anmerkungstextfelds

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

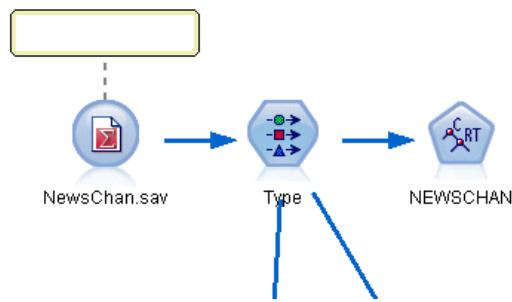
Wenn Sie einen neuen freistehenden Kommentar erstellen, wird dieser anfangs in der oberen linken Ecke des Stream-Zeichenbereichs angezeigt.

Abbildung 5-35
Neuer freistehender Kommentar



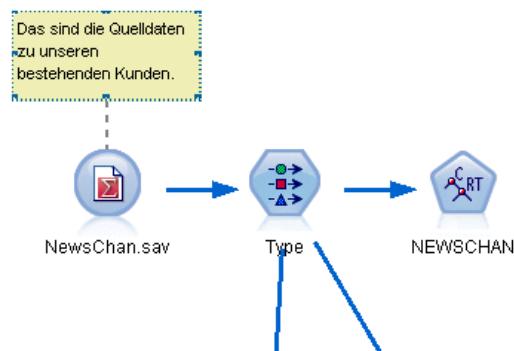
Wenn Sie einen Kommentar mit einem Knoten oder ein Nugget verknüpfen, wird der Kommentar anfangs über dem Stream-Objekt angezeigt, mit dem er verbunden ist.

Abbildung 5-36
Neuer mit Knoten verbundener Kommentar



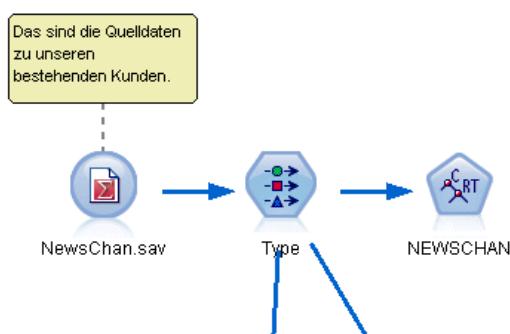
Das Textfeld ist weiß, um zu zeigen, dass sich Text eingeben lässt. Wenn Sie den Text eingegeben haben, klicken Sie außerhalb des Textfeldes. Der Kommentarhintergrund wird gelb, um zu zeigen, dass der Texteingriff abgeschlossen ist. Der Kommentar bleibt ausgewählt, damit Sie ihn gleich verschieben, löschen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

Abbildung 5-37
Kommentar im Bearbeitungsmodus



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

Abbildung 5-38
Abgeschlossener Kommentar



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

Außerdem können Sie Kommentare mit Superknoten verknüpfen. Für weitere Informationen siehe Thema [Anmerkungen für Superknoten und Umbenennen von Superknoten in Kapitel 9 in IBM SPSS Modeler 15 Quellen-, Prozess- und Ausgabeknoten](#).

Operationen mit Kommentaren

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

- Hinzufügen eines freistehenden Kommentars
- Verknüpfen eines Kommentars mit einem Knoten oder Nugget
- Bearbeiten eines Kommentars
- Größenänderung eines Kommentars
- Verschieben eines Kommentars
- Lösen der Verbindung eines Kommentars
- Löschen eines Kommentars
- Alle Kommentare für einen Stream ein-/ausblenden

So fügen Sie einen freistehenden Kommentar hinzu:

- ▶ Stellen Sie sicher, dass im Stream keine Auswahl getroffen ist.
- ▶ Führen Sie eine der folgenden Aktionen aus:
 - Klicken Sie im Hauptmenü auf Folgendes:
Einfügen > Neuer Kommentar
 - Klicken Sie mit der rechten Maustaste auf den Stream-Hintergrund und klicken Sie im Popup-Menü auf Neuer Kommentar.
 - Klicken Sie in der Symbolleiste auf die Schaltfläche für Neuer Kommentar.
- ▶ Geben Sie den Kommentartext ein (oder fügen Sie Text aus der Zwischenablage ein).
- ▶ Klicken Sie auf einen Knoten im Stream, um den Kommentar zu speichern.

So verknüpfen Sie einen Kommentar mit einem Knoten oder Nugget:

- ▶ Wählen Sie einen oder mehrere Knoten oder Nuggets im Stream-Zeichenbereich aus.
- ▶ Führen Sie eine der folgenden Aktionen aus:
 - Klicken Sie im Hauptmenü auf Folgendes:
Einfügen > Neuer Kommentar
 - Klicken Sie mit der rechten Maustaste auf den Stream-Hintergrund und klicken Sie im Popup-Menü auf Neuer Kommentar.
 - Klicken Sie in der Symbolleiste auf die Schaltfläche für Neuer Kommentar.

- ▶ Geben Sie den Kommentartext ein.
- ▶ Klicken Sie auf einen anderen Knoten im Stream, um den Kommentar zu speichern.
Sie haben außerdem folgende Möglichkeiten:
 - ▶ Einfügen eines freistehenden Kommentars (siehe vorherigen Abschnitt)
 - ▶ Führen Sie eine der folgenden Aktionen aus:
 - Wählen Sie den Kommentar aus, drücken Sie F2 und wählen Sie dann den Knoten oder das Nugget aus.
 - Wählen Sie den Knoten oder das Nugget aus, drücken Sie F2 und wählen Sie dann den Kommentar aus.
 - (Nur mit 3-Tasten-Maus) Verschieben Sie den Mauszeiger über den Kommentar, halten Sie die mittlere Taste gedrückt, ziehen Sie den Mauszeiger auf den Knoten oder das Nugget und lassen Sie die Maustaste los.

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

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

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

So bearbeiten Sie einen bestehenden Kommentar:

- ▶ Führen Sie eine der folgenden Aktionen aus:
 - Doppelklicken Sie auf das Kommentar-Textfeld.
 - Wählen Sie das Textfeld aus und drücken Sie die Eingabetaste.
 - Klicken Sie mit der rechten Maustaste auf das Textfeld, um sein Menü zu öffnen, und klicken Sie auf "Bearbeiten".
- ▶ Bearbeiten Sie den Kommentartext. Sie können die Windows-Standardtastenkombinationen zur Bearbeitung verwenden, z. B. Strg+C, um Text zu kopieren. Andere Optionen während der Bearbeitung werden im Popup-Menü für den Kommentar aufgelistet.
- ▶ Klicken Sie einmal außerhalb des Textfelds, um die Größenänderungssteuerungen anzuzeigen, und noch einmal, um den Kommentar abzuschließen.

So ändern Sie die Größe eines Kommentar-Textfelds:

- ▶ Wählen Sie den Kommentar aus, um die Größenänderungssteuerungen anzuzeigen.

- ▶ Klicken Sie auf eine Steuerung und ziehen Sie sie, um die Größe des Felds zu ändern.
- ▶ Klicken Sie außerhalb des Textfelds, um die Änderung zu speichern.

So verschieben Sie einen bestehenden Kommentar:

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

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

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

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

So lösen Sie einen Kommentar von einem Knoten oder Nugget:

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

So löschen Sie einen Kommentar:

- ▶ Wählen Sie einen oder mehrere Kommentare aus, die Sie löschen möchten.
- ▶ Führen Sie eine der folgenden Aktionen aus:
 - Drücken Sie die Löschtaste.
 - Klicken Sie mit der rechten Maustaste auf einen ausgewählten Kommentar und klicken Sie in dessen Menü auf “Löschen”.

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

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

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

- ▶ Führen Sie eine der folgenden Aktionen aus:
 - Klicken Sie im Hauptmenü auf Folgendes:
Ansicht > Kommentare
 - Klicken Sie in der Symbolleiste auf die Schaltfläche Kommentare ein-/ausblenden.

Auflisten von Stream-Kommentaren

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

Diese Liste bietet die folgenden Möglichkeiten:

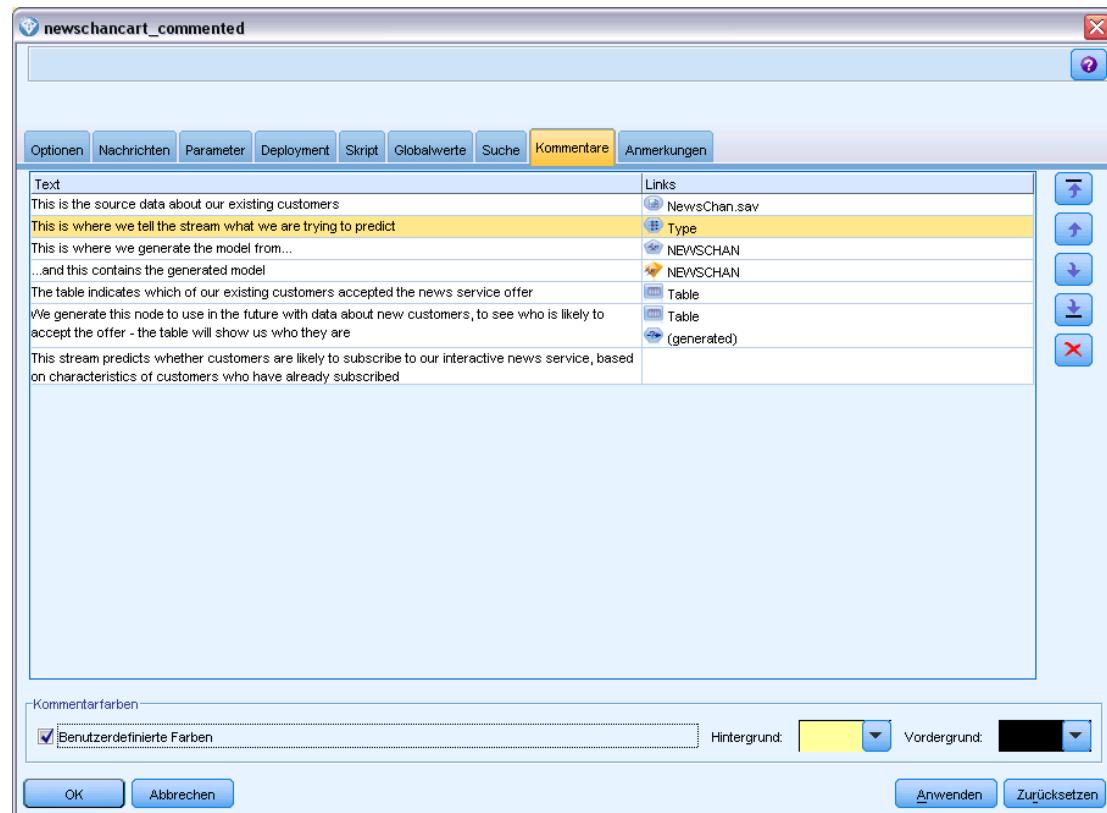
- Ändern der Reihenfolge der Kommentare
- Bearbeiten des Kommentartexts
- Ändern der Vorder- oder Hintergrundfarbe eines Kommentars

Auflisten von Kommentaren

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

- Klicken Sie im Hauptmenü auf Folgendes:
Werkzeuge > Stream-Eigenschaften > Kommentare
- Klicken Sie mit der rechten Maustaste auf einen Stream im Manager-Bereich und klicken Sie auf Stream-Eigenschaften und dann auf Kommentare.
- Klicken Sie mit der rechten Maustaste in einen Stream-Hintergrund im Zeichenbereich und klicken Sie auf Stream-Eigenschaften und dann auf Kommentare.

Abbildung 5-39
Auflisten von Kommentaren für einen Stream



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

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

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

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

Umwandeln von Anmerkungen in Kommentare

Anmerkungen zu Streams oder Superknoten können in Kommentare umgewandelt werden.

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

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

So wandeln Sie eine Stream-Anmerkung in einen Kommentar um:

- ▶ Klicken Sie im Menü “Extras” auf Stream-Eigenschaften. (Als Alternative können Sie mit der rechten Maustaste auf einen Stream im Manager-Bereich und dann auf Stream-Eigenschaften klicken.)
- ▶ Klicken Sie auf die Registerkarte Anmerkungen.
- ▶ Aktivieren Sie das Kontrollkästchen Anmerkung als Kommentar anzeigen.
- ▶ Klicken Sie auf OK.

So wandeln Sie eine Superknoten-Anmerkung in einen Kommentar um:

- ▶ Doppelklicken Sie auf das Symbol des Superknotens im Zeichenbereich.
- ▶ Klicken Sie auf die Registerkarte Anmerkungen.
- ▶ Aktivieren Sie das Kontrollkästchen Anmerkung als Kommentar anzeigen.
- ▶ Klicken Sie auf OK.

Anmerkungen

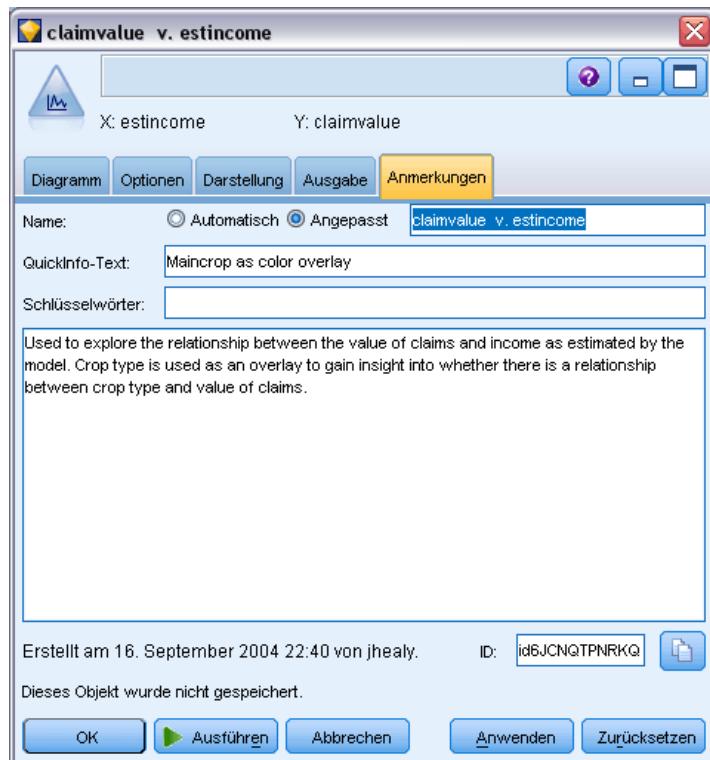
Knoten, Streams und Modelle können auf verschiedene Art und Weise mit Anmerkungen versehen werden. Sie können beschreibende Anmerkungen hinzufügen und einen benutzerdefinierten Namen angeben. Diese Optionen sind besonders hilfreich, wenn Sie Berichte für Streams generieren, die dem Projektbereich hinzugefügt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

Hinzufügen von Anmerkungen

Wenn Sie einen Knoten oder ein Modell-Nugget bearbeiten, wird ein Dialogfeld mit Registerkarten geöffnet, das eine Registerkarte “Anmerkungen” enthält, mit der Sie eine Vielzahl von Anmerkung festlegen können. Sie können die Registerkarte “Anmerkungen” auch direkt aufrufen.

- ▶ Um Anmerkungen für einen Knoten oder ein Nugget zu erstellen, klicken Sie mit der rechten Maustaste auf den Knoten bzw. das Nugget im Stream-Zeichenbereich und dann auf Umbenennen und mit Anmerkung versehen. Das Bearbeitungsdialogfeld wird so angezeigt, dass die Registerkarte “Anmerkungen” sichtbar ist.
- ▶ Um einen Stream mit Anmerkungen zu versehen, klicken Sie im Menü “Extras” auf Stream-Eigenschaften. (Als Alternative können Sie mit der rechten Maustaste auf einen Stream im Manager-Bereich und dann auf Stream-Eigenschaften klicken.) Klicken Sie auf die Registerkarte “Anmerkungen”.

Abbildung 5-40
Optionen auf der Registerkarte "Anmerkungen"



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

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

Schlüsselwörter. Geben Sie Schlüsselwörter an, die in Projektberichten und bei der Suche nach Knoten in einem Stream oder beim Verfolgen von Objekten verwendet werden sollen, die im Repository gespeichert sind (siehe Informationen zu IBM SPSS Collaboration and Deployment Services Repository auf S. 180). Mehrere Schlüsselwörter können durch Strichpunkte voneinander getrennt werden, z. B. Einkommen; Feldfruchtyp; Subventionshöhe. Leerzeichen am Anfang und am Ende von Schlüsselwörtern werden ausgelassen; Einkommen; Feldfruchtyp führt z. B. zum selben Ergebnis wie Einkommen;Feldfruchtyp. (Die Leerzeichen innerhalb von Schlüsselwörtern werden jedoch nicht ausgelassen. Zum Beispiel sind Höhe der Subventionen mit einem Leerzeichen vor "der" und Höhe der Subventionen mit zwei Leerzeichen vor "der" nicht dasselbe.)

Der Haupttextbereich kann verwendet werden, um lange Anmerkungen zu den Operationen des Knotens oder Entscheidungen im Knoten einzugeben. Wenn Sie z. B. Streams gemeinsam nutzen und erneut verwenden, ist es hilfreich, Anmerkungen zu Entscheidungen zu machen, z. B. Verwerfen eines Felds mit zahlreichen Leerzeichen unter Verwendung eines Filterknotens. Wenn Sie den Knoten mit Anmerkungen versehen, werden diese Informationen zusammen

mit dem Knoten gespeichert. Sie können diese Anmerkungen auch in einem Projektbericht berücksichtigen, der im Projektbereich erstellt wurde. [Für weitere Informationen siehe Thema Einführung in Projekte in Kapitel 11 auf S. 225.](#)

Anmerkung als Kommentar anzeigen. (Nur für Stream- und Superknoten-Anmerkungen) Markieren Sie dieses Kontrollkästchen, um die Anmerkung in einen freistehenden Kommentar umzuwandeln, der im Stream-Zeichenbereich sichtbar ist. [Für weitere Informationen siehe Thema Hinzufügen von Kommentaren und Anmerkungen zu Knoten und Streams auf S. 88.](#)

ID. Zeigt eine eindeutige ID an, die zur Referenzierung des Knotens für Skripterstellung oder Automatisierung verwendet werden kann. Dieser Wert wird bei der Erstellung des Knotens automatisch generiert und ändert sich nicht. Beachten Sie außerdem, dass zur Vermeidung einer Verwechslung mit dem Buchstaben "O" in Knoten-IDs keine Nullen verwendet werden. Mit der Kopierschaltfläche auf der rechten Seite können Sie die ID kopieren und nach Bedarf in Skripten oder anderswo einfügen. [Für weitere Informationen siehe Thema Referenzieren von Knoten in Kapitel 3 in IBM SPSS Modeler 15 Handbuch für Skripterstellung und Automatisierung.](#)

Speichern von Daten-Streams

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

So speichern Sie einen Stream:

- ▶ Klicken Sie im Menü “Datei” auf Stream speichern oder Stream speichern unter.
- ▶ Blättern Sie im Dialogfeld “Speichern” zu dem Ordner, in dem die Stream-Datei gespeichert werden soll.
- ▶ Geben Sie im Textfeld “Dateiname” einen Namen für den Stream ein.
- ▶ Wählen Sie Zu Projekt hinzufügen, wenn Sie den gespeicherten Stream zum aktuellen Projekt hinzufügen möchten.

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

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

Speichern von Statusangaben

Zusätzlich zu Streams können Sie **Statusangaben** speichern, u. a. das aktuell angezeigte Stream-Diagramm und alle Modell-Nuggets, die Sie erstellt haben (auf der Registerkarte “Modelle” im Manager-Bereich aufgelistet).

So speichern Sie einen Status:

- ▶ Klicken Sie im Menü “Datei” auf:
State > Status speichern oder Status speichern unter
 - ▶ Blättern Sie im Dialogfeld “Speichern” zu dem Ordner, in dem die Statusdatei gespeichert werden soll.
- Wenn Sie auf Speichern klicken, wird der Status mit der Erweiterung *.cst im angegebenen Verzeichnis gespeichert.

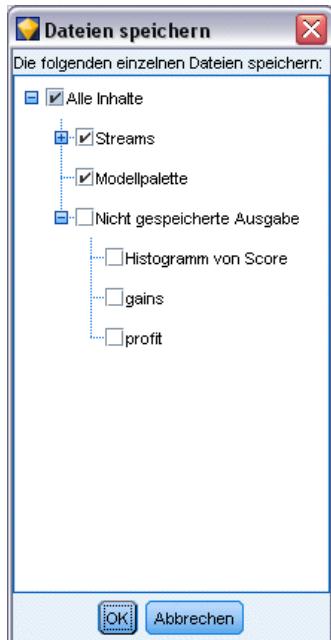
Speichern von Knoten

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

Speichern mehrerer Stream-Objekte

Wenn Sie IBM® SPSS® Modeler verlassen, obwohl mehrere Objekte, wie Streams, Projekte oder Modell-Nuggets, nicht gespeichert sind, werden Sie zum Speichern aufgefordert, bevor Sie die Software vollständig beenden. Wenn Sie auswählen, dass Objekte gespeichert werden sollen, wird ein Dialogfeld mit Optionen zum Speichern der einzelnen Objekte geöffnet.

Abbildung 5-41
Speichern mehrerer Objekte



- ▶ Aktivieren Sie einfach die Kontrollkästchen für die Objekte, die Sie speichern möchten.
- ▶ Klicken Sie auf OK, um die Objekte am gewünschten Speicherort zu speichern.

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

Speichern der Ausgabe

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

- ▶ Um die zu speichernde Ausgabe anzuzeigen, klicken Sie in den Menüs des Ausgabefensters auf: Datei > Speichern
- ▶ Geben Sie einen Namen und einen Speicherort für die Ausgabedatei an.
- ▶ Wählen Sie optional Datei zu Projekt hinzufügen im Dialogfeld “Speichern”, um die Datei in das aktuelle Projekt aufzunehmen. Für weitere Informationen siehe [Thema Einführung in Projekte in Kapitel 11 auf S. 225](#).

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

Verschlüsseln und Entschlüsseln von Informationen

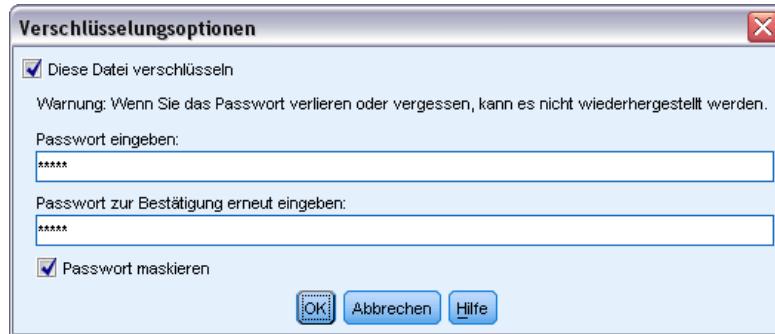
Beim Speichern von Streams, Knoten, Projekten, Ausgabedateien oder Modell-Nuggets können Sie das betreffende Element verschlüsseln, um es vor Verwendung durch Unbefugte zu schützen. Dazu wählen Sie beim Speichern eine zusätzliche Option aus und legen ein Passwort für das zu speichernde Element fest.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

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

So verschlüsseln Sie ein Objekt:

- ▶ Klicken Sie im Dialogfeld “Speichern” für das zu verschlüsselnde Element auf Optionen. Das Dialogfeld “Verschlüsselungsoptionen” wird geöffnet.

Abbildung 5-42
Verschlüsselungsoptionen beim Speichern einer Datei



- ▶ Wählen Sie die Option **Diese Datei verschlüsseln**.
- ▶ Für noch größere Sicherheit können Sie optional **Passwort maskieren** wählen. Dadurch werden alle Ihre Eingaben nur als eine Reihe von Punkten angezeigt.
- ▶ Geben Sie das Passwort ein. *Achtung:* Wenn Sie das Passwort vergessen, kann die Datei bzw. das Modell nicht mehr geöffnet werden.
- ▶ Wenn Sie **Passwort maskieren** ausgewählt haben, müssen Sie das Passwort erneut eingeben, um sicherzustellen, dass die erste Eingabe korrekt war.
- ▶ Klicken Sie auf **OK**, um zum Dialogfeld “Speichern” zurückzukehren.

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

Laden von Dateien

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

- Streams (.str)
- Statusangaben (.cst)
- Modelle (.gm)
- Modellpalette (.gen)
- Knoten (.nod)
- Ausgabe (.cou)
- Projekte (.cpj)

Öffnen neuer Dateien

Streams können direkt im Menü “Datei” geladen werden.

- ▶ Klicken Sie im Menü “Datei” auf Stream öffnen.

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

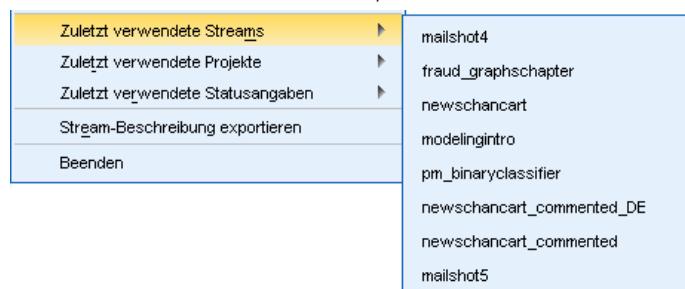
Models > Modell öffnen oder Modellpalette laden

Öffnen der zuletzt verwendeten Dateien

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

Abbildung 5-43

Öffnen der zuletzt verwendeten Optionen im Menü “Datei”



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

Zuordnen von Daten-Streams

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

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

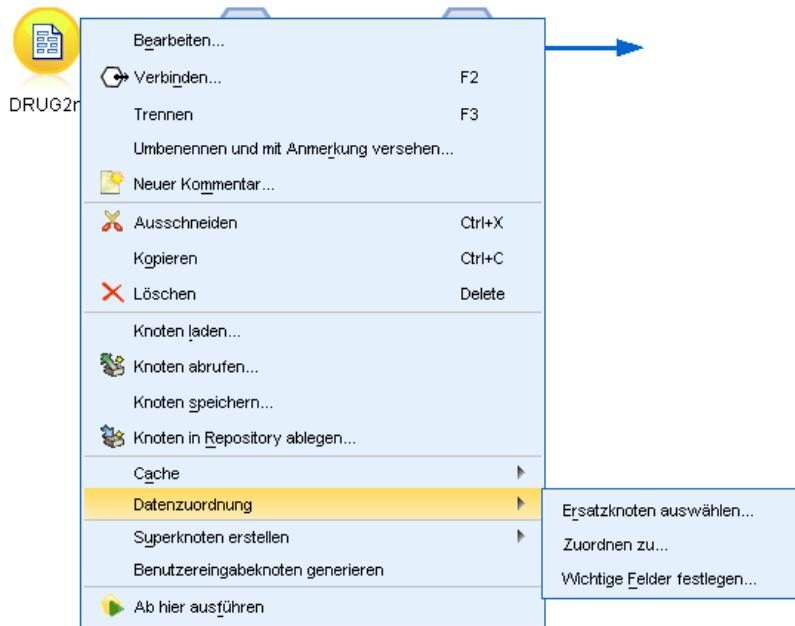
Es gibt zwei gleichwertige Möglichkeiten zum Zuordnen von Daten:

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

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

Stattdessen sollte der Stream auf die normale Art und Weise mit dem Mergeknoten verbunden werden.

Abbildung 5-44
Auswählen von Datenzuordnungsoptionen



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

Zuordnen von Daten zu einer Vorlage

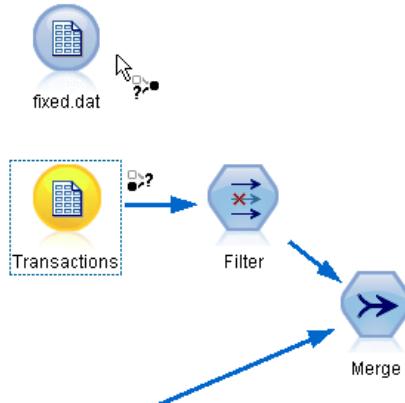
Um die Datenquelle für einen Vorlagen-Stream durch einen neuen Quellenknoten zu ersetzen, der Ihre eigenen Daten in IBM® SPSS® Modeler einbringt, sollten Sie aus dem Popup-Menü “Datenzuordnung” die Option Ersatzknoten auswählen verwenden. Diese Option steht für alle Knoten mit Ausnahme von Merge-, Aggregat-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

Schritt 1: Geben Sie die erforderlichen Felder im ursprünglichen Quellenknoten an. Damit die Stream-Operationen ordnungsgemäß ausgeführt werden können, müssen die erforderlichen Felder angegeben werden. [Für weitere Informationen siehe Thema Festlegen von wichtigen Feldern auf S. 107.](#)

Schritt 2: Fügen Sie dem Stream-Zeichenbereich eine neue Datenquelle hinzu. Importieren Sie mithilfe eines Quellenknotens die neuen Ersatzdaten.

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

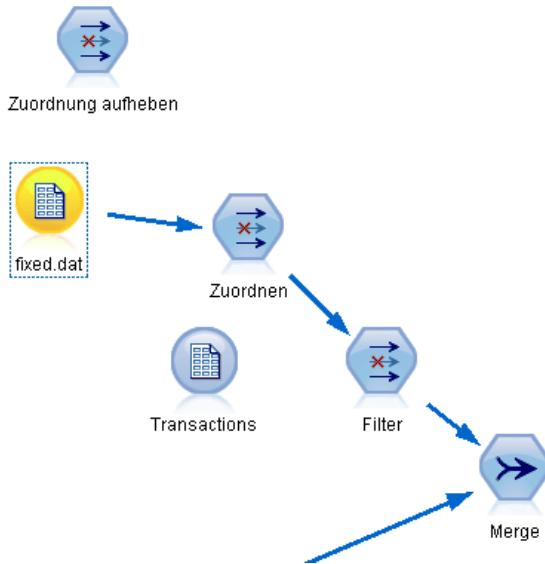
Abbildung 5-45
Auswahl eines Ersatzquellenknotens



Schritt 4: Prüfen Sie die zugeordneten Felder. Prüfen Sie in dem Dialogfeld, das nun geöffnet wird, ob die Software die Felder aus der Ersatzdatenquelle dem Stream ordnungsgemäß zuordnet. Sämtliche nicht zugeordnete, aber erforderlichen Felder werden rot dargestellt. Diese Felder werden in Stream-Operationen verwendet und müssen durch ein ähnliches Feld in der neuen Datenquelle ersetzt werden, damit die weiter unten im Stream liegenden Operationen ordnungsgemäß ausgeführt werden können. Für weitere Informationen siehe Thema Prüfen von zugeordneten Feldern auf S. 108.

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

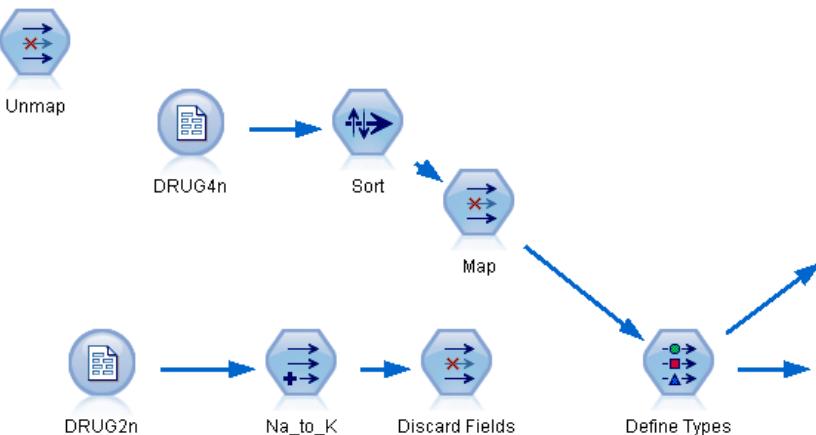
Abbildung 5-46
Neue, dem Stream erfolgreich zugeordnete Datenquelle



Zuordnen zwischen Streams

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

Abbildung 5-47
Zuordnen eines Streams von seinem Sortierknoten zum Typknoten eines anderen Streams



So ordnen Sie Daten zwischen Streams zu:

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

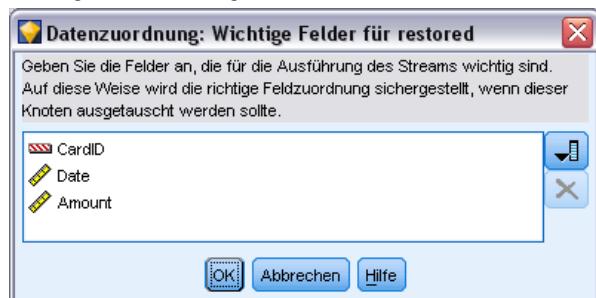
Festlegen von wichtigen Feldern

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

So legen Sie die wichtigen Felder fest:

- ▶ Klicken Sie mit der rechten Maustaste auf den Quellenknoten des Vorlagen-Streams, der ersetzt werden soll.
- ▶ Klicken Sie im Menü auf:
Datenuordnung > Festlegen von wichtigen Feldern

Abbildung 5-48
Festlegen von wichtigen Feldern

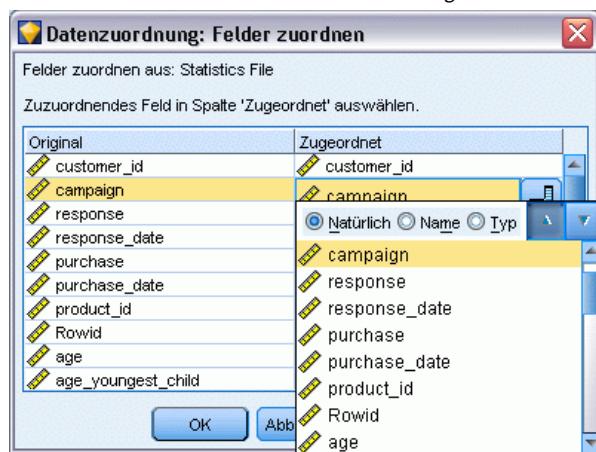


- ▶ Mit der Feldauswahl-Schaltfläche können Sie der Liste Felder hinzufügen oder entfernen. Um die Feldauswahl zu aktivieren, klicken Sie auf das Symbol rechts neben der Feldliste.

Prüfen von zugeordneten Feldern

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

Abbildung 5-49
Auswählen von Feldern für die Zuordnung



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

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

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

Tipps und Direktzugriffstasten

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

- **Erstellen Sie Streams schnell durch Doppelklicken.** Doppelklicken Sie einfach auf einen Knoten in der Palette, um ihn zum aktuellen Stream hinzuzufügen und zu verbinden.
- **Verwenden Sie Tastenkombinationen zur Auswahl von weiter unten liegenden Knoten.** Drücken Sie Strg+Q und Strg+W, um alle weiter unten liegenden Knoten aus- bzw. abzuwählen.

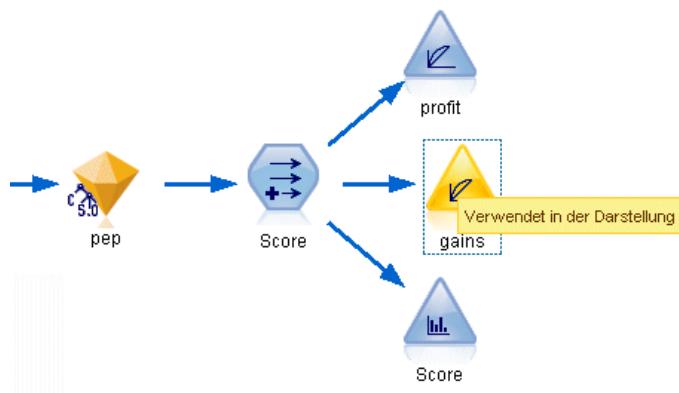
- **Verwenden Sie Direktzugriffstasten, um Knoten zu verbinden und zu trennen.** Wenn ein Knoten im Zeichenbereich ausgewählt ist, drücken Sie F2, um eine Verbindung zu beginnen, drücken Sie die Tabulatortaste, um zum gewünschten Knoten zu gelangen, und Umschalt+Leertaste, um die Verbindung fertig zu stellen. Drücken Sie F3, um alle Ein- und Ausgaben vom ausgewählten Knoten zu trennen.
- **Passen Sie die Registerkarte „Knotenpalette“ mit Ihren Lieblingsknoten an.** Klicken Sie im Menü „Extras“ auf Paletten verwalten, um ein Dialogfeld zu öffnen, in dem Sie die in der Knotenpalette angezeigten Knoten hinzufügen, entfernen bzw. verschieben können.

Abbildung 5-50
Paletten-Manager



- **Umbenennen von Knoten und Hinzufügen von QuickInfo.** Jedes Knotendialogfeld enthält eine Registerkarte „Anmerkungen“, auf der Sie einen benutzerdefinierten Namen für Knoten im Zeichenbereich angeben können sowie QuickInfo hinzufügen können, wodurch Ihr Stream einfacher organisiert wird. Sie können auch lange Anmerkungen zur Fortschrittsverfolgung einfügen, Prozessdetails speichern und sämtliche Geschäftsentscheidungen notieren, die erforderlich sind oder umgesetzt wurden.

Abbildung 5-51
QuickInfo und benutzerdefinierter Knotenname



- **Automatisches Einfügen von Werten in einen CLEM-Ausdruck.** Mit dem Expression Builder, den Sie in einer Vielzahl von Dialogfeldern aufrufen können (z. B. Dialogfelder für Ableitungs- und Filterknoten), können Sie automatisch Feldwerte in einen CLEM-Ausdruck einfügen. Klicken Sie auf die Werteschaltfläche im Expression Builder, um eine Auswahl aus den bereits vorhandenen Feldwerten zu treffen.

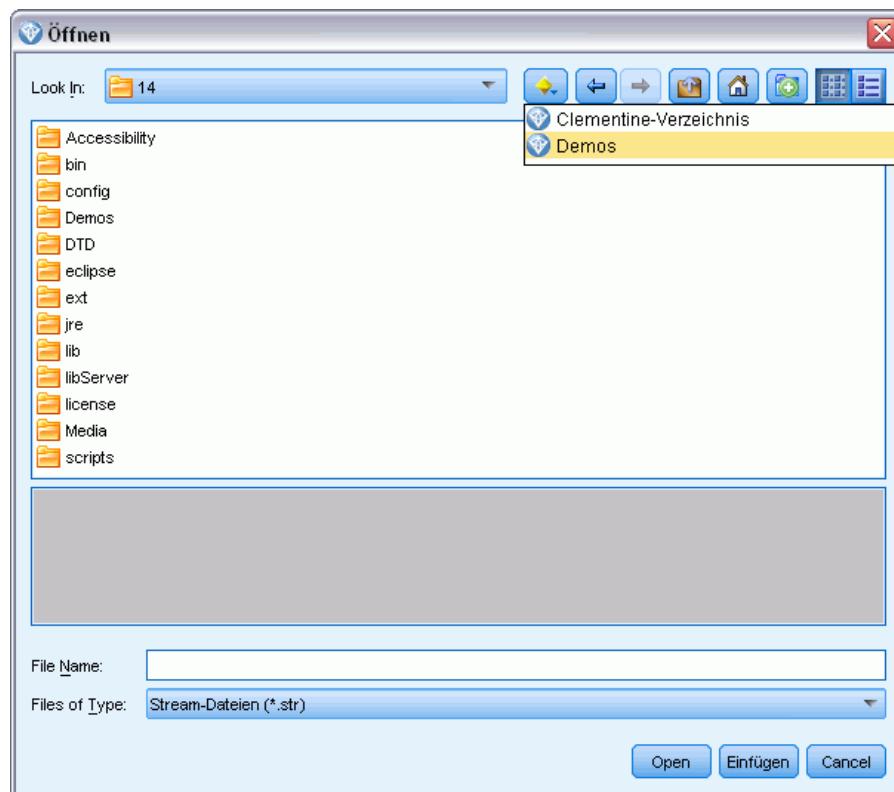
Abbildung 5-52
Werteschaltfläche



- **Schnelles Suchen nach Dateien.** Beim Suchen von Dateien in einem Öffnen-Dialogfeld verwenden Sie die Liste “Datei” (klicken Sie auf das gelbe Rautensymbol), um zuvor verwendete Verzeichnisse sowie IBM® SPSS® Modeler-Standardverzeichnisse aufzurufen. Verwenden Sie die Vor- und Zurückschaltflächen, um durch die Verzeichnisse zu blättern.

Abbildung 5-53

Auswahl des Ordners "Demos" in der Liste der zuletzt angezeigten Verzeichnisse



- **Minimieren überflüssiger Elemente im Ausgabefenster.** Sie können die Ausgabe schnell schließen und entfernen, indem Sie die rote X-Schaltfläche verwenden, die Sie in allen Ausgabefenstern in der rechten oberen Ecke finden. So können Sie nur die vielversprechenden und interessanten Ergebnisse auf der Registerkarte "Ausgaben" des Manager-Bereichs beibehalten.

Für die Software steht eine vollständige Palette mit Direktzugriffstasten zur Verfügung. Für weitere Informationen siehe Thema Tastatureingabehilfen in Anhang A auf S. 266.

Wussten Sie, dass Sie ...

- ... mithilfe der Maus eine Gruppe von Knoten in den Stream-Zeichenbereich ziehen und auswählen können?
- ... Knoten von einem Stream zu einem anderen kopieren und einfügen können?
- ... in jedem Dialogfeld und Ausgabefenster die Hilfe aufrufen können?
- ... Hilfe zu CRISP-DM, dem Standardverfahren für das Data Mining, das dem Industriestandard entspricht, aufrufen können? (Klicken Sie dazu im Hilfemenü auf CRISP-DM-Hilfe.)

Umgang mit fehlenden Werten

Überblick über fehlende Werte

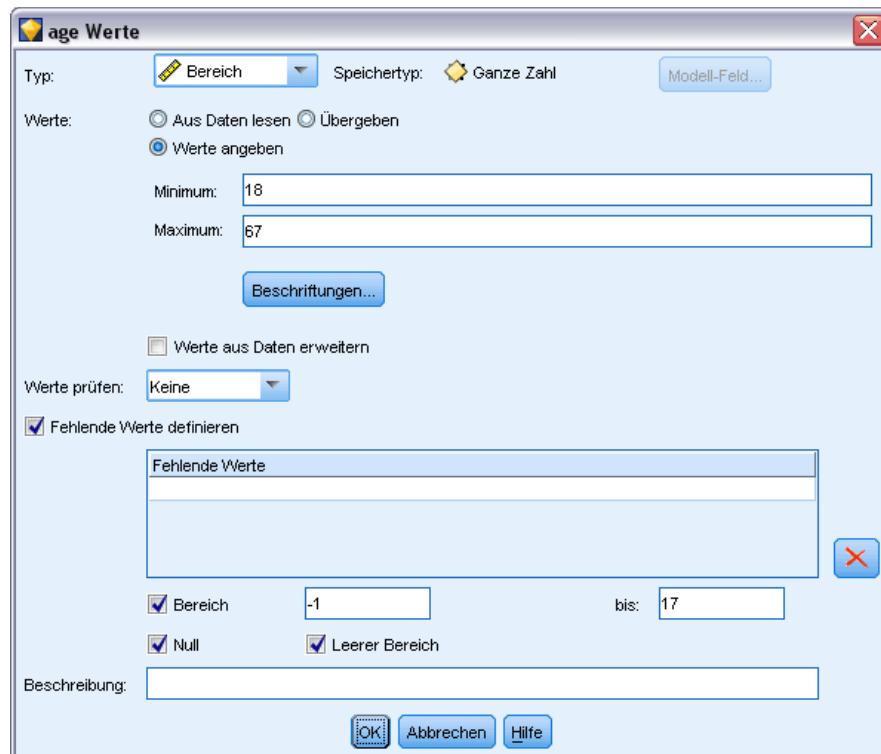
In der Datenvorbereitungsphase des Data-Mining kann es häufig vorkommen, dass Sie fehlende Werte in den Daten ersetzen möchten. **Fehlende Werte** sind Werte im Daten-Set, die unbekannt, nicht erfasst oder falsch eingegeben sind. In der Regel sind solche Werte für die entsprechenden Felder ungültig.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

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

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

- **Nullwerte oder systemdefiniert fehlende Werte.** Bei diesen Werten handelt es sich um Nicht-Zeichenketten-Werte, die in der Datenbank bzw. der Quelldatei leer gelassen und nicht speziell in einem Quellen- oder Typknoten als "fehlend" definiert wurden. Systemdefinierte fehlende Werte werden als \$null\$ angezeigt. Beachten Sie, dass leere Zeichenketten in SPSS Modeler nicht als Nullen betrachtet werden, auch wenn sie von bestimmten Datenbanken (siehe unten) als Nullen behandelt werden können.
- **Leere Zeichenketten und leere Bereiche.** Leere Zeichenkettenwerte und leere Bereiche (Zeichenketten ohne sichtbare Zeichen) werden anders als Nullwerte behandelt. Leere Zeichenketten werden in den meisten Fällen als äquivalent mit leeren Bereichen (Leerzeichen) behandelt. Beispiel: Wenn Sie die Option auswählen, dass leere Bereiche in einem Quellen- oder Typknoten als Leerstellen behandelt werden sollen, gilt diese Einstellung auch für leere Zeichenketten.
- **Leere oder benutzerdefiniert fehlende Werte.** Es handelt sich hierbei um Werte wie unbekannt, 99 oder -1, die in einem Quellen- oder Typknoten ausdrücklich als fehlend definiert sind. Optional können Sie auch auswählen dass Nullen und leere Bereiche als Leerzeichen behandelt werden sollen. Dadurch können sie mit Flags für eine spezielle Behandlung versehen und aus den meisten Berechnungen ausgeschlossen werden. Beispielsweise können Sie die Funktion @BLANK verwenden, um diese Werte gemeinsam mit anderen Arten von fehlenden Werten als Leerstellen zu behandeln. [Für weitere Informationen siehe Thema Verwenden des Dialogfelds "Werte" in Kapitel 4 in IBM SPSS Modeler 15 Quellen-, Prozess- und Ausgabeknoten.](#)

Abbildung 6-1
Festlegen von fehlenden Werten für eine stetige Variable



Einlesen gemischter Daten. Beachten Sie, dass beim Einlesen von Feldern mit numerischem Speichertyp (ganze Zahl, reelle Zahl, Zeit, Zeitstempel oder Datum) alle nicht numerischen Werte auf *null* oder auf *systemdefiniert fehlend* gesetzt werden. Dies liegt daran, dass im Gegensatz zu einigen anderen Anwendungen hier keine gemischten Speichertypen innerhalb eines Felds zulässig sind. Um dies zu vermeiden, sollten alle Felder mit gemischten Daten als Zeichenketten eingelesen werden, indem der Speichertyp im Quellenknoten bzw. in der externen Anwendung nach Bedarf geändert wird. Für weitere Informationen siehe Thema [Festlegen von Feldspeichertyp und Formatierung in Kapitel 2 in IBM SPSS Modeler 15 Quellen-, Prozess- und Ausgabeknoten](#).

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

Umgang mit fehlenden Werten

Entscheiden Sie vor dem Hintergrund Ihres Geschäfts- oder Domänenwissens, wie fehlende Werte behandelt werden sollen. Um die Trainingszeit zu verkürzen und die Genauigkeit zu erhöhen, können Sie beispielsweise einerseits Leerzeichen aus Ihrem Daten-Set entfernen. Andererseits kann das Vorhandensein leerer Werte neue Geschäftschancen eröffnen oder für zusätzliche

Einblicke sorgen. Bei der Auswahl des besten Verfahrens sollten Sie folgende Aspekte Ihrer Daten berücksichtigen:

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

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

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

Beide Ansätze lassen sich mithilfe des Data Audit-Knotens weitgehend automatisieren. Beispielsweise können Sie einen Filterknoten generieren, der Felder ausschließt, die so viele fehlende Werte aufweisen, dass sie bei der Modellierung nicht mehr sinnvoll einsetzbar sind, und einen Superknoten generieren, der fehlende Werte für bestimmte oder alle verbleibenden Felder imputiert. Hier zeigt sich die wahre Stärke des Audit: Sie können nicht nur den aktuellen Status Ihrer Daten einschätzen, sondern auch aufgrund dieser Einschätzung aktiv werden. [Für weitere Informationen siehe Thema Vorbereiten von Daten für die Analyse \(Data Audit\) in Kapitel 7 in IBM SPSS Modeler 15- Benutzerhandbuch.](#)

Behandeln von Datensätzen mit fehlenden Werten

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

- Sie können die Personaldatensätze mithilfe eines Auswahlknotens entfernen.
- Bei einem großen Daten-Set können Sie alle Datensätze mit Leerstellen verwerten. [Für weitere Informationen siehe Thema Auswählen von Datensätzen mit fehlenden Daten in Kapitel 6 in IBM SPSS Modeler 15 Quellen-, Prozess- und Ausgabeknoten.](#)

Behandeln von Feldern mit fehlenden Werten

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

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

Feld-Messniveau

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

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

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

Sichten bzw. Entfernen von Feldern

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

- Mit dem Data Audit-Knoten können Sie Felder nach Qualität filtern. [Für weitere Informationen siehe Thema Filtern von Feldern mit fehlenden Daten in Kapitel 6 in IBM SPSS Modeler 15 Quellen-, Prozess- und Ausgabeknoten.](#)
- Mithilfe eines Merkmalsauswahlknotens, können Sie Felder ausschließen, die mehr als einen angegebenen Prozentsatz fehlender Werte enthalten und eine Rangordnung der Felder auf der Grundlage der Wichtigkeit in Bezug auf ein angegebenes Ziel erstellen. [Für weitere Informationen siehe Thema Merkmalsauswahlknoten in Kapitel 4 in IBM SPSS Modeler 15 Modellierungsknoten.](#)
- Statt die Felder zu entfernen, können Sie die Feldrolle anhand eines Typknotens auf Keine festlegen. Dadurch werden die Felder im Daten-Set belassen, aber aus den Modellierungsprozessen ausgeschlossen.

Imputieren oder Auffüllen fehlender Werte

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

Folgende Methoden stehen zum Imputieren fehlender Werte zur Verfügung:

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

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

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

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

Für weitere Informationen siehe Thema *Fehlende Werte imputieren* in Kapitel 6 in *IBM SPSS Modeler 15 Quellen-, Prozess- und Ausgabeknoten*.

Um alternativ Werte für bestimmte Felder zu erzwingen, können Sie mithilfe eines Typknotens sicherstellen, dass die Feldtypen nur zulässige Werte abdecken, und dann die Spalte *Überprüfen* für die Felder, deren leere Werte ersetzt werden müssen, auf Erzwingen setzen. Für weitere Informationen siehe Thema *Typknoten* in Kapitel 4 in *IBM SPSS Modeler 15 Quellen-, Prozess- und Ausgabeknoten*.

CLEM-Funktionen für fehlende Werte

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

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

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

Sie können Nullen folgendermaßen in einer Liste von Feldern zählen:

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

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

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

Abbildung 6-2

Verwenden eines Füllerknotens, um im ausgewählten Feld leere Werte durch 0 zu ersetzen



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

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

Mit dieser Anweisung werden alle Werte ersetzt, die nicht mit einem systemdefiniert fehlenden Wert, der als `$null$` angezeigt wird, in einem Bereich liegen. Mit der Funktion `not()` können Sie alle anderen numerischen Werte, einschließlich negativer Werte, erfassen. Für weitere Informationen siehe Thema [Funktionen zum Umgang mit Leerstellen und Nullwerten](#) in Kapitel 8 auf S. 176.

Hinweis zum Verwerfen von Datensätzen

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

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

In früheren Versionen wurden Nullwerte in solchen Fällen ausgeschlossen.

Erstellen von CLEM-Ausdrücken

Informationen zu CLEM

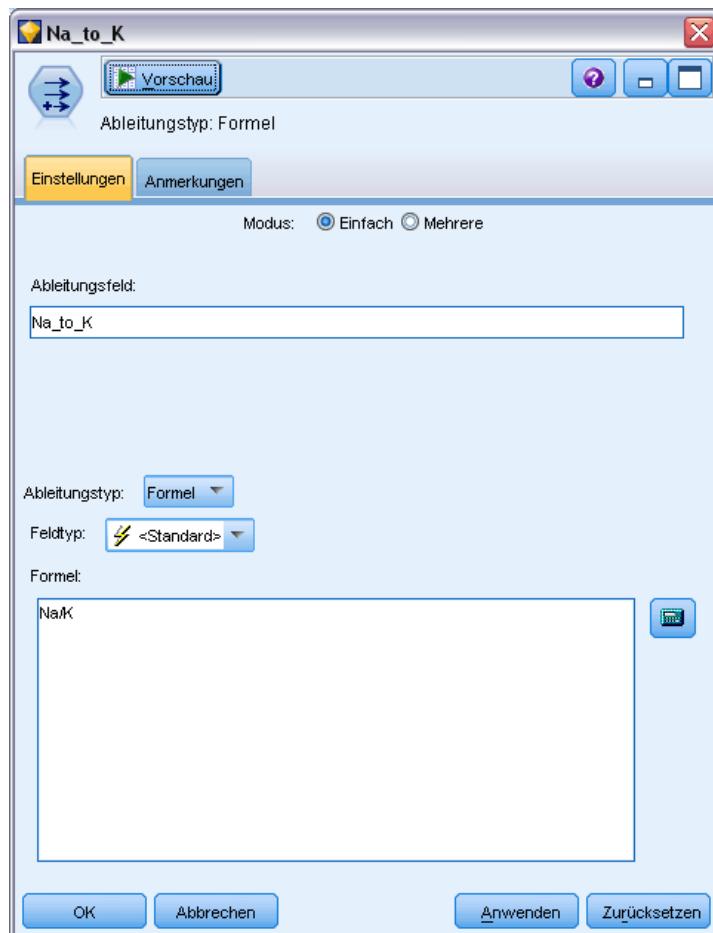
CLEM (Control Language for Expression Manipulation) ist eine leistungsfähige Sprache zur Analyse und Bearbeitung von Daten, die durch IBM® SPSS® Modeler-Streams fließen. Data-Mining-Experten verwenden CLEM in hohem Maße für Stream-Operationen, um so einfache Aufgaben durchzuführen, wie das Ableiten von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

CLEM wird innerhalb von SPSS Modeler für folgende Zwecke eingesetzt:

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

Skript. Eine Untergruppe der CLEM-Sprache kann außerdem in Skripts in der Benutzeroberfläche verwendet werden. Dadurch können Sie viele gleichartige Datenbearbeitungen automatisieren. Für weitere Informationen siehe Thema *Skripterstellung – Überblick* in Kapitel 2 in *IBM SPSS Modeler 15 Handbuch für Skripterstellung und Automatisierung*.

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

Abbildung 7-1*Ableiten eines Knotens durch Erstellen eines neuen Felds, das auf einer Formel basiert*

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

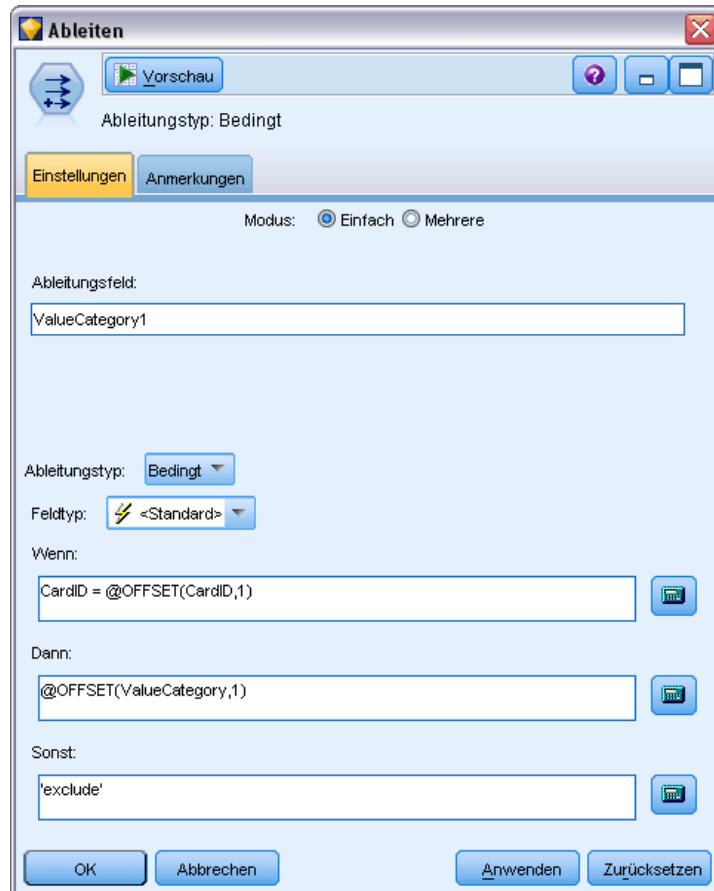
Abbildung 7-2
Füllerknoten ersetzt systemdefinierte fehlende Werte durch 0



Es können auch komplexere CLEM-Ausdrücke erstellt werden. Beispielsweise können Sie neue Felder auf der Grundlage einer bedingten Regelmenge ableiten.

Abbildung 7-3

Bedigte Ableitung zum Vergleich der Werte eines Felds mit denen des Felds davor



CLEM-Beispiele

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

Einfache Ausdrücke

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

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

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

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

```
log(Gehalt)
```

Komplexe Ausdrücke

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

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

Einige Funktionen, wie Zeichenkettenfunktionen, erfordern die Eingabe mehrerer Parameter in der korrekten Syntax. Die Funktion *subscrs* in folgendem Beispiel gibt das erste Zeichen eines Felds *produce_ID* zurück und gibt an, ob ein Element organisch, genetisch verändert oder herkömmlich ist. Die Ergebnisse eines Ausdrucks werden durch -> `result` beschrieben.

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

Ebenso lautet der folgende Ausdruck:

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

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

Kombinieren von Funktionen in einem Ausdruck

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

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

Der gleiche Ausdruck kann in Kurzform wie folgt geschrieben werden:

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

Eine andere häufig verwendete Kombination von Funktionen lautet:

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

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

Sonderfunktionen

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

@BLANK('referrer ID') -> T

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

@BLANK(@FIELD)-> T

Weitere Beispiele werden in der CLEM-Dokumentation erörtert. Für weitere Informationen siehe [Thema CLEM-Referenz – Überblick in Kapitel 8 auf S. 141](#).

Werte und Datentypen

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

3
1.79
'banana'

Feldnamen sind zum Beispiel:

Product_ID
'\$P-NextField'

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

- Zeichenketten, z. B. "c1", "Type 2", "a piece of free text"
- Ganze Zahlen, z. B. 12, 0, -189
- Reelle Zahlen, z. B. 12.34, 0.0, -0.0045
- Datums-/Zeitfelder, z. B. 05/12/2002, 12/05/2002, 12/05/02

Es können auch folgende Elemente verwendet werden:

- Zeichencodes, z. B. `a` or 3
- Liste mit Elementen, z. B. [1 2 3], ['Type 1' 'Type 2']

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

Regeln für Anführungszeichen

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

- **Zeichenketten** – Setzen Sie Zeichenketten immer in doppelte Anführungszeichen ("Type 2" oder "value"). Bei einfachen Anführungszeichen besteht die Gefahr, dass sie mit in Anführungszeichen gesetzten Feldern verwechselt werden.
- **Zeichen** – Verwenden Sie immer einfache umgekehrte Anführungszeichen wie z. B. ` . Beachten Sie beispielsweise das Zeichen d in der Funktion `stripchar(`d`,"drugA")`. Die einzige Ausnahme zu dieser Regel ist die Verwendung einer Ganzzahl für den Verweis auf ein bestimmtes Zeichen in einer Zeichenkette. Beachten Sie beispielsweise das Zeichen 5 in der Funktion `lowertoupper("drugA"(5)) —> "A"`. *Hinweis:* Auf einer englischen und US-amerikanischen Standardtastatur befindet sich die Taste für das umgekehrte Anführungszeichen (Accent grave, Unicode 0060) unterhalb der Esc-Taste, auf europäischen Standardtastaturen unterhalb der F10-Taste.
- **Felder** – Felder werden in CLEM-Ausdrücken in der Regel nicht in Anführungszeichen gesetzt (`subscr(2,arrayID)`) —> CHAR). Wenn Leerzeichen oder andere Sonderzeichen eingeschlossen werden müssen, verwenden Sie einfache Anführungszeichen ('Order Number'). Felder, die in Anführungszeichen gesetzt sind, im Daten-Set jedoch nicht definiert sind, werden irrtümlich als Zeichenketten interpretiert.
- **Parameter** – Verwenden Sie immer einfache Anführungszeichen ('\$P-threshold').

Ausdrücke und Bedingungen

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

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

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

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

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

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

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

- `abs(Signal)` wird ausgewertet und `sqrt` wird auf dessen Ergebnis angewendet.
- `max(T1, T2)` wird ausgewertet.

- Beide Ergebnisse werden multipliziert: \times hat Vorrang vor $+$.
- Schließlich wird **Baseline** zum Ergebnis addiert.

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

- Funktionsargumente
- Funktionsaufrufe
- $\times \times$
- $x / \text{mod} \text{ div} \text{ rem}$
- $+ -$
- $> < >= <= / == == = /=$

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

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

Stream-, Sitzungs- und Superknoten-Parameter

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

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

- Stream-Parameter können in einem Stream-Skript oder im Dialogfeld “Stream-Eigenschaften” festgelegt werden und sie stehen für alle Knoten im Stream zur Verfügung. Sie werden in der Parameterliste im Expression Builder angezeigt.
- Sitzungsparameter können in einem Standalone-Skript oder im Dialogfeld “Sitzungsparameter” festgelegt werden. Sie stehen allen Streams zur Verfügung, die in der aktuellen Sitzung verwendet werden (alle Streams auf der Registerkarte “Streams” im Manager-Bereich).

Parameter können auch für Superknoten festgelegt werden. In diesem Fall sind sie nur für Knoten sichtbar, die in dem betreffenden Superknoten verkapselt sind. Für weitere Informationen siehe Thema Festlegen von Superknoten-Parametern in Kapitel 9 in *IBM SPSS Modeler 15 Quellen-, Prozess- und Ausgabeknoten*.

Verwenden von Parametern in CLEM-Ausdrücken

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

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

- ▶ Klicken Sie in einem beliebigen Dialogfeld, in das CLEM-Ausdrücke eingegeben werden können, auf die Schaltfläche des Expression Builder.
- ▶ Wählen Sie in der Liste “Felder” die Option Parameter.

Sie können in der Liste Parameter auswählen, um sie in den CLEM-Ausdruck einzufügen. Für weitere Informationen siehe [Thema Auswählen von Feldern, Parametern und globalen Variablen auf S. 135](#).

Arbeiten mit Zeichenketten

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

- Konvertieren einer Zeichenkette zu Groß- oder Kleinbuchstaben – uppertolower(CHAR)
- Entfernen bestimmter Zeichen, wie z. B. `ID_` oder `\$', aus einer Zeichenkettenvariablen – stripchar(CHAR,STRING)
- Bestimmen der Länge (Zeichenanzahl) einer Zeichenkettenvariablen – length(STRING).
- Überprüfen der alphabetischen Sortierung von Zeichenketten – alphabefore(STRING1, STRING2)
- Entfernen führender oder nachgestellter Leerzeichen aus Werten – trim(STRING), trim_start(STRING) oder trimend(STRING)
- Extrahieren der ersten oder letzten *n* Zeichen aus einer Zeichenkette – startstring(LENGTH, STRING) oder endstring(LENGTH, STRING). Angenommen, es liegt beispielsweise ein Feld mit der Bezeichnung *Element* vor, in dem ein Produktnamen und ein 4-stelliger ID-Code kombiniert sind (ACME CAMERA-D109). Geben Sie die folgende Formel in einem Ableitungsknoten an, um ein neues Feld zu erstellen, das nur den 4-stelligen Code enthält:

endstring(4, Element)

- Abgleichen eines bestimmten Musters – STRING matches PATTERN. Um beispielsweise Personen mit dem Wort “market” an einer beliebigen Stelle in ihrer Stellenbezeichnung auszuwählen, können Sie Folgendes in einem Auswahlknoten angeben:

job_title matches “*market*”

- Ersetzen aller Instanzen einer Teilzeichenkette innerhalb einer Zeichenkette – replace(SUBSTRING, NEWSUBSTRING, STRING). Um vor dem Text-Mining zum Beispiel alle Instanzen eines nicht unterstützten Zeichens (etwa einen senkrechten Strich [|]) mit einem Strichpunkt zu ersetzen, verwenden Sie in einem Füllerknoten die Funktion replace. Wählen Sie unter Felder ausfüllen alle Felder aus, in denen das Zeichen möglicherweise vorkommt.

Wählen Sie für die Bedingung Ersetzen: die Option Immer aus und geben Sie unter Ersetzen mit folgende Bedingung an:

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

- Ableiten eines Flag-Felds basierend auf Vorhandensein einer bestimmten Teilzeichenkette. Sie können beispielsweise eine Zeichenkettenfunktion in einem Ableitungsknoten verwenden, um ein gesondertes Flag-Feld für jede Antwort mit einem Ausdruck der folgenden Art zu erstellen:

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

Für weitere Informationen siehe Thema String-Funktionen in Kapitel 8 auf S. 159.

Umgang mit Leerzeichen und fehlenden Werten

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

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

Abbildung 7-4

Füllerknoten ersetzt systemdefiniert fehlende Werte durch 0



Für weitere Informationen siehe Thema Funktionen zum Umgang mit Leerstellen und Nullwerten in Kapitel 8 auf S. 176.

Arbeiten mit Zahlen

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

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

Für weitere Informationen siehe Thema Numerische Funktionen in Kapitel 8 auf S. 154.

Arbeiten mit Zeit- und Datumsangaben

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

Berechnen der verstrichenen Zeit

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

`date_in_months(Date)`

Vergleichen von Datums-/Uhrzeitwerten

Die Werte in Datums- und Uhrzeitfeldern können in den Datensätzen anhand von Funktionen wie der folgenden verglichen werden. Diese Funktion gibt den Wert *Wahr* zurück, wenn die Datumszeichenkette DATE1 ein Datum darstellt, das vor dem durch die Datumszeichenkette DATE2 repräsentierten Datum liegt. Andernfalls ergibt die Funktion den Wert 0.

`date_before(Datum1, Datum2)`

Berechnen von Differenzen

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

`date_weeks_difference(Datum1, Datum2)`

Diese Funktion gibt die Zeit in Wochen ab dem durch die Datumszeichenkette DATE1 repräsentierten Datum bis zu dem durch die Datumszeichenkette DATE2 angegebenen Datum als reelle Zahl zurück. Der Wert basiert auf einer Woche mit 7,0 Tagen. Wenn DATE2 vor DATE1 liegt, wird eine negative Zahl zurückgegeben.

Heutiges Datum

Das aktuelle Datum kann mit der Funktion @TODAY zum Daten-Set hinzugefügt werden. Das heutige Datum wird als Zeichenkette zum angegebenen oder neuen Feld im Datumsformat hinzugefügt, das im Dialogfeld “Stream-Eigenschaften” ausgewählt wurde. Für weitere Informationen siehe Thema [Datums- und Uhrzeitfunktionen in Kapitel 8 auf S. 165](#).

Zusammenfassen mehrerer Felder

Die CLEM-Sprache beinhaltet eine Reihe von Funktionen, mit denen Übersichtsstatistiken ausgegeben werden, die mehrere Felder umfassen. Diese Funktionen sind besonders nützlich bei der Analyse von Umfragedaten, bei denen mehrere Antworten auf eine Frage in mehreren Feldern gespeichert werden. Für weitere Informationen siehe Thema [Arbeiten mit Mehrfachantwortdaten auf S. 130](#).

Vergleichsfunktionen

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

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

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

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

So können Sie Nullwerte in derselben Gruppe von Feldern zählen:

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

Beachten Sie, dass in diesem Beispiel die Anzahl der Karten im Besitz gezählt werden, nicht die Anzahl der Personen, die sie besitzen. Für weitere Informationen siehe Thema [Vergleichsfunktionen in Kapitel 8 auf S. 151](#).

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

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

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

Answer1	Answer2	Answer3	Anzahl
J	X	J	2
J	X	X	1

Numerische Funktionen

Mit den Funktionen `sum_n`, `mean_n` und `sdev_n` können Sie Statistiken für mehrere Felder erhalten. Beispiel:

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

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

Für weitere Informationen siehe Thema Numerische Funktionen in Kapitel 8 auf S. 154.

Generieren von Feldlisten

Wenn Sie eine der Funktionen verwenden, die eine Liste von Feldern als Eingabe akzeptieren, können die Sonderfunktionen `@FIELDS_BETWEEN(start, end)` und `@FIELDS_MATCHING(pattern)` als Eingabe verwendet werden. Nehmen wir beispielsweise an, dass die Reihenfolge der Felder der im Beispiel für `sum_n` oben entspricht. Folgendes wäre äquivalent:

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

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

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

Für weitere Informationen siehe Thema Sonderfelder in Kapitel 8 auf S. 177.

Arbeiten mit Mehrfachantwortdaten

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

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

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

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

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

Kunden-ID	Auto1	Auto2	Auto3
101	1	3	2
102	3	2	1
103	2	3	1

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

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

Für weitere Informationen siehe Thema [Vergleichsfunktionen in Kapitel 8 auf S. 151.](#)

Referenzieren von Mehrfachantworten-Sets

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

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

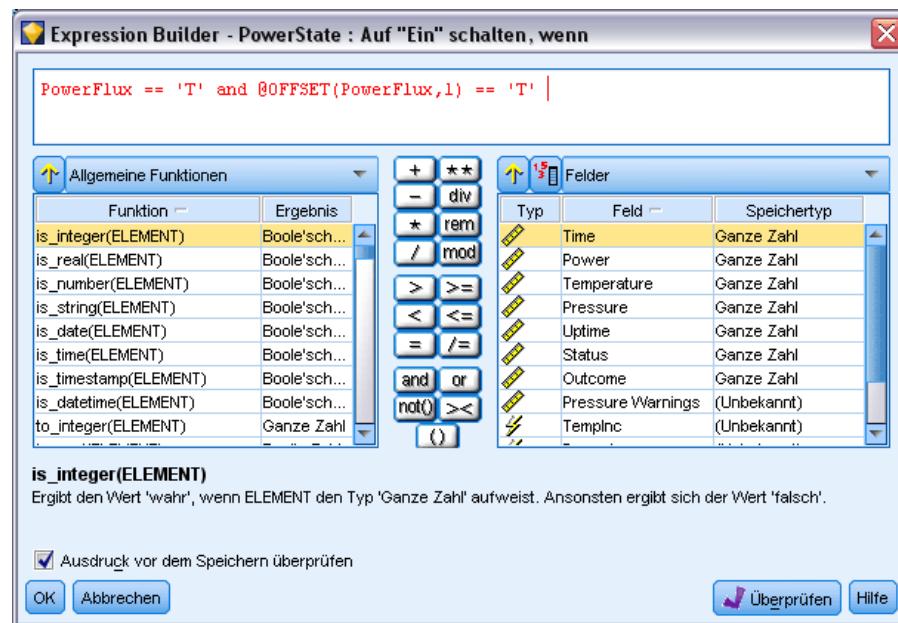
Für weitere Informationen siehe Thema [Bearbeiten von Mehrfachantworten-Sets in Kapitel 4 in IBM SPSS Modeler 15 Quellen-, Prozess- und Ausgabeknoten.](#)

Expression Builder

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

automatisch die entsprechenden Anführungszeichen für Felder und Werte hinzufügt, sodass es einfacher wird, syntaktisch korrekte Ausdrücke zu erstellen.

Abbildung 7-5
Dialogfeld "Expression Builder"

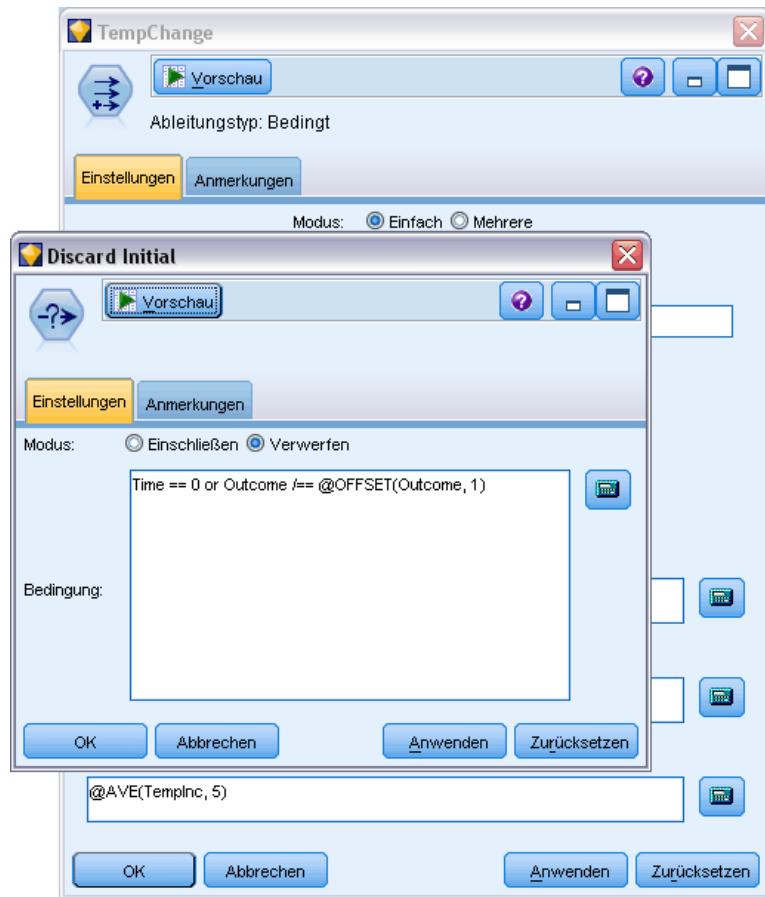


Hinweis: Expression Builder wird in Skript- oder Parametereinstellungen nicht unterstützt.

Zugriff auf Expression Builder

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

Abbildung 7-6
Diverse Knoten mit der Schaltfläche "Expression Builder"



Erstellen von Ausdrücken

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

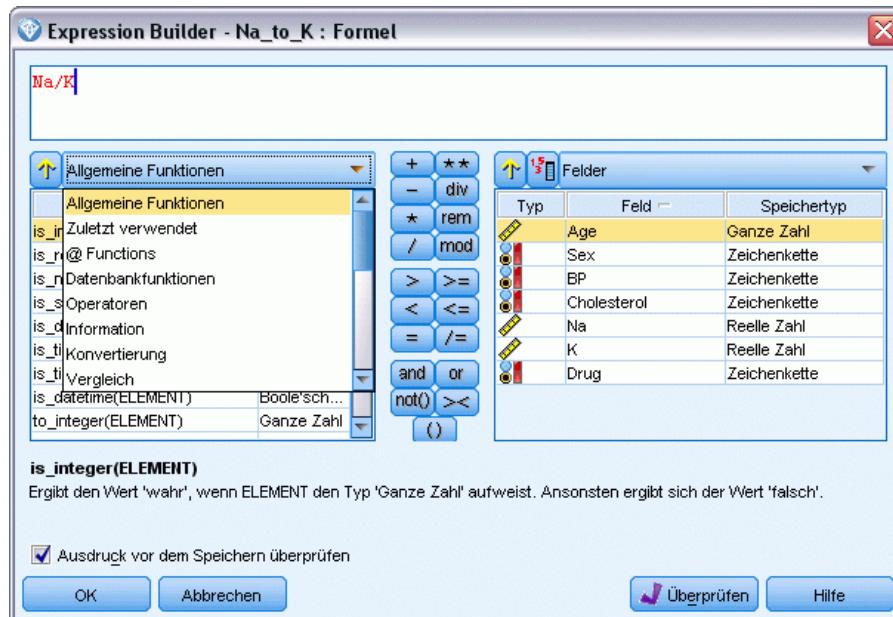
So erstellen Sie einen Ausdruck mit Expression Builder:

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

Auswählen von Funktionen

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

Abbildung 7-7
Dropdownliste "Funktionen"



Die meisten dieser Kategorien sind im Referenzabschnitt der CLEM-Sprachbeschreibung zu finden. Für weitere Informationen siehe Thema **Funktionsreferenz** in Kapitel 8 auf S. 148.

Die anderen Kategorien lauten wie folgt.

- **Allgemeine Funktionen** enthält eine Auswahl von einigen der am häufigsten verwendeten Funktionen.

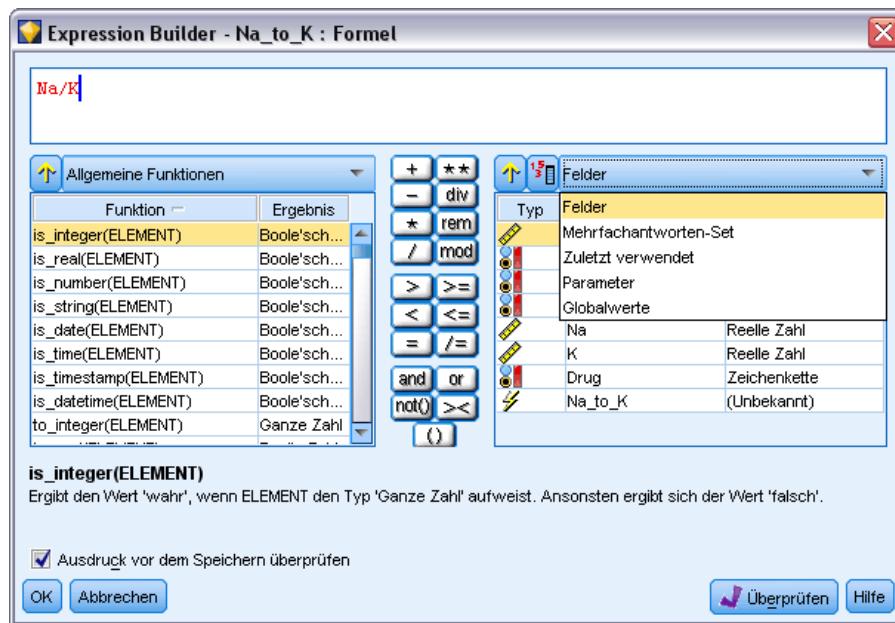
- **Zuletzt verwendet** – Enthält eine Liste der CLEM-Funktionen, die in der aktuellen Sitzung verwendet wurden.
- **@ Functions** enthält eine Liste aller Sonderfunktionen, deren Namen das Zeichen “@” vorangestellt ist.
- **Datenbankfunktionen**. Wenn der Stream eine Datenbankverbindung beinhaltet (mithilfe eines Datenbankquellenknotens) listet diese Auswahl die Funktionen auf, die in dieser Datenbank zur Verfügung stehen, einschließlich benutzerdefinierter Funktionen (UDFs).
- **Operatoren** listet alle Operatoren auf, die Sie beim Erstellen von Ausdrücken verwenden können. Operatoren stehen auch über die Schaltflächen in der Mitte des Dialogfelds zur Verfügung.
- **Alle Funktionen** – Enthält eine vollständige Liste der verfügbaren CLEM-Funktionen.

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

Auswählen von Feldern, Parametern und globalen Variablen

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

Abbildung 7-8
Expression Builder: Feldliste



Für weitere Informationen siehe Thema Stream-, Sitzungs- und Superknoten-Parameter auf S. 125.

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

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

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

Parameter. Für weitere Informationen siehe Thema Stream-, Sitzungs- und Superknoten-Parameter auf S. 125.

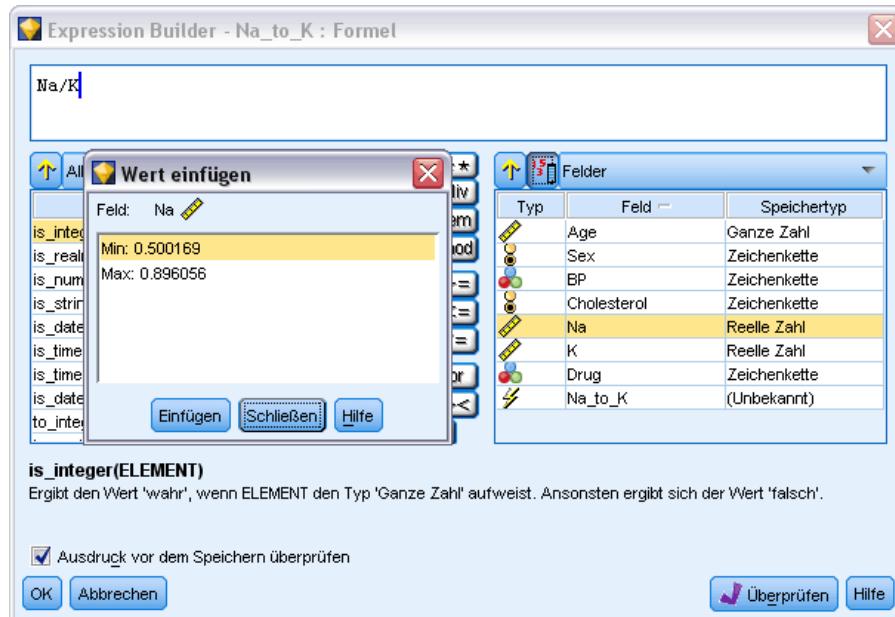
Globalwerte. Weitere Informationen finden Sie im *IBM SPSS Modeler-Handbuch zu Quellenknoten, Prozessknoten und Ausgabeknoten*.

Anzeigen oder Auswählen von Werten

Feldwerte können an verschiedenen Stellen im System angezeigt werden. Dazu gehören Expression Builder, Data Audit-Berichte und beim Bearbeiten zukünftiger Werte in einem Zeitintervallknoten. Beachten Sie, dass die Daten vollständig in einem Quellen- oder Typknoten instanziert sein müssen (Speichertypen, Typen und Werte sind bekannt), um diese Funktion verwenden zu können. Für weitere Informationen siehe Thema Verwenden des Dialogfelds "Werte" in Kapitel 4 in *IBM SPSS Modeler 15 Quellen-, Prozess- und Ausgabeknoten*.

Abbildung 7-9

Feldliste mit den für das ausgewählte Feld angezeigten Werten



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

Abbildung 7-10
Schaltfläche zur Wertauswahl

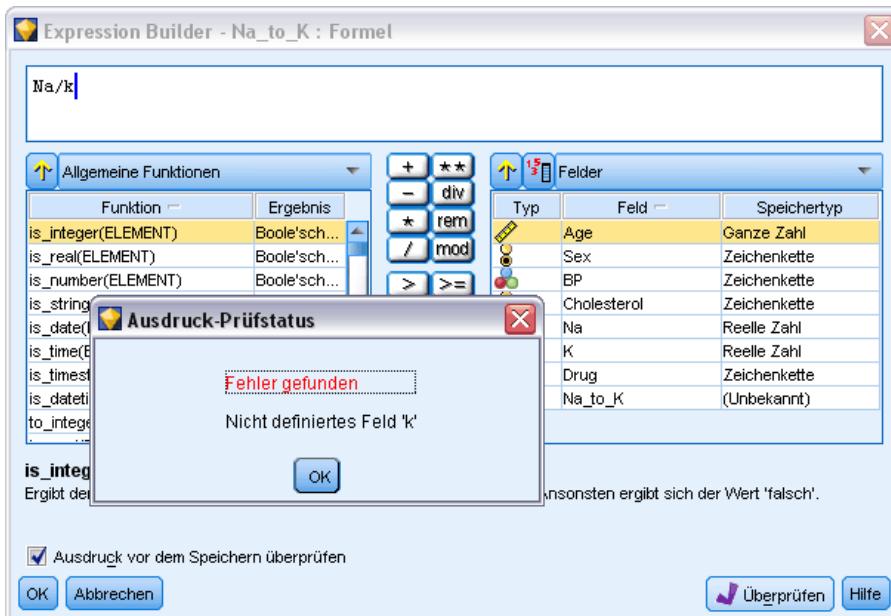


Bei Flag-Feldern und nominalen Feldern werden alle definierten Werte aufgeführt. Bei stetigen Feldern (numerischer Bereich) werden die Minimal- und Maximalwerte angezeigt.

Überprüfen von CLEM-Ausdrücken

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

Abbildung 7-11
Ungültiger CLEM-Ausdruck



Folgende Elemente werden überprüft:

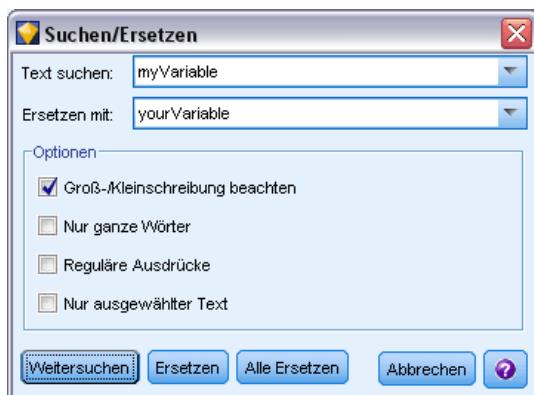
- Richtige Anführungszeichen für Werte und Feldnamen
- Richtige Verwendung von Parametern und globalen Variablen
- Zulässige Verwendung von Operatoren
- Vorhandensein von referenzierten Feldern
- Vorhandensein und Definition von referenzierten Globalwerten

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

Suchen und ersetzen

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

Abbildung 7-12
Dialogfeld “Suchen/Ersetzen”



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

Suchoptionen

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

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

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

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

Syntax für reguläre Ausdrücke

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

Zeichenübereinstimmungen

Zeichen	Übereinstimmung/Bedeutung
x	Das Zeichen x
\\"	Umgekehrter Schrägstrich
\0n	Das Zeichen mit dem Oktalwert 0n (0 <= n <= 7)
\0nn	Das Zeichen mit dem Oktalwert 0nn (0 <= n <= 7)
\0mnn	Das Zeichen mit dem Oktalwert 0mnn (0 <= m <= 3; 0 <= n <= 7)
\xhh	Das Zeichen mit dem Hexadezimalwert 0xhh
\uhhhh	Das Zeichen mit dem Hexadezimalwert 0uhhhh
\t	Das Tabulatorzeichen ('\u0009')
\n	Das Zeichen für Zeilenvorschub ('\u000A')
\r	Das Zeichen für Wagenrücklauf ('\u000D')
\f	Das Zeichen für Seitenvorschub ('\u000C')
\a	Das Alarmzeichen ('\u0007')
\e	Das Escape-Zeichen ('\u001B')
\cx	Das Steuerzeichen, das x entspricht

Übereinstimmende Zeichenklassen

Zeichenklassen	Übereinstimmung/Bedeutung
[abc]	a, b oder c (einfache Klasse)
[^abc]	Beliebiges Zeichen mit Ausnahme von a, b oder c (Differenzmenge)
[a-zA-Z]	a bis einschließlich z bzw. A bis einschließlich Z (Bereich)

Zeichenklassen	Übereinstimmung/Bedeutung
[a-d[m-p]]	a bis d bzw. m bis P (Vereinigungsmenge) Alternative Angabemöglichkeit: [a-dm-p]
[a-z&&[def]]	a bis z und d, e bzw. f (Schnittmenge)
[a-z&&[^bc]]	a bis z mit Ausnahme von b und c (Differenzmenge) Alternative Angabemöglichkeit: [ad-z]
[a-z&&[^m-p]]	a bis z mit Ausnahme von m bis p (Differenzmenge) Alternative Angabemöglichkeit: [a-lq-z]

Vordefinierte Zeichenklassen

Vordefinierte Zeichenklassen	Übereinstimmung/Bedeutung
.	Beliebiges Zeichen (evtl. Übereinstimmung mit Zeilenabschlusszeichen)
\d	Beliebige Ziffer: [0-9]
\D	Nichtziffer: [^0-9]
\s	Ein Leerzeichen: [\t\n\x0B\f\r]
\S	Ein Nicht-Leerzeichen: [^\s]
\w	Buchstabe: [a-zA-Z_0-9]
\W	Nichtbuchstabe: [^\w]

Übereinstimmungen mit Begrenzungszeichen

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

CLEM-Sprachreferenz

CLEM-Referenz – Überblick

In diesem Abschnitt wird CLEM (Control Language for Expression Manipulation) beschrieben, ein leistungsstarkes Tool zur Analyse und Bearbeitung der in IBM® SPSS® Modeler-Streams verwendeten Daten. CLEM kann in Knoten zur Ausführung verschiedener Aufgaben eingesetzt werden, vom Auswerten von Bedingungen über das Ableiten von Werten bis hin zum Einfügen von Daten in Berichte. [Für weitere Informationen siehe Thema Informationen zu CLEM in Kapitel 7 auf S. 118.](#)

Eine Untergruppe der CLEM-Sprache kann außerdem in Skripts in der Benutzeroberfläche verwendet werden. Dadurch können Sie viele gleichartige Datenbearbeitungen automatisieren. [Für weitere Informationen siehe Thema CLEM-Ausdrücke in Skripts in Kapitel 3 in IBM SPSS Modeler 15 Handbuch für Skripterstellung und Automatisierung.](#)

CLEM-Ausdrücke bestehen aus Werten, Feldnamen, Operatoren und Funktionen. Mit der richtigen Syntax können Sie eine Vielzahl leistungsstarker Datenoperationen erstellen. [Für weitere Informationen siehe Thema CLEM-Beispiele in Kapitel 7 auf S. 121.](#)

CLEM-Datentypen

CLEM-Datentypen können aus folgenden Elementen bestehen:

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

Regeln für die Verwendung von Anführungszeichen

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

- Strings – Setzen Sie Zeichenketten immer in doppelte Anführungszeichen, wie beispielsweise "Type 2". Bei einfachen Anführungszeichen besteht die Gefahr, dass sie mit in Anführungszeichen gesetzten Feldern verwechselt werden.

- Fields – Verwenden Sie einzelne Anführungszeichen, nur wenn es erforderlich ist, um Leerzeichen oder andere Sonderzeichen einzuschließen. Beispiel: 'Order Number'. Felder, die in Anführungszeichen gesetzt sind, im Daten-Set jedoch nicht definiert sind, werden irrtümlich als Zeichenketten interpretiert.
- Parameter – Setzen Sie Parameter immer in einfache Anführungszeichen, wie beispielsweise '\$P-threshold'.
- Zeichen – Zeichen müssen stets in einfache umgekehrte Anführungszeichen (`) gesetzt werden. Beispiel: stripchar(`d`, "drugA").

Für weitere Informationen siehe Thema Werte und Datentypen in Kapitel 7 auf S. 123. Diese Regeln werden außerdem in folgenden Themenabschnitten eingehend behandelt.

Ganze Zahlen

Ganze Zahlen werden als Folge von Dezimalziffern dargestellt. Optional können Sie der ganzen Zahl ein Minuszeichen (–) voranstellen, um eine negative Zahl anzugeben. Beispiele: 1234, 999, –77.

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

Reelle Zahlen

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

Optional können Sie der reellen Zahl ein Minuszeichen (–) voranstellen, um eine negative Zahl anzugeben. Beispiele: 1.234, 0.999, –77.001. Mit dem Format <Zahl> e <Exponent> können Sie eine reelle Zahl in Exponentialnotation angeben. Beispiel: 1234.0e5, 1.7e–2. Wenn die IBM® SPSS® Modeler-Anwendung numerische Zeichenketten aus Dateien einliest und sie automatisch in Zahlen umwandelt, werden Zahlen ohne führende Ziffer vor dem Dezimalpunkt oder ohne Ziffer nach dem Punkt akzeptiert. Beispiel: 999. oder .11. Bei CLEM-Ausdrücken sind diese Formate jedoch nicht zulässig.

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

Na > 0.6

Und nicht:

Na > 0,6

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

Zeichen

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

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

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

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

Zeichenketten

Im Allgemeinen sollten Zeichenketten in doppelte Anführungszeichen gesetzt werden. Beispiele: "c35product2" und "referrerID". Zur Kennzeichnung von Sonderzeichen in Zeichenketten werden umgekehrte Schrägstriche verwendet, z. B. "\\$65443". (Zur Kennzeichnung eines umgekehrten Schrägstrichs wird ein doppelter umgekehrter Schrägstrich verwendet: \\.) Zeichenketten können auch in einfache Anführungszeichen gesetzt werden. Das Ergebnis lässt sich jedoch nicht von einem in Anführungszeichen gesetzten Feld unterscheiden ('referrerID'). [Für weitere Informationen siehe Thema String-Funktionen auf S. 159.](#)

Listen

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

Felder

Namen in CLEM-Ausdrücken, bei denen es sich nicht um Namen von Funktionen handelt, werden als Feldnamen betrachtet. Diese können einfach als **Power**, **val27**, **state_flag** usw. geschrieben werden, wenn der Name jedoch mit einer Ziffer beginnt oder nicht alphabetische Zeichen wie Leerzeichen enthält (mit Ausnahme des Unterstrichs), setzen Sie den Namen in einfache Anführungszeichen. Beispiel: 'Power Increase', '2nd answer', '#101', '\$P-NextField'.

Hinweis: Felder, die in Anführungszeichen gesetzt sind, im Daten-Set jedoch nicht definiert sind, werden irrtümlich als Zeichenketten interpretiert.

Datumsangaben

Die Datumsberechnungen beruhen auf einem “Basisdatum”, das im Dialogfeld “Stream-Eigenschaften” angegeben wird. Standardmäßig wird als Basisdatum der 1. Januar 1900 verwendet. Für weitere Informationen siehe [Thema Festlegen von allgemeinen Optionen für Streams in Kapitel 5 auf S. 63.](#)

Die CLEM-Sprache unterstützt folgende Datumsformate.

Format	Beispiele
DDMMYY	150163
MMDDYY	011563
YYMMDD	630115
YYYYMMDD	19630115
YYYYDDD	Vierstellige Jahreszahl, gefolgt von einer dreistelligen Zahl für den Tag im Jahr – so steht 2000032 für den 32. Tag des Jahres 2000, also für den 1. Februar 2000.
DAY	Tag der Woche in der aktuellen Ländereinstellung – z. B. Monday, Tuesday usw. für Deutsch.
MONTH	Monat in der aktuellen Ländereinstellung – z. B. January, February,
DD/MM/YY	15/01/63
DD/MM/YYYY	15/01/1963
MM/DD/YY	01/15/63
MM/DD/YYYY	01/15/1963
DD-MM-YY	15-01-63
DD-MM-YYYY	15-01-1963
MM-DD-YY	01-15-63
MM-DD-YYYY	01-15-1963
DD.MM.YY	15.01.63
DD.MM.YYYY	15.01.1963
MM.DD.YY	01.15.63
MM.DD.YYYY	01.15.1963
DD-MON-YY	15-JAN-63, 15-jan-63, 15-Jan-63
DD/MON/YY	15/JAN/63, 15/jan/63, 15/Jan/63
DD.MON.YY	15.JAN.63, 15.jan.63, 15.Jan.63
DD-MON-YYYY	15-JAN-1963, 15-jan-1963, 15-Jan-1963
DD/MON/YYYY	15/JAN/1963, 15/jan/1963, 15/Jan/1963
DD.MON.YYYY	15.JAN.1963, 15.jan.1963, 15.Jan.1963
MON YYYY	Jan 2004

Format	Beispiele
q Q YYYY	Datum als Ziffer (1–4) für das Quartal, gefolgt vom Buchstaben <i>Q</i> und einer vierstelligen Jahresangabe. So wird beispielsweise das Datum 25. Dezember 2004 als 4 Q 2004 dargestellt.
ww WK YYYY	Zweistellige Zahl für die Kalenderwoche, gefolgt von den Buchstaben <i>WK</i> und dann von einer vierstelligen Jahreszahl. Bei der Berechnung der Kalenderwoche wird davon ausgegangen, dass Montag der erste Wochentag ist, und dass mindestens ein Tag in der ersten Kalenderwoche liegt.

Zeit

Die CLEM-Sprache unterstützt folgende Zeitformate.

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

CLEM Operatoren

Folgende Operatoren stehen zur Verfügung.

Operation	Kommentare	Rangfolge (siehe nächsten Abschnitt)
or	Wird zwischen zwei CLEM-Ausdrücken verwendet. Ergibt den Wert "Wahr", wenn mindestens einer der beiden Ausdrücke wahr ist.	10
and	Wird zwischen zwei CLEM-Ausdrücken verwendet.	9

Operation	Kommentare	Rangfolge (siehe nächsten Abschnitt)
=	Ergibt den Wert "Wahr", wenn beide Ausdrücke "wahr" sind.	
==	Wird zwischen zwei beliebigen miteinander vergleichbaren Elementen verwendet. Ergibt den Wert "Wahr", wenn ELEMENT1 gleich ELEMENT2 ist.	7
!=	Identisch mit =.	7
/=	Wird zwischen zwei beliebigen miteinander vergleichbaren Elementen verwendet. Ergibt den Wert "Wahr", wenn ELEMENT1 <i>nicht</i> gleich ELEMENT2 ist.	7
/==	Identisch mit /=.	7
>	Wird zwischen zwei beliebigen miteinander vergleichbaren Elementen verwendet. Ergibt den Wert "Wahr", wenn ELEMENT1 größer als ELEMENT2 ist.	6
>=	Wird zwischen zwei beliebigen miteinander vergleichbaren Elementen verwendet. Ergibt den Wert "Wahr", wenn ELEMENT1 größer oder gleich ELEMENT2 ist.	6
<	Wird zwischen zwei beliebigen miteinander vergleichbaren Elementen verwendet. Ergibt den Wert "Wahr", wenn ELEMENT1 kleiner als ELEMENT2 ist.	6
<=	Wird zwischen zwei beliebigen miteinander vergleichbaren Elementen verwendet. Ergibt den Wert "Wahr", wenn ELEMENT1 kleiner oder gleich ELEMENT2 ist.	6
&&=_0	Wird zwischen zwei ganzen Zahlen verwendet. Entspricht dem Boole'schen Ausdruck GANZZ1 && GANZZ2 = 0.	6
&&/=_0	Wird zwischen zwei ganzen Zahlen verwendet. Entspricht dem Boole'schen Ausdruck GANZZ1 && GANZZ2 /= 0.	6
+	Addiert zwei Zahlen: ZAHL1 + ZAHL2.	5
><	Verkettet zwei Zeichenfolgen; beispielsweise STRING1 >< STRING2.	5
-	Subtrahiert eine Zahl von einer anderen Zahl: ZAHL1 - ZAHL2. Kann außerdem vor einer Zahl verwendet werden: -ZAHL.	5
*	Dient zur Multiplikation zweier Zahlen: ZAHL1 * ZAHL2.	4

Operation	Kommentare	Rangfolge (siehe nächsten Abschnitt)
<code>&&</code>	Wird zwischen zwei ganzen Zahlen verwendet. Das Ergebnis liefert das bitweise "Und" der ganzen Zahlen GANZZ1 und GANZZ2.	4
<code>&&~</code>	Wird zwischen zwei ganzen Zahlen verwendet. Das Ergebnis liefert das bitweise "Und" von GANZZ1 und das bitweise Komplement von GANZZ2.	4
<code> </code>	Wird zwischen zwei ganzen Zahlen verwendet. Das Ergebnis liefert das bitweise "Oder" von GANZZ1 und GANZZ2.	4
<code>~~</code>	Wird vor einer ganzen Zahl verwendet. Liefert das bitweise Komplement von GANZZ.	4
<code> /&</code>	Wird zwischen zwei ganzen Zahlen verwendet. Das Ergebnis liefert das bitweise "exklusive Oder" von GANZZ1 und GANZZ2.	4
<code>INT1 << N</code>	Wird zwischen zwei ganzen Zahlen verwendet. Ergibt das Bitmuster von GANZZ, um N Positionen nach links verschoben.	4
<code>INT1 >> N</code>	Wird zwischen zwei ganzen Zahlen verwendet. Ergibt das Bitmuster von GANZZ, um N Positionen nach rechts verschoben.	4
<code>/</code>	Dient zur Division zwischen zwei Zahlen: ZAHL1 / ZAHL2.	4
<code>**</code>	Wird zwischen zwei Zahlen verwendet. BASIS ** POTENZ. Ergibt BASIS hoch POTENZ.	3
<code>rem</code>	Wird zwischen zwei ganzen Zahlen verwendet: GANZZ1 rem GANZZ2. Ergibt den Rest: GANZZ1 - (GANZZ1 div GANZZ2) * GANZZ2.	2
<code>div</code>	Wird zwischen zwei ganzen Zahlen verwendet: GANZZ1 div GANZZ2. Führt eine Division zwischen zwei ganzen Zahlen durch.	2

Rangfolge der Operatoren

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

`3 + 4 * 5`

Analysiert als $3 + (4 * 5)$ statt als $(3 + 4) * 5$, da die relativen Rangfolgen vorgeben, dass $*$ vor $+$ analysiert wird. Jedem Operator in der CLEM-Sprache ist ein Rangfolgewert zugeordnet; je niedriger dieser Wert ist, desto wichtiger ist er auf der Analyseliste, was bedeutet, dass er vor anderen Operatoren mit höheren Rangfolgewerten verarbeitet wird.

Funktionsreferenz

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

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

Konventionen bei Funktionsbeschreibungen

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

Konvention	Beschreibung
<i>BOOL</i>	Ein Boole'scher Wert bzw. Flag, wie "wahr" oder "falsch".
<i>ZAHL, ZAHL1, ZAHL2</i>	Eine beliebige Zahl.
<i>REELL, REELL1, REELL2</i>	Eine reelle Zahl, wie 1.234 oder -77.01.
<i>GANZZ, GANZZ1, GANZZ2</i>	Eine ganze Zahl, wie 1 oder -77.
<i>CHAR</i>	Eine Zeichencode, wie beispielsweise 'A'.
<i>STRING</i>	Ein Zeichenkette, wie beispielsweise "referrerID".
<i>LISTE</i>	Eine Liste mit Elementen, wie beispielsweise ["abc" "def"].
<i>ELEMENT</i>	Ein Feld, wie beispielsweise Customer oder extract_concept.
<i>DATE</i>	Ein Datumsfeld, wie beispielsweise start_date, mit Werten in einem Format wie DD-MÖN-YYYY.
<i>ZEIT</i>	Ein Zeitfeld, wie beispielsweise power_flux, mit Werten in einem Format wie HHMMSS.

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

Funktion	Ergebnis	Beschreibung
INT1 rem INT2	Number	Ergibt den Rest des Quotienten aus <i>GANZZ1</i> und <i>GANZZ2</i> . Beispiel: INT1 – (INT1 div INT2) * INT2.

Einzelheiten zur den Verwendungskonventionen, beispielsweise zur Auflistung von Elementen oder zur Angabe von Zeichen in einer Funktion, werden an anderer Stelle beschrieben. Für weitere Informationen siehe Thema [CLEM-Datentypen auf S. 141](#).

Informationsfunktionen

Informationsfunktionen werden verwendet, um einen Einblick in die Werte eines bestimmten Felds zu gewinnen. Sie werden üblicherweise verwendet, um Flag-Felder abzuleiten. Beispiel: Mit der Funktion @BLANK können Sie ein Flag-Feld erstellen, das die Datensätze angibt, deren

Werte für das ausgewählte Feld leer sind. In ähnlicher Weise können Sie den Speichertyp für ein Feld mit einer der Speichertypfunktionen überprüfen, beispielsweise mit `is_string`.

Funktion	Ergebnis	Beschreibung
<code>@BLANK(FIELD)</code>	<i>Boolesch</i>	Ergibt den Wert "Wahr" für alle Datensätze, deren Werte gemäß den Regeln zum Umgang mit Leerstellen, die in einem weiter oben im Stream gelegenen Typknoten oder Quellenknoten (Registerkarte "Typen") festgelegt wurden, Leerstellen sind. Beachten Sie, dass diese Funktion nicht über ein Skript aufgerufen werden kann. Für weitere Informationen siehe Thema CLEM-Ausdrücke in Skripts in Kapitel 3 in IBM SPSS Modeler 15 Handbuch für Skripterstellung und Automatisierung .
<code>@NULL(ITEM)</code>	<i>Boolesch</i>	Ergibt den Wert "Wahr" für alle Datensätze, deren Werte nicht definiert sind. Nichtdefinierte Werte sind systemdefinierte Nullwerte, die in IBM® SPSS® Modeler als <code>\$null\$</code> angezeigt werden. Beachten Sie, dass diese Funktion nicht über ein Skript aufgerufen werden kann. Für weitere Informationen siehe Thema CLEM-Ausdrücke in Skripts in Kapitel 3 in IBM SPSS Modeler 15 Handbuch für Skripterstellung und Automatisierung .
<code>is_date(ITEM)</code>	<i>Boolesch</i>	Ergibt den Wert "Wahr" für alle Datensätze, deren Typ ein Datum ist.
<code>is_datetime(ITEM)</code>	<i>Boolesch</i>	Ergibt den Wert "Wahr" für alle Datensätze, deren Typ "Datum", "Zeit" oder "Zeitstempel" ist.
<code>is_integer(ITEM)</code>	<i>Boolesch</i>	Ergibt den Wert "Wahr" für alle Datensätze, deren Typ eine ganze Zahl ist.
<code>is_number(ITEM)</code>	<i>Boolesch</i>	Ergibt den Wert "Wahr" für alle Datensätze, deren Typ eine Zahl ist.
<code>is_real(ITEM)</code>	<i>Boolesch</i>	Ergibt den Wert "Wahr" für alle Datensätze, deren Typ eine reelle Zahl ist.
<code>is_string(ITEM)</code>	<i>Boolesch</i>	Ergibt den Wert "Wahr" für alle Datensätze, deren Typ eine Zeichenkette ist.
<code>is_time(ITEM)</code>	<i>Boolesch</i>	Ergibt den Wert "Wahr" für alle Datensätze, deren Typ eine Zeitangabe ist.
<code>is_timestamp(ITEM)</code>	<i>Boolesch</i>	Ergibt den Wert "Wahr" für alle Datensätze, deren Typ ein Zeitstempel ist.

Konvertierungsfunktionen

Mit Konvertierungsfunktionen können Sie neue Felder erstellen und den Speichertyp bestehender Felder umwandeln. Beispielsweise können Sie neue Zeichenketten bilden, indem Sie Zeichenketten miteinander verbinden oder zerlegen. Zwei Zeichenketten können mit dem Operator `><` miteinander verbunden werden. Wenn das Feld `Site` beispielsweise den Wert "BRAMLEY" hat, dann ergibt "xx" `><` `Site` als Ergebnis "xxBRAMLEY". Das Ergebnis von `><` ist immer eine Zeichenkette, selbst wenn die einzelnen Argumente keine Zeichenketten sind. Wenn das Feld `V1` den Wert 3 und das Feld `V2` den Wert 5 hat, dann gibt `V1 >< V2` folglich als Ergebnis "35" zurück (eine Zeichenkette, keine Zahl).

Konvertierungsfunktionen (und alle anderen Funktionen, für die ein spezieller Eingabetyp, wie beispielsweise ein Wert für Datum oder Uhrzeit, erforderlich ist) hängen von den aktuell im Dialogfeld für die Stream-Optionen angegebenen Formaten ab. Wenn Sie beispielsweise ein Zeichenkettenfeld mit den Werten *Jan 2003*, *Feb 2003* usw. konvertieren möchten, müssen Sie das zugehörige Datumsformat, MON JJJJ, als Standard-Datumsformat für den Stream auswählen. Für weitere Informationen siehe Thema Festlegen von allgemeinen Optionen für Streams in Kapitel 5 auf S. 63.

Funktion	Ergebnis	Beschreibung
ITEM1 >< ITEM2	Zeichenfolge	Zieht die Werte der beiden Felder zusammen und ergibt die resultierende Zeichenkette als ELEMENT1ELEMENT2.
to_integer(ITEM)	Ganzzahl	Konvertiert den Speichertyp des angegebenen Felds in eine ganze Zahl.
to_real(ITEM)	Reelle Zahl	Konvertiert den Speichertyp des angegebenen Felds in eine reelle Zahl.
to_number(ITEM)	Number	Konvertiert den Speichertyp des angegebenen Felds in eine Zahl.
to_string(ITEM)	Zeichenfolge	Konvertiert den Speichertyp des angegebenen Felds in eine Zeichenkette.
to_time(ITEM)	Zeit	Konvertiert den Speichertyp des angegebenen Felds in eine Zeit.
to_date(ITEM)	Datum	Konvertiert den Speichertyp des angegebenen Felds in ein Datum.
to_timestamp(ITEM)	Zeitstempel	Konvertiert den Speichertyp des angegebenen Felds in einen Zeitstempel.
to_datetime(ITEM)	Datetime	Konvertiert den Speichertyp des angegebenen Felds in einen Wert vom Typ "Datum", "Zeit" oder "Zeitstempel".
datetime_date(ITEM)	Datum	Gibt den Datumswert für eine <i>Zahl</i> , eine <i>Zeichenkette</i> oder einen <i>Zeitstempel</i> zurück. Beachten Sie, dass dies die einzige Funktion ist, mit der Sie eine Zahl (in wenigen Sekunden) wieder in ein Datum zurückkonvertieren können. Wenn es sich bei ITEM um eine Zeichenkette handelt, wird ein Datum erstellt, indem eine Zeichenkette im aktuellen Datumsformat analysiert wird. Das im Dialogfeld für die Stream-Eigenschaften angegebene Datumsformat muss korrekt sein, damit diese Funktion erfolgreich ausgeführt werden kann. Wenn es sich bei ITEM um eine Zahl handelt, wird sie als Anzahl von Sekunden seit dem Basisdatum (Epochen) interpretiert. Die Bruchteile eines Tages werden gekürzt. Wenn es sich bei ITEM um einen Zeitstempel handelt, wird der Datumsteil des Zeitstempels ausgegeben. Wenn es sich bei ITEM um ein Datum handelt, wird es unverändert ausgegeben.

Vergleichsfunktionen

Mit Vergleichsfunktionen werden Feldwerte miteinander oder mit einer angegebenen Zeichenkette verglichen. Beispielsweise können Sie Zeichenketten mit = auf Gleichheit überprüfen. Ein Beispiel für die Überprüfung der Zeichenkettengleichheit ist: *Class = "class 1"*.

Bei numerischen Vergleichen bedeutet *größer* näher an der positiven Unendlichkeit und *kleiner* näher an der negativen Unendlichkeit. Negativen Zahlen sind also stets kleiner als jede positive Zahl.

Funktion	Ergebnis	Beschreibung
count_equal(ITEM1, LIST)	Ganzzahl	Ergibt die Anzahl der Werte aus einer Felderliste, die gleich <i>ELEMENT1</i> sind, bzw. den Wert null, wenn <i>ELEMENT1</i> gleich null ist. Für weitere Informationen siehe Thema Zusammenfassen mehrerer Felder in Kapitel 7 auf S. 129 .
count_greater_than(ITEM1, LIST)	Ganzzahl	Ergibt die Anzahl der Werte aus einer Felderliste, die größer als <i>ELEMENT1</i> sind, bzw. den Wert null, wenn <i>ELEMENT1</i> gleich null ist.
count_less_than(ITEM1, LIST)	Ganzzahl	Ergibt die Anzahl der Werte aus einer Felderliste, die kleiner als <i>ELEMENT1</i> sind, bzw. den Wert null, wenn <i>ELEMENT1</i> gleich null ist.
count_not_equal(ITEM1, LIST)	Ganzzahl	Ergibt die Anzahl der Werte aus einer Felderliste, die ungleich <i>ELEMENT1</i> sind, bzw. den Wert null, wenn <i>ELEMENT1</i> gleich null ist.
count_nulls(LIST)	Ganzzahl	Ergibt die Anzahl der Nullwerte aus einer Felderliste.
count_non_nulls(LIST)	Ganzzahl	Ergibt die Anzahl der Nicht-Nullwerte aus einer Felderliste.
date_before(DATE1, DATE2)	Boolesch	Dient zur Überprüfung der alphabetischen Sortierung von Datumswerten. Ergibt den Wert "Wahr", wenn <i>DATUM1</i> vor <i>DATUM2</i> liegt.
first_index(ITEM, LIST)	Ganzzahl	Ergibt den Index des ersten Felds aus einer LISTE von Feldern, das ELEMENT enthält, oder "0", wenn der Wert nicht gefunden wird. Nur für die Typen "Zeichenkette", "Ganze Zahl" und "Reelle Zahl" unterstützt. Für weitere Informationen siehe Thema Arbeiten mit Mehrfachantwortdaten in Kapitel 7 auf S. 130 .
first_non_null(LIST)	Jede	Ergibt den ersten Nicht-Null-Wert in der angegebenen Felderliste. Alle Speichertypen werden unterstützt.
first_non_null_index(LIST)	Ganzzahl	Ergibt den Index des erstens Felds in der angegebenen LISTE, das einen Nicht-Null-Wert enthält, oder "0", wenn alle Werte null sind. Alle Speichertypen werden unterstützt.
ITEM1 = ITEM2	Boolesch	Ergibt den Wert "Wahr" für Datensätze, bei denen <i>ELEMENT1</i> gleich <i>ELEMENT2</i> ist.
ITEM1 /= ITEM2	Boolesch	Ergibt den Wert "Wahr", wenn die beiden Zeichenketten nicht identisch sind, oder "0", wenn sie identisch sind.
ITEM1 < ITEM2	Boolesch	Ergibt den Wert "Wahr" für Datensätze, bei denen <i>ELEMENT1</i> kleiner als <i>ELEMENT2</i> ist.
ITEM1 <= ITEM2	Boolesch	Ergibt den Wert "Wahr" für Datensätze, bei denen <i>ELEMENT1</i> kleiner oder gleich <i>ELEMENT2</i> ist.
ITEM1 > ITEM2	Boolesch	Ergibt den Wert "Wahr" für Datensätze, bei denen <i>ELEMENT1</i> größer als <i>ELEMENT2</i> ist.
ITEM1 >= ITEM2	Boolesch	Ergibt den Wert "Wahr" für Datensätze, bei denen <i>ELEMENT1</i> größer oder gleich <i>ELEMENT2</i> ist.
last_index(ITEM, LIST)	Ganzzahl	Ergibt den Index des letzten Felds aus einer LISTE von Feldern, das ELEMENT enthält, oder "0", wenn der Wert nicht gefunden wird. Nur für die Typen "Zeichenkette", "Ganze Zahl" und "Reelle Zahl" unterstützt. Für weitere Informationen siehe Thema Arbeiten mit Mehrfachantwortdaten in Kapitel 7 auf S. 130 .

Funktion	Ergebnis	Beschreibung
<code>last_non_null(LIST)</code>	<i>Jede</i>	Ergibt den letzten Nicht-Null-Wert in der angegebenen Felderliste. Alle Speichertypen werden unterstützt.
<code>last_non_null_index(LIST)</code>	<i>Ganzzahl</i>	Ergibt den Index des letzten Felds in der angegebenen LISTE, das einen Nicht-Null-Wert enthält, oder “0”, wenn alle Werte null sind. Alle Speichertypen werden unterstützt.
<code>max(ITEM1, ITEM2)</code>	<i>Jede</i>	Ergibt das größere der beiden Elemente: <i>ELEMENT1</i> oder <i>ELEMENT2</i> .
<code>max_index(LIST)</code>	<i>Ganzzahl</i>	Ergibt den Index des Felds mit dem größten Wert aus einer Liste numerischer Felder bzw. “0”, wenn alle Feldwerte null sind. Wenn beispielsweise das dritte Feld in der Liste den größten Wert enthält, wird der Indexwert 3 ausgegeben. Wenn mehrere Felder den größten Wert enthalten, wird das zuerst aufgelistete (von links nach rechts) ausgegeben. Für weitere Informationen siehe Thema Arbeiten mit Mehrfachantwortdaten in Kapitel 7 auf S. 130.
<code>max_n(LIST)</code>	<i>Number</i>	Ergibt den größten Wert aus einer Liste numerischer Felder bzw. den Wert null, wenn alle Feldwerte gleich null sind. Für weitere Informationen siehe Thema Zusammenfassen mehrerer Felder in Kapitel 7 auf S. 129.
<code>member(ITEM, LIST)</code>	<i>Boolesch</i>	Ergibt den Wert “Wahr”, wenn <i>ELEMENT</i> ein Mitglied der angegebenen <i>LISTE</i> ist. Andernfalls ergibt sich ein Falsch-Wert. Auch eine Liste mit Feldnamen kann angegeben werden. Für weitere Informationen siehe Thema Zusammenfassen mehrerer Felder in Kapitel 7 auf S. 129.
<code>min(ITEM1, ITEM2)</code>	<i>Jede</i>	Ergibt das kleinere der beiden Elemente: <i>ELEMENT1</i> oder <i>ELEMENT2</i> .
<code>min_index(LIST)</code>	<i>Ganzzahl</i>	Ergibt den Index des Felds mit dem kleinsten Wert aus einer Liste numerischer Felder bzw. “0”, wenn alle Feldwerte null sind. Wenn beispielsweise das dritte Feld in der Liste den kleinsten Wert enthält, wird der Indexwert 3 ausgegeben. Wenn mehrere Felder den kleinsten Wert enthalten, wird das zuerst aufgelistete (von links nach rechts) ausgegeben. Für weitere Informationen siehe Thema Arbeiten mit Mehrfachantwortdaten in Kapitel 7 auf S. 130.
<code>min_n(LIST)</code>	<i>Number</i>	Ergibt den kleinsten Wert aus einer Liste numerischer Felder bzw. den Wert null, wenn alle Feldwerte gleich null sind.
<code>time_before(TIME1, TIME2)</code>	<i>Boolesch</i>	Dient zur Überprüfung der alphabetischen Sortierung von Zeitwerten. Ergibt den Wert “Wahr”, wenn <i>ZEIT1</i> vor <i>ZEIT2</i> liegt.
<code>value_at(INT, LIST)</code>		Gibt den Wert jedes aufgelisteten Felds bei Offset GANZZ zurück oder NULL, wenn das Offset außerhalb des Bereichs der gültigen Werte liegt (d. h. kleiner als 1 oder größer als die Anzahl der aufgelisteten Felder ist). Alle Speichertypen werden unterstützt.

Logische Funktionen

CLEM-Ausdrücke können zur Durchführung logischer Operationen verwendet werden.

Funktion	Ergebnis	Beschreibung
COND1 and COND2	Boolesch	Bei dieser Operation handelt es sich um eine logische Konjunktion, die einen Wahr-Wert ergibt, wenn sowohl <i>BEDG1</i> als auch <i>BEDG2</i> wahr sind. Wenn <i>BEDG1</i> falsch ist, wird <i>BEDG2</i> nicht ausgewertet; dadurch sind Konjunktionen möglich, bei denen <i>BEDG1</i> zuerst testet, ob eine Operation in <i>BEDG2</i> zulässig ist. Beispiel: <code>length(Label) >=6</code> und <code>Label(6) = 'x'</code> .
COND1 or COND2	Boolesch	Bei dieser Operation handelt es sich um eine logische (inklusive) Disjunktion, die einen Wahr-Wert ergibt, wenn entweder <i>BEDG1</i> oder <i>BEDG2</i> wahr ist oder wenn beide wahr sind. Wenn <i>BEDG1</i> wahr ist, wird <i>BEDG2</i> nicht ausgewertet.
not(COND)	Boolesch	Bei dieser Operation handelt es sich um eine logische Negation, die einen Wahr-Wert ergibt, wenn <i>BEDG</i> falsch ist. Andernfalls gibt diese Operation den Wert 0 zurück.
if COND then EXPR1 else EXPR2 endif	Jede	Bei dieser Operation handelt es sich um eine bedingte Auswertung. Wenn <i>BEDG</i> wahr ist, ergibt die Operation das Ergebnis von <i>AUSDR1</i> . Andernfalls ergibt sich das Ergebnis der Auswertung von <i>AUSDR2</i> .
if COND1 then EXPR1 elseif COND2 then EXPR2 else EXPR_N endif	Jede	Bei dieser Operation handelt es sich um eine bedingte Auswertung mit mehreren Zweigen. Wenn <i>BEDG1</i> wahr ist, ergibt die Operation das Ergebnis von <i>AUSDR1</i> . Andernfalls, wenn <i>BEDG2</i> wahr ist, ergibt die Operation das Ergebnis der Auswertung von <i>AUSDR2</i> . Andernfalls ergibt sich das Ergebnis der Auswertung von <i>AUSDR_N</i> .

Numerische Funktionen

CLEM enthält eine Reihe häufig verwendeter numerischer Funktionen.

Funktion	Ergebnis	Beschreibung
-NUM	Number	Wird zur Umkehrung des Vorzeichens von <i>ZAHL</i> verwendet. Gibt die entsprechende Zahl mit dem entgegengesetzten Vorzeichen zurück.
NUM1 + NUM2	Number	Ergibt die Summe von <i>ZAHL1</i> und <i>ZAHL2</i> .
Code -NUM2	Number	Ergibt die Differenz von <i>ZAHL1</i> und <i>ZAHL2</i> .
NUM1 * NUM2	Number	Ergibt das Produkt aus <i>ZAHL1</i> und <i>ZAHL2</i> .
NUM1 / NUM2	Number	Ergibt den Quotienten aus <i>ZAHL1</i> und <i>ZAHL2</i> .
INT1 div INT2	Number	Wird zur Durchführung einer Division zwischen zwei ganzen Zahlen verwendet. Ergibt den Wert von <i>GANZZ1</i> dividiert durch <i>GANZZ2</i> .
INT1 rem INT2	Number	Ergibt den Rest des Quotienten aus <i>GANZZ1</i> und <i>GANZZ2</i> . Beispiel: <code>INT1 - (INT1 div INT2) * INT2</code> .
INT1 mod INT2	Number	Diese Funktion wurde verworfen. Verwenden Sie stattdessen die Funktion <code>rem</code> .

Funktion	Ergebnis	Beschreibung
BASE ** POWER	Number	Ergibt den Wert von <i>BASIS</i> hoch <i>POTENZ</i> . Für beide Werte ist eine beliebige Zahl möglich (mit der Ausnahme, dass <i>BASIS</i> nicht null sein darf, wenn <i>POTENZ</i> einen anderen Nullwert als die ganze Zahl 0 aufweist). Wenn <i>POTENZ</i> eine ganze Zahl ist, erfolgt die Berechnung, indem <i>BASIS</i> mehrmals nacheinander mit sich selbst multipliziert wird. Wenn <i>BASIS</i> also eine ganze Zahl ist, ist das Ergebnis ebenfalls eine ganze Zahl. Wenn <i>POTENZ</i> die ganze Zahl 0 ist, ist das Ergebnis immer eine 1 vom selben Typ wie <i>BASIS</i> . Andernfalls, wenn <i>POTENZ</i> keine ganze Zahl ist, wird das Ergebnis als $\exp(\text{POWER} * \log(\text{BASE}))$ berechnet.
abs(NUM)	Number	Ergibt den Absoluten Wert von <i>Zahl</i> , bei dem es sich immer um eine Zahl desselben Typs handelt.
exp(NUM)	Reelle Zahl	Ergibt den Wert von <i>e</i> hoch <i>Zahl</i> . Dabei ist <i>e</i> die Basis eines natürlichen Logarithmus.
fracof(NUM)	Reelle Zahl	Ergibt den Bruchbereich (Nachkommastellen) von <i>Zahl</i> , der definiert ist als NUM-intof(NUM).
intof(NUM)	Ganzzahl	Kürzt das Argument zu einer ganzen Zahl. Ergibt eine ganze Zahl mit demselben Vorzeichen wie <i>Zahl</i> und mit dem größten Absolutwert, sodass abs(INT) <= abs(NUM).
log(NUM)	Reelle Zahl	Ergibt den natürlichen Logarithmus (Basis <i>e</i>) von <i>Zahl</i> (darf keine Art von Nullwert sein).
log10(NUM)	Reelle Zahl	Ergibt den Logarithmus zur Basis 10 von <i>Zahl</i> (darf keine Art von Nullwert sein). Diese Funktion ist definiert als $\log(\text{NUM}) / \log(10)$.
negate(NUM)	Number	Wird zur Umkehrung des Vorzeichens von <i>Zahl</i> verwendet. Ergibt die entsprechende Zahl mit dem entgegengesetzten Vorzeichen.
round(NUM)	Ganzzahl	Dient zum Runden von <i>Zahl</i> auf eine ganze Zahl. Dabei wird intof(NUM+0.5) verwendet, wenn <i>Zahl</i> positiv ist, oder intof(NUM-0.5), wenn <i>Zahl</i> negativ ist.
sign(NUM)	Number	Wird zur Ermittlung des Vorzeichens von <i>Zahl</i> verwendet. Diese Operation ergibt -1, 0 oder 1, wenn <i>Zahl</i> eine ganze Zahl ist. Wenn <i>Zahl</i> eine reelle Zahl ist, ergibt sich -1,0; 0,0 oder 1,0, je nachdem ob <i>Zahl</i> negativ, null oder positiv ist.
sqrt(NUM)	Reelle Zahl	Ergibt die Quadratwurzel von <i>Zahl</i> . <i>Zahl</i> muss positiv sein.
sum_n(LIST)	Number	Ergibt die Summe der Werte aus einer Liste numerischer Felder bzw. den Wert null, wenn alle Feldwerte gleich null sind. Für weitere Informationen siehe Thema Zusammenfassen mehrerer Felder in Kapitel 7 auf S. 129 .
mean_n(LIST)	Number	Ergibt den Mittelwert aus einer Liste numerischer Felder bzw. den Wert null, wenn alle Feldwerte gleich null sind.
sdev_n(LIST)	Number	Ergibt die Standardabweichung aus einer Liste numerischer Felder bzw. den Wert null, wenn alle Feldwerte gleich null sind.

Trigonometrische Funktionen

Alle Funktionen in diesem Abschnitt verwenden entweder einen Winkel als Argument oder ergeben einen Winkel als Ergebnis. In beiden Fällen richtet sich die für den Winkel verwendete Einheit (Radianen oder Grad) nach die Einstellung in der entsprechenden Stream-Option.

Funktion	Ergebnis	Beschreibung
arccos(NUM)	Reelle Zahl	Berechnet den Arcuscosinus des angegebenen Winkels.
arccosh(NUM)	Reelle Zahl	Berechnet den hyperbolischen Arcuscosinus des angegebenen Winkels.
arcsin(NUM)	Reelle Zahl	Berechnet den Arcussinus des angegebenen Winkels.
arcsinh(NUM)	Reelle Zahl	Berechnet den hyperbolischen Arcussinus des angegebenen Winkels.
arctan(NUM)	Reelle Zahl	Berechnet den Arcustangens des angegebenen Winkels.
arctan2(NUM_Y, NUM_X)	Reelle Zahl	Berechnet den Arcustangens von NUM_Y / NUM_X und verwendet die Vorzeichen der beiden Zahlen zur Ableitung von Quadranteninformationen. Das Ergebnis ist eine reelle Zahl im Bereich von - pi < ANGLE <= pi (radians) – 180 < ANGLE <= 180 (degrees).
arctanh(NUM)	Reelle Zahl	Berechnet den hyperbolischen Arcustangens des angegebenen Winkels.
cos(NUM)	Reelle Zahl	Berechnet den Cosinus des angegebenen Winkels.
cosh(NUM)	Reelle Zahl	Berechnet den hyperbolischen Cosinus des angegebenen Winkels.
pi	Reelle Zahl	Diese Konstante ist die beste reelle Annäherung an Pi.
sin(NUM)	Reelle Zahl	Berechnet den Sinus des angegebenen Winkels.
sinh(NUM)	Reelle Zahl	Berechnet den hyperbolischen Sinus des angegebenen Winkels.
tan(NUM)	Reelle Zahl	Berechnet den Tangens des angegebenen Winkels.
tanh(NUM)	Reelle Zahl	Berechnet den hyperbolischen Tangens des angegebenen Winkels.

Wahrscheinlichkeitsfunktionen

Wahrscheinlichkeitsfunktionen ergeben Wahrscheinlichkeiten auf der Grundlage verschiedener Verteilungen, z. B. die Wahrscheinlichkeit, dass ein Wert aus der t -Verteilung in der Studentenversion kleiner ist als ein bestimmter Wert.

Funktion	Ergebnis	Beschreibung
cdf_chisq(NUM, DF)	Reelle Zahl	Ergibt die Wahrscheinlichkeit, dass ein Wert aus der Chi-Quadrat-Verteilung mit den angegebenen Freiheitsgraden kleiner ist als die angegebene Zahl.
cdf_f(NUM, DF1, DF2)	Reelle Zahl	Ergibt die Wahrscheinlichkeit, dass ein Wert aus der F -Verteilung mit den angegebenen Freiheitsgraden $DF1$ und $DF2$ kleiner ist als die angegebene Zahl.

Funktion	Ergebnis	Beschreibung
<code>cdf_normal(NUM, MEAN, STDDEV)</code>	<i>Reelle Zahl</i>	Ergibt die Wahrscheinlichkeit, dass ein Wert aus der Normalverteilung mit dem angegebenen Mittelwert und der angegebenen Standardabweichung kleiner ist als die angegebene Zahl.
<code>cdf_t(NUM, DF)</code>	<i>Reelle Zahl</i>	Ergibt die Wahrscheinlichkeit, dass ein Wert aus der Chi-Quadrat-Verteilung aus einer <i>t</i> -Verteilung in der Studentenversion mit den angegebenen Freiheitsgraden kleiner ist als die angegebene Zahl.

Bitweise Operationen mit ganzen Zahlen

Mit diesen Funktionen können Ganzzahlen als Bitmuster behandelt werden, die die Zweierkomplement-Werte darstellen, bei denen die Bit-Position N den Wert 2^{**N} aufweist. Bits werden von 0 aufwärts nummeriert. Diese Operationen funktionieren so, als ob das Vorzeichen-Bit einer ganzen Zahl endlos nach links erweitert wird. Daher weisen positive ganze Zahlen überall oberhalb des werthöchsten 0-Bits auf, negative ganze Zahlen 1-Bits.

Hinweis: Die bitweisen Funktionen können nicht aus Skripts heraus aufgerufen werden. Für weitere Informationen siehe Thema [CLEM-Ausdrücke in Skripts in Kapitel 3 in IBM SPSS Modeler 15 Handbuch für Skripterstellung und Automatisierung](#).

Funktion	Ergebnis	Beschreibung
<code>~~ INT1</code>	<i>Ganzzahl</i>	Liefert das bitweise Komplement der ganzen Zahl <i>GANZZ1</i> . Im Ergebnis steht also eine 1 für jede Bitposition, bei der <i>GANZZ1</i> den Wert 0 aufweist. Es gilt immer: <code>~~ INT = -(INT + 1)</code> . Beachten Sie, dass diese Funktion nicht über ein Skript aufgerufen werden kann. Für weitere Informationen siehe Thema CLEM-Ausdrücke in Skripts in Kapitel 3 in IBM SPSS Modeler 15 Handbuch für Skripterstellung und Automatisierung .
<code>INT1 INT2</code>	<i>Ganzzahl</i>	Das Ergebnis dieser Operation ist das bitweise “inklusive Oder” von <i>GANZZ1</i> und <i>GANZZ2</i> . Im Ergebnis steht also eine 1 für jede Bitposition, bei der bei <i>GANZZ1</i> und/oder bei <i>GANZZ2</i> der Wert 1 steht.
<code>INT1 & INT2</code>	<i>Ganzzahl</i>	Das Ergebnis dieser Operation ist das bitweise “exklusive Oder” von <i>GANZZ1</i> und <i>GANZZ2</i> . Im Ergebnis steht also eine 1 für jede Bitposition, bei der entweder bei <i>GANZZ1</i> oder bei <i>GANZZ2</i> der Wert 1 steht, nicht jedoch bei beiden.
<code>INT1 && INT2</code>	<i>Ganzzahl</i>	Liefert das bitweise “Und” der ganzen Zahlen <i>GANZZ1</i> und <i>GANZZ2</i> . Im Ergebnis steht also eine 1 für jede Bitposition, bei der sowohl bei <i>GANZZ1</i> als auch bei <i>GANZZ2</i> der Wert 1 steht.
<code>INT1 &&~~ INT2</code>	<i>Ganzzahl</i>	Liefert das bitweise “Und” von <i>GANZZ1</i> und das bitweise Komplement von <i>GANZZ2</i> . Im Ergebnis steht also eine 1 für jede Bitposition, bei der bei <i>GANZZ1</i> der Wert 1 und bei <i>GANZZ2</i> der Wert 0 steht. Diese Operation ist identisch mit <code>INT1&& (~INT2)</code> und ist sinnvoll zum Bereinigen von in <i>GANZZ2</i> gesetzten Bits von <i>GANZZ1</i> .

Funktion	Ergebnis	Beschreibung
INT << N	Ganzzahl	Ergibt das Bitmuster von <i>GANZZI</i> , um <i>N</i> Positionen nach links verschoben. Ein negativer Wert von <i>N</i> führt zu einer Verschiebung nach rechts.
INT >> N	Ganzzahl	Ergibt das Bitmuster von <i>GANZZI</i> , um <i>N</i> Positionen nach rechts verschoben. Ein negativer Wert von <i>N</i> führt zu einer Verschiebung nach links.
INT1 &&=_0 INT2	Boolesch	Entspricht dem Boole'schen Ausdruck INT1 && INT2 /== 0, ist jedoch effizienter.
INT1 &&/_=0 INT2	Boolesch	Entspricht dem Boole'schen Ausdruck INT1 && INT2 == 0, ist jedoch effizienter.
integer_bitcount(INT)	Ganzzahl	Zählt die Anzahl der 1- oder 0-Bits in der Zweierkomplement-Darstellung von <i>GANZZ</i> . Wenn <i>GANZZ</i> nichtnegativ ist, ist <i>N</i> die Anzahl der 1-Bits. Wenn <i>GANZZ</i> negativ ist, gibt dieser Wert die Anzahl der 0-Bits an. Aufgrund der Vorzeichenerweiterung gibt es unendlich Viele 0-Bits in nichtnegativen Ganzzahlen bzw. 1-Bits in negativen Ganzzahlen. Es gilt immer: integer_bitcount(INT) = integer_bitcount(-(INT+1)).
integer_leastbit(INT)	Ganzzahl	Gibt die Bitposition <i>N</i> des wertniedrigsten Bits an, das in der ganzen Zahl <i>GANZZ</i> gesetzt wurde. <i>N</i> ist die höchste Potenz von 2, durch die sich <i>GANZZ</i> ohne Rest teilen lässt.
integer_length(INT)	Ganzzahl	Ergibt die Länge von <i>GANZZ</i> als ganze Zweierkomplement-Zahl in Bit. Im Ergebnis ist also <i>N</i> die kleinste Ganzzahl, sodass INT < (1 << N) if INT >= 0 INT >= (-1 << N) if INT < 0. Wenn <i>GANZZ</i> nichtnegativ ist, ist für die Darstellung von <i>GANZZ</i> als vorzeichenlose ganze Zahl ein Feld mit mindestens <i>N</i> Bit erforderlich. Alternativ sind, unabhängig vom Vorzeichen, mindestens <i>N+1</i> Bit erforderlich, um <i>GANZZ</i> als ganze Zahl mit Vorzeichen darzustellen.
testbit(INT, N)	Boolesch	Testet das Bit an Position <i>N</i> in der ganzen Zahl <i>GANZZ</i> und ergibt den Status von Bit <i>N</i> als Boole'schen Wert, der bei 1 "wahr" und bei 0 "falsch" ist.

Zufallsfunktionen

Folgende Funktionen dienen zur zufälligen Auswahl von Elementen oder zur Generierung von Zufallszahlen.

Funktion	Ergebnis	Beschreibung
oneof(LIST)	Jede	Ergibt ein zufällig ausgewähltes Element von <i>LISTE</i> . Die Listenelemente sollten wie folgt eingegeben werden [ITEM1,ITEM2,...,ITEM_N]. Auch eine Liste mit Feldnamen kann angegeben werden. Für weitere Informationen siehe Thema Zusammenfassen mehrerer Felder in Kapitel 7 auf S. 129 .
random(NUM)	Number	Ergibt eine gleichmäßig verteilte Zufallszahl desselben Typs (<i>GANZZ</i> oder <i>REELL</i>). Es gilt der Bereich von 1 bis <i>ZAHL</i> . Bei Verwendung einer ganzen Zahl ergeben sich nur ganze Zahlen. Bei Verwendung einer reellen Zahl (Dezimalzahl), ergeben sich reelle Zahlen (die Anzahl der Dezimalstellen werden in den Stream-Optionen festgelegt). Die größte Zufallszahl, die die Funktion ergibt, kann gleich <i>ZAHL</i> sein.

Funktion	Ergebnis	Beschreibung
random0(NUM)	Number	Diese Operation hat dieselben Eigenschaften wie random(NUM), der Bereich beginnt jedoch bei 0. Die größte Zufallszahl, die die Funktion ergibt, ist nie gleich X.

String-Funktionen

In CLEM können Sie folgende Operationen an Zeichenketten durchführen:

- Vergleichen von Zeichenketten
- Erstellen von Zeichenketten
- Zugreifen auf Zeichen

In CLEM-Ausdrücken ist eine Zeichenkette jede Folge von Zeichen zwischen zusammengehörenden (doppelten) Anführungszeichen ("string quotes"). Zeichen (CHAR) können beliebige alphanumerische Zeichen sein. Sie werden in CLEM-Ausdrücken unter Verwendung von einfachen umgekehrten Anführungszeichen in folgendem Format deklariert: `<Zeichen>`. Beispiele: `z`, `A` oder `2`. Zeichen, die außerhalb des zulässigen Bereichs liegen, oder negative Indizes bei einer Zeichenkette führen zu einem undefinierten Verhalten.

Anmerkung. Vergleiche zwischen Zeichenketten, die SQL-Pushback verwenden oder nicht, können unterschiedliche Ergebnisse generieren, wenn nachgestellte Leerzeichen vorhanden sind.

Funktion	Ergebnis	Beschreibung
allbutfirst(N, STRING)	Zeichenfolge	Ergibt eine Zeichenkette, bei der es sich um ZEICHENKETTE nach der Entfernung der ersten N Zeichen handelt.
allbutlast(N, STRING)	Zeichenfolge	Ergibt eine Zeichenkette, bei der es sich um ZEICHENKETTE nach der Entfernung der letzten N Zeichen handelt.
alphabefore(STRING1, STRING2)	Boolesch	Dient zur Überprüfung der alphabetischen Sortierung von Zeichenketten. Ergibt den Wert "Wahr", wenn ZEICHENKETTE1 der ZEICHENKETTE2 vorausgeht.
endstring(LENGTH, STRING)	Zeichenfolge	Extrahiert die letzten N Zeichen aus der angegebenen Zeichenkette. Ist eine Zeichenkette kleiner oder gleich der angegebenen Länge, dann bleibt die Zeichenkette unverändert.
hasendstring(STRING, SUBSTRING)	Ganzzahl	Diese Funktion entspricht isendstring(SUBSTRING, STRING)..
hasmidstring(STRING, SUBSTRING)	Ganzzahl	Diese Funktion entspricht ismidstring(SUBSTRING, STRING) (eingebettete Teilzeichenkette).
hasstartstring(STRING, SUBSTRING)	Ganzzahl	Diese Funktion entspricht isstartstring(SUBSTRING, STRING).
hassubstring(STRING, N, SUBSTRING)	Ganzzahl	Diese Funktion entspricht issubstring(SUBSTRING, N, STRING), wobei für N standardmäßig der Wert 1 verwendet wird.

Funktion	Ergebnis	Beschreibung
count_substring(STRING, SUBSTRING)	Ganzzahl	Ergibt den Wert, wie oft die angegebene Teilzeichenkette in der Zeichenkette auftritt. Beispiel: count_substring("foooo.txt", "oo") ergibt 3.
hassubstring(STRING, SUBSTRING)	Ganzzahl	Diese Funktion entspricht issubstring(SUBSTRING, 1, STRING), wobei für N standardmäßig der Wert 1 verwendet wird.
isalphacode(CHAR)	Boolesch	Ergibt den Wert "Wahr", wenn $CHAR$ ein Zeichen in der angegebenen Zeichenkette (häufig ein Feldname) ist, dessen Zeichencode ein Buchstabe ist. Andernfalls ergibt diese Funktion den Wert 0. Beispiel: isalphacode(produce_num(1)).
isendstring(SUBSTRING, STRING)	Ganzzahl	Wenn die Zeichenkette ZEICHENKETTE mit der Teilzeichenkette TEILZEICHENKETTE endet, ergibt diese Funktion den ganzzahligen Index von TEILZEICHENKETTE in ZEICHENKETTE. Andernfalls ergibt die Funktion den Wert 0.
islowercode(CHAR)	Boolesch	Ergibt den Wert "Wahr", wenn $CHAR$ ein Kleinbuchstabe der angegebenen Zeichenkette (häufig ein Feldname) ist. Andernfalls ergibt die Funktion den Wert 0. Beispielsweise sind sowohl islowercode(`) als auch islowercode(country_name(2)) gültige Ausdrücke.
ismidstring(SUBSTRING, STRING)	Ganzzahl	Wenn TEILZEICHENKETTE eine Teilzeichenkette von ZEICHENKETTE ist, jedoch nicht beim ersten Zeichen von ZEICHENKETTE beginnt oder mit dem letzten Zeichen endet, ergibt diese Funktion den Index, bei dem die Teilzeichenkette beginnt. Andernfalls gibt die Funktion den Wert 0 zurück.
isnumbercode(CHAR)	Boolesch	Ergibt den Wert "Wahr", wenn $CHAR$ bei der angegebenen Zeichenkette (häufig ein Feldname) ein Zeichen ist, dessen Zeichencode eine Ziffer ist. Andernfalls ergibt diese Funktion den Wert 0. Beispiel: isnumbercode(product_id(2)).
isstartstring(SUBSTRING, STRING)	Ganzzahl	Wenn die Zeichenkette ZEICHENKETTE mit der Teilzeichenkette TEILZEICHENKETTE beginnt, ergibt diese Funktion den Index 1. Andernfalls ergibt die Funktion den Wert 0.
issubstring(SUBSTRING, N, STRING)	Ganzzahl	Durchsucht die Zeichenkette ZEICHENKETTE beginnend beim N . Zeichen nach einer Teilzeichenkette, die mit TEILZEICHENKETTE übereinstimmt. Wenn die Teilzeichenkette gefunden wird, ergibt diese Funktion den Index, bei dem die Teilzeichenkette beginnt. Andernfalls ergibt diese Funktion den Wert 0. Wenn N nicht angegeben ist, wird standardmäßig der Wert 1 verwendet.

Funktion	Ergebnis	Beschreibung
issubstring(SUBSTRING, STRING)	Ganzzahl	Durchsucht die Zeichenkette <i>ZEICHENKETTE</i> beginnend beim <i>N</i> . Zeichen nach einer Teilzeichenkette, die mit <i>TEILZEICHENKETTE</i> übereinstimmt. Wenn die Teilzeichenkette gefunden wird, ergibt diese Funktion den Index, bei dem die Teilzeichenkette beginnt. Andernfalls ergibt diese Funktion den Wert 0. Wenn <i>N</i> nicht angegeben ist, wird standardmäßig der Wert 1 verwendet.
issubstring_count(SUBSTRING, N, STRING):	Ganzzahl	Ergibt den Index für das <i>N</i> -te Vorkommen von <i>TEILZEICHENKETTE</i> in der angegebenen <i>ZEICHENKETTE</i> . Falls <i>TEILZEICHENKETTE</i> weniger als <i>N</i> -mal auftritt, ergibt sich der Wert 0.
issubstring_lim(SUBSTRING, N, STARTLIM, ENDLIM, STRING)	Ganzzahl	Diese Funktion entspricht <i>issubstring</i> , der Treffer muss jedoch beim oder vor dem Index <i>STARTLIM</i> beginnen und beim oder vor dem Index <i>ENDLIM</i> enden. Die Beschränkungen <i>STARTLIM</i> bzw. <i>ENDLIM</i> können deaktiviert werden, indem ein Falsch-Wert für die Argumente angegeben wird. So ist beispielsweise <i>issubstring_lim(SUBSTRING, N, false, false, STRING)</i> mit <i>issubstring</i> identisch.
isuppercode(CHAR)	Boolesch	Ergibt den Wert "Wahr", wenn <i>CHAR</i> ein Großbuchstabe ist. Andernfalls ergibt die Funktion den Wert 0. Beispielsweise sind sowohl <i>isuppercode(``)</i> als auch <i>isuppercode(country_name(2))</i> gültige Ausdrücke.
last(CHAR)	Zeichenfolge	Ergibt das letzte Zeichen <i>CHAR</i> von <i>ZEICHENKETTE</i> (muss mindestens ein Zeichen lang sein).
length(STRING)	Ganzzahl	Ergibt die Länge der Zeichenkette <i>ZEICHENKETTE</i> , d. h. die Anzahl der Zeichen in der Zeichenkette.
locchar(CHAR, N, STRING)	Ganzzahl	Mit dieser Funktion kann die Position der Zeichen in symbolischen Feldern ermittelt werden. Mit dieser Funktion wird die Zeichenkette <i>ZEICHENKETTE</i> nach dem Zeichen <i>CHAR</i> durchsucht, wobei der Suchvorgang beim <i>N</i> . Zeichen von <i>ZEICHENKETTE</i> gestartet wird. Diese Funktion ergibt einen Wert (mit Startwert <i>N</i>), der die Position angibt, an der das Zeichen gefunden wurde. Wenn das Zeichen nicht gefunden wurde, ergibt die Funktion den Wert 0. Wenn die Funktion ein ungültiges Offset (<i>N</i>) aufweist (z. B. ein Offset, das außerhalb der Länge des Strings liegt), ergibt die Funktion \$null\$. <i>locchar(`n`, 2, web_page)</i> beispielsweise durchsucht das Feld <i>Webseite</i> nach dem Zeichen `n` ausgehend vom zweiten Zeichen im Feldwert.

Funktion	Ergebnis	Beschreibung
		<i>Hinweis:</i> Achten Sie darauf, das angegebene Zeichen in einzelne umgekehrte Anführungszeichen einzuschließen.
locchar_back(CHAR, N, STRING)	Ganzzahl	Ähnlich wie locchar, außer dass die Suche ausgehend vom <i>N</i> -ten Zeichen rückwärts durchgeführt wird. Beispiel: <code>locchar_back('n', 9, web_page)</code> durchsucht das Feld <i>Webseite</i> ausgehend vom 9. Zeichen rückwärts bis zum Anfang der Zeichenkette. Wenn die Funktion ein ungültiges Offset aufweist (z. B. ein Offset, das außerhalb der Länge des Strings liegt), ergibt die Funktion \$null\$. Idealerweise sollte locchar_back in Verbindung mit der Funktion length(<field>) verwendet werden, um die Länge des aktuellen Wertes des Feldes dynamisch zu verwenden. Beispiel: <code>locchar_back('n', length(web_page), web_page)</code> .
lowertoupper(CHAR) lowertoupper (STRING)	CHAR oder Zeichenkette	Die Eingabe kann entweder eine Zeichenkette oder ein Zeichen sein. Diese wird in dieser Funktion verwendet, um ein neues Element desselben Typs auszugeben, wobei alle Kleinbuchstaben in die entsprechenden Großbuchstaben konvertiert wurden. <code>lowertoupper('a')</code> , <code>lowertoupper("My string")</code> und <code>lowertoupper(field_name(2))</code> sind beispielsweise alle gültige Ausdrücke.
matches	Boolesch	Ergibt den Wert "wahr", wenn eine Zeichenkette mit einem bestimmten Muster übereinstimmt. Das Muster muss aus einer Literalzeichenkette bestehen; der Name eines Felds, das ein Muster enthält, ist nicht zulässig. Das Fragezeichen (?) als Platzhalterzeichen kann die Stelle von genau einem beliebigen Zeichen einnehmen, das Sternchen (*) entspricht keinem, einem oder mehreren Zeichen. Soll ein Fragezeichen oder ein Sternchen nicht als Platzhalter fungieren, sondern bei der Suche berücksichtigt werden, geben Sie den umgekehrten Schrägstrich () als Escape-Zeichen ein.
replace(SUBSTRING, NEWSUBSTRING, STRING)	Zeichenfolge	Innerhalb der angegebenen ZEICHENKETTE werden alle Instanzen von TEILZEICHENKETTE durch NEUE_TEILZEICHENKETTE ersetzt.
replicate(COUNT, STRING)	Zeichenfolge	Ergibt eine Zeichenkette, die aus der ursprünglichen Zeichenkette besteht, die so oft wie angegeben kopiert wurde.

Funktion	Ergebnis	Beschreibung
stripchar(CHAR, STRING)	Zeichenfolge	Mit dieser Funktion können Sie die angegebenen Zeichen aus einer Zeichenkette oder einem Feld entfernen. Mit dieser Funktion können Sie beispielsweise zusätzliche Symbole, wie Währungsangaben, aus den Daten entfernen, um eine einfache Zahl oder einen einfachen Namen zu erhalten. Mit der Syntax <code>stripchar('\$', 'Cost')</code> beispielsweise erhalten Sie ein neues Feld, bei dem das Dollar-Zeichen aus allen Werten entfernt ist. <i>Hinweis:</i> Achten Sie darauf, das angegebene Zeichen in einzelne umgekehrte Anführungszeichen einzuschließen.
skipchar(CHAR, N, STRING)	Ganzzahl	Durchsucht die Zeichenkette <i>ZEICHENKETTE</i> nach allen Zeichen außer <i>CHAR</i> und beginnt dabei beim <i>N</i> . Zeichen. Diese Funktion ergibt eine ganzzahlige Teilzeichenkette, die den Punkt angibt, an dem ein solches Zeichen gefunden wurde, oder 0, wenn es sich bei jedem Zeichen vom <i>N</i> . Zeichen an um <i>CHAR</i> handelt. Wenn die Funktion ein ungültiges Offset aufweist (z. B. ein Offset, das außerhalb der Länge des Strings liegt), ergibt die Funktion \$null\$. locchar wird häufig in Verbindung mit den Funktionen vom Typ skipchar verwendet, um den Wert von <i>N</i> (die Stelle, an der mit der Suche begonnen werden soll) zu bestimmen. Beispiel: <code>skipchar('s', (locchar('s', 1, "MyString")), "MyString")</code> .
skipchar_back(CHAR, N, STRING)	Ganzzahl	Ähnlich wie skipchar, außer dass die Suche ausgehend vom <i>N</i> -ten Zeichen rückwärts durchgeführt wird.
startstring(LENGTH, STRING)	Zeichenfolge	Extrahiert die ersten <i>N</i> Zeichen aus der angegebenen Zeichenkette. Ist eine Zeichenkette kleiner oder gleich der angegebenen Länge, dann bleibt die Zeichenkette unverändert.
strmember(CHAR, STRING)	Ganzzahl	Entspricht locchar(CHAR, 1, STRING). Ergibt eine ganzzahlige Teilzeichenkette, die entweder den Punkt angibt, an dem <i>CHAR</i> zum ersten Mal auftritt, oder den Wert 0. Wenn die Funktion ein ungültiges Offset aufweist (z. B. ein Offset, das außerhalb der Länge des Strings liegt), ergibt diese Funktion \$null\$.
subscrs(N, STRING)	CHAR	Ergibt das <i>N</i> . Zeichen <i>CHAR</i> der Eingabezeichenkette <i>ZEICHENKETTE</i> . Diese Funktion kann auch in einer abgekürzten Form geschrieben werden, nämlich als <code>STRING(N)</code> . <code>lowertoupper("name"(1))</code> beispielsweise ist ein gültiger Ausdruck.

Funktion	Ergebnis	Beschreibung
substring(N, LEN, STRING)	Zeichen-folge	Ergibt die Zeichenkette <i>TEILZEICHEN-KETTE</i> , die aus den <i>LÄNGE</i> Zeichen der Zeichenkette <i>ZEICHENKETTE</i> besteht, wobei mit dem Zeichen bei Index <i>N</i> begonnen wird.
substring_between(N1, N2, STRING)	Zeichen-folge	Ergibt die Teilzeichenkette von <i>ZEICHENKETTE</i> , die bei Index <i>N1</i> beginnt und bei Index <i>N2</i> endet.
trim(STRING)	Zeichen-folge	Entfernt führende und nachgestellte Leerzeichen aus der angegebenen Zeichenkette.
trim_start(STRING)	Zeichen-folge	Entfernt führende Leerzeichen aus der angegebenen Zeichenkette.
trimend(STRING)	Zeichen-folge	Entfernt nachgestellte Leerzeichen aus der angegebenen Zeichenkette.
unicode_char(NUM)	CHAR	Ergibt das Zeichen mit dem Unicode-Wert <i>ZAHL</i> .
unicode_value(CHAR)	NUM	Ergibt den Unicode-Wert für das Zeichen <i>CHAR</i> .
uppertolower(CHAR) uppertolower (STRING)	CHAR oder Zeichenkette	Die Eingabe kann entweder eine Zeichenkette oder ein Zeichen sein und wird in dieser Funktion verwendet, um ein neues Element desselben Typs auszugeben, wobei alle Großbuchstaben in die entsprechenden Kleinbuchstaben konvertiert wurden. <i>Hinweis:</i> Achten Sie darauf, Zeichenketten in doppelte Anführungszeichen und Zeichen in einzelne, umgekehrte Anführungszeichen zu setzen. Einzelne Feldnamen sollten ohne Anführungszeichen angegeben werden.

SoundEx-Funktionen

SoundEx ist eine Methode zum Auffinden von Zeichenketten, deren Klang bekannt ist, nicht jedoch deren genaue Schreibweise. Die Methode wurde im Jahr 1918 entwickelt. Hiermit werden Wörter mit ähnlichen Lauten ermittelt, und zwar auf der Grundlage phonetischer Annahmen, wie bestimmte Buchstaben ausgesprochen werden. Suchen Sie beispielsweise Namen in einer Datenbank, bei denen die Schreibweise und die Betonung ähnlicher Namen voneinander abweichen können. Der grundlegende SoundEx-Algorithmus ist in zahlreichen Schriften dokumentiert. Trotz bekannter Einschränkungen (führende Buchstaben und Buchstabenfolgen wie

ph und f gelten beispielsweise nicht als Übereinstimmung, obwohl sie gleich klingen) wird diese Methode in den meisten Datenbanken in irgendeiner Form genutzt.

Funktion	Ergebnis	Beschreibung
soundex(STRING)	Ganzzahl	Ergibt den vierbuchstabigen SoundEx-Code für die angegebene ZEICHENKETTE.
soundex_difference(STRING1, STRING2)	Ganzzahl	Ergibt eine ganze Zahl zwischen 0 und 4, aus der die Anzahl der Zeichen hervorgeht, die im SoundEx-Code für die beiden Zeichenketten übereinstimmen. Der Wert 0 bedeutet dabei, dass gar keine Ähnlichkeit zwischen den Zeichenketten vorliegt, der Wert 4 weist auf sehr ähnliche oder sogar identische Zeichenketten hin.

Datums- und Uhrzeitfunktionen

CLEM beinhaltet eine Gruppe von Funktionen zum Umgang mit Feldern mit dem Speichertyp "Datum/Uhrzeit" von Stringvariablen, die für Datums- und Zeitangaben stehen. Die Formate für Datum und Uhrzeit sind für jeden Stream spezifisch und werden im Dialogfeld "Stream-Eigenschaften" angegeben. Die Datums- und Zeitfunktionen analysieren Datums- und Zeitzeichenketten gemäß dem aktuell ausgewählten Format.

Wenn Sie ein Jahr in einem Datumsformat angeben, bei dem nur zwei Ziffern verwendet werden (d. h., das Jahrhundert ist nicht angegeben), verwendet IBM® SPSS® Modeler das im Dialogfeld "Stream-Eigenschaften" angegebene Standard-Jahrhundert.

Hinweis: Die Datums- und Zeitfunktionen können nicht aus Skripts heraus aufgerufen werden. Für weitere Informationen siehe Thema CLEM-Ausdrücke in Skripts in Kapitel 3 in *IBM SPSS Modeler 15 Handbuch für Skripterstellung und Automatisierung*.

Funktion	Ergebnis	Beschreibung
@TODAY	Zeichenfolge	Bei Auswahl von Tage/Minuten übertragen im Dialogfeld "Stream-Eigenschaften" ergibt diese Funktion das aktuelle Datum als Zeichenkette im aktuellen Datumsformat. Wenn Sie ein zweistelliges Datumsformat verwenden und nicht die Option Tage/Minuten übertragen verwenden, ergibt diese Funktion \$null\$ auf dem aktuellen Server. Beachten Sie, dass diese Funktion nicht über ein Skript aufgerufen werden kann. Für weitere Informationen siehe Thema CLEM-Ausdrücke in Skripts in Kapitel 3 in <i>IBM SPSS Modeler 15 Handbuch für Skripterstellung und Automatisierung</i> .
to_time(ITEM)	Zeit	Konvertiert den Speichertyp des angegebenen Felds in eine Zeit.
to_date(ITEM)	Datum	Konvertiert den Speichertyp des angegebenen Felds in ein Datum.
to_timestamp(ITEM)	Zeitstempel	Konvertiert den Speichertyp des angegebenen Felds in einen Zeitstempel.
to_datetime(ITEM)	Datetime	Konvertiert den Speichertyp des angegebenen Felds in einen Wert vom Typ "Datum", "Zeit" oder "Zeitstempel".

Funktion	Ergebnis	Beschreibung
datetime_date(ITEM)	Datum	Gibt den Datumswert für eine <i>Zahl</i> , eine <i>Zeichenkette</i> oder einen <i>Zeitstempel</i> zurück. Beachten Sie, dass dies die einzige Funktion ist, mit der Sie eine Zahl (in wenigen Sekunden) wieder in ein Datum zurückkonvertieren können. Wenn es sich bei ITEM um eine Zeichenkette handelt, wird ein Datum erstellt, indem eine Zeichenkette im aktuellen Datumsformat analysiert wird. Das im Dialogfeld für die Stream-Eigenschaften angegebene Datumsformat muss korrekt sein, damit diese Funktion erfolgreich ausgeführt werden kann. Wenn es sich bei ITEM um eine Zahl handelt, wird sie als Anzahl von Sekunden seit dem Basisdatum (Epocha) interpretiert. Die Bruchteile eines Tages werden gekürzt. Wenn es sich bei ITEM um einen Zeitstempel handelt, wird der Datumsteil des Zeitstempels ausgegeben. Wenn es sich bei ITEM um ein Datum handelt, wird es unverändert ausgegeben.
date_before(DATE1, DATE2)	Boolesch	Ergibt den Wert "Wahr", wenn DATUM1 ein Datum oder einen Zeitstempel darstellt, das/der vor dem durch DATUM2 dargestellten Datum liegt. Andernfalls ergibt die Funktion den Wert 0.
date_days_difference(DATE1, DATE2)	Ganzzahl	Ergibt die Zeit in Tagen (als ganze Zahl), die zwischen dem durch DATUM1 und dem durch DATUM2 angegebenen Datum bzw. Zeitstempel liegt. Wenn DATUM2 vor DATUM1 liegt, ergibt diese Funktion eine negative Zahl.
date_in_days(DATE)	Ganzzahl	Ergibt die Zeit in Tagen (als ganze Zahl), die zwischen dem Basisdatum und dem durch DATUM angegebenen Datum oder Zeitstempel liegt. Wenn DATUM vor dem Basisdatum liegt, ergibt diese Funktion eine negative Zahl. Es muss ein gültiges Datum angegeben werden, damit die Berechnung ordnungsgemäß erfolgen kann. Beispielsweise sollten Sie nicht den 29. Februar 2001 als Datum angeben. Da 2001 kein Schaltjahr war, gibt es dieses Datum nicht.
date_in_months(DATE)	Reelle Zahl	Ergibt die Zeit in Monaten (als reelle Zahl), die zwischen dem Basisdatum und dem durch DATUM angegebenen Datum oder Zeitstempel liegt. Dies ist ein Näherungswert, der auf einem Monat mit 30.4375 Tagen beruht. Wenn DATUM vor dem Basisdatum liegt, ergibt diese Funktion eine negative Zahl. Es muss ein gültiges Datum angegeben werden, damit die Berechnung ordnungsgemäß erfolgen kann. Beispielsweise sollten Sie nicht den 29. Februar 2001 als Datum angeben. Da 2001 kein Schaltjahr war, gibt es dieses Datum nicht.
date_in_weeks(DATE)	Reelle Zahl	Ergibt die Zeit in Wochen (als reelle Zahl), die zwischen dem Basisdatum und dem durch DATUM angegebenen Datum oder Zeitstempel liegt. Der Wert basiert auf einer Woche mit 7,0 Tagen. Wenn DATUM vor dem Basisdatum liegt, ergibt diese Funktion eine negative Zahl. Es muss ein gültiges Datum angegeben werden, damit die Berechnung ordnungsgemäß erfolgen kann. Beispielsweise sollten Sie nicht den 29. Februar 2001 als Datum angeben. Da 2001 kein Schaltjahr war, gibt es dieses Datum nicht.

Funktion	Ergebnis	Beschreibung
date_in_years(DATE)	Reelle Zahl	Ergibt die Zeit in Jahren (als reelle Zahl), die zwischen dem Basisdatum und dem durch <i>DATUM</i> angegebenen Datum oder Zeitstempel liegt. Dies ist ein Näherungswert, der auf einem Jahr mit 365.25 Tagen beruht. Wenn <i>DATUM</i> vor dem Basisdatum liegt, ergibt diese Funktion eine negative Zahl. Es muss ein gültiges Datum angegeben werden, damit die Berechnung ordnungsgemäß erfolgen kann. Beispielsweise sollten Sie nicht den 29. Februar 2001 als Datum angeben. Da 2001 kein Schaltjahr war, gibt es dieses Datum nicht.
date_months_difference(DATE1, DATE2)	Reelle Zahl	Ergibt die Zeit in Monaten (als reelle Zahl), die zwischen dem durch <i>DATUM1</i> und dem durch <i>DATUM2</i> angegebenen Datum bzw. Zeitstempel liegt. Dies ist ein Näherungswert, der auf einem Monat mit 30.4375 Tagen beruht. Wenn <i>DATUM2</i> vor <i>DATUM1</i> liegt, ergibt diese Funktion eine negative Zahl.
datetime_date(YEAR, MONTH, DAY)	Datum	Erstellt einen Datumswert für die angegeben Wert für <i>JAHR</i> , <i>MONAT</i> und <i>TAG</i> . Bei den Argumenten muss es sich um ganze Zahlen handeln.
datetime_day(DATE)	Ganzzahl	Ergibt den Tag im Monat ausgehend von einem angegebenen <i>DATUM</i> oder Zeitstempel. Das Ergebnis ist eine Ganzzahl im Bereich von 1 bis 31.
datetime_day_name(DAY)	Zeichenfolge	Ergibt den vollständigen Namen für den angegeben <i>TAG</i> . Das Argument muss eine ganze Zahl im Bereich von 1 (Sonntag) bis 7 (Samstag) sein.
datetime_hour(TIME)	Ganzzahl	Ergibt die Stunde aus einer <i>ZEIT</i> oder einem Zeitstempel. Das Ergebnis ist eine Ganzzahl im Bereich von 0 bis 23.
datetime_in_seconds(TIME)	Reelle Zahl	Ergibt den in <i>ZEIT</i> gespeicherten Sekundenabschnitt.
datetime_in_seconds(DATE), datetime_in_seconds(DATE-TIME)	Reelle Zahl	Ergibt die aus der Differenz zwischen dem aktuellen <i>DATUM</i> oder <i>DATUM/UHRZEIT</i> und dem Basisdatum (1900-01-01) berechneten Gesamtwert (umgewandelt in Sekunden).
datetime_minute(TIME)	Ganzzahl	Ergibt die Minute aus einer <i>ZEIT</i> oder einem Zeitstempel. Das Ergebnis ist eine Ganzzahl im Bereich von 0 bis 59.
datetime_month(DATE)	Ganzzahl	Ergibt den Monat aus einem <i>DATUM</i> oder einem Zeitstempel. Das Ergebnis ist eine Ganzzahl im Bereich von 1 bis 12.
datetime_month_name(MONTH)	Zeichenfolge	Ergibt den vollständigen Namen für den angegeben <i>MONAT</i> . Das Argument muss eine ganze Zahl im Bereich von 1 bis 12 sein.
datetime_now	Zeitstempel	Ergibt die aktuelle Zeit als Zeitstempel.
datetime_second(TIME)	Ganzzahl	Ergibt die Sekunde aus einer <i>ZEIT</i> oder einem Zeitstempel. Das Ergebnis ist eine Ganzzahl im Bereich von 0 bis 59.
datetime_day_short_name(DAY)	Zeichenfolge	Ergibt den abgekürzten Namen für den angegeben <i>TAG</i> . Das Argument muss eine ganze Zahl im Bereich von 1 (Sonntag) bis 7 (Samstag) sein.
datetime_month_short_name(MONTH)	Zeichenfolge	Ergibt den abgekürzten Namen für den angegeben <i>MONAT</i> . Das Argument muss eine ganze Zahl im Bereich von 1 bis 12 sein.

Funktion	Ergebnis	Beschreibung
<code>datetime_time(HOUR, MINUTE, SECOND)</code>	<i>Zeit</i>	Ergibt den Zeitwert für die angegebene <i>STUNDE</i> , <i>MINUTE</i> und <i>SEKUNDE</i> . Bei den Argumenten muss es sich um ganze Zahlen handeln.
<code>datetime_time(ITEM)</code>	<i>Zeit</i>	Ergibt den Zeitwert für das angegebene <i>ELEMENT</i> .
<code>datetime_timestamp(YEAR, MONTH, DAY, HOUR, MINUTE, SECOND)</code>	<i>Zeitstempel</i>	Ergibt den Zeitstempelwert für die angegeben Wert für <i>JAHR</i> , <i>MONAT</i> , <i>TAG</i> , <i>STUNDE</i> , <i>MINUTE</i> und <i>SEKUNDE</i> .
<code>datetime_timestamp(DATE, TIME)</code>	<i>Zeitstempel</i>	Ergibt den Zeitstempelwert für das angegebene <i>DATUM</i> und die angegebene <i>ZEIT</i> .
<code>datetime_timestamp (NUMBER)</code>	<i>Zeitstempel</i>	Gibt den Zeitstempelwert für die angegebene Anzahl von Sekunden zurück.
<code>datetime_weekday(DATE)</code>	<i>Ganzzahl</i>	Ergibt den Tag in der Woche t ausgehend von einem angegebenen <i>DATUM</i> oder Zeitstempel.
<code>datetime_year(DATE)</code>	<i>Ganzzahl</i>	Ergibt das Jahr aus einem <i>DATUM</i> oder einem Zeitstempel. Das Ergebnis ist eine ganze Zahl, wie beispielsweise 2002.
<code>date_weeks_difference (DATE1, DATE2)</code>	<i>Reelle Zahl</i>	Ergibt die Zeit in Wochen (als reelle Zahl), die zwischen dem durch <i>DATUM1</i> und dem durch <i>DATUM2</i> angegebenen Datum bzw. Zeitstempel liegt. Der Wert basiert auf einer Woche mit 7,0 Tagen. Wenn <i>DATUM2</i> vor <i>DATUM1</i> liegt, ergibt diese Funktion eine negative Zahl.
<code>date_years_difference (DATE1, DATE2)</code>	<i>Reelle Zahl</i>	Ergibt die Zeit in Jahren (als reelle Zahl), die zwischen dem durch <i>DATUM1</i> und dem durch <i>DATUM2</i> angegebenen Datum bzw. Zeitstempel liegt. Dies ist ein Näherungswert, der auf einem Jahr mit 365,25 Tagen beruht. Wenn <i>DATUM2</i> vor <i>DATUM1</i> liegt, ergibt diese Funktion eine negative Zahl.
<code>time_before(TIME1, TIME2)</code>	<i>Boolesch</i>	Ergibt den Wert "Wahr", wenn <i>ZEIT1</i> eine Uhrzeit oder einen Zeitstempel darstellt, die/der vor der durch <i>ZEIT2</i> dargestellten Zeit liegt. Andernfalls ergibt die Funktion den Wert 0.
<code>time_hours_difference (TIME1, TIME2)</code>	<i>Reelle Zahl</i>	Ergibt den Zeitunterschied in Stunden (als reelle Zahl) zwischen den durch <i>ZEIT1</i> und <i>ZEIT2</i> angegebenen Uhrzeiten oder Zeitstempeln. Bei Auswahl von Tage/Minuten übertragen im Dialogfeld "Stream-Eigenschaften" wird ein höherer Wert von <i>ZEIT1</i> als Verweis auf den vorhergehenden Tag interpretiert. Wenn die Übertragungsoption nicht ausgewählt wird, führt ein höherer Wert von <i>ZEIT1</i> dazu, dass der ausgegebene Wert negativ ist.
<code>time_in_hours(TIME)</code>	<i>Reelle Zahl</i>	Ergibt die durch <i>ZEIT</i> angegebene Zeit in Stunden als reelle Zahl. So wird beispielsweise beim Zeitformat HHMM der Ausdruck <code>time_in_hours('0130')</code> als 1,5 ausgewertet. <i>ZEIT</i> kann für eine Uhrzeit oder einen Zeitstempel stehen.
<code>time_in_mins(TIME)</code>	<i>Reelle Zahl</i>	Ergibt die durch <i>ZEIT</i> angegebene Zeit in Minuten als reelle Zahl. <i>ZEIT</i> kann für eine Uhrzeit oder einen Zeitstempel stehen.
<code>time_in_secs(TIME)</code>	<i>Ganzzahl</i>	Ergibt die durch <i>ZEIT</i> angegebene Zeit in Sekunden als ganze Zahl. <i>ZEIT</i> kann für eine Uhrzeit oder einen Zeitstempel stehen.

Funktion	Ergebnis	Beschreibung
time_mins_difference(TIME1, TIME2)	Reelle Zahl	Ergibt den Zeitunterschied in Minuten (als reelle Zahl) zwischen den durch ZEIT1 und ZEIT2 angegebenen Uhrzeiten oder Zeitstempeln. Bei Auswahl von Tage/Minuten übertragen im Dialogfeld "Stream-Eigenschaften" wird ein höherer Wert von ZEIT1 als Verweis auf den vorhergehenden Tag (bzw. die vorangegangene Stunde, sofern nur Minuten und Sekunden im aktuellen Format angegeben werden) interpretiert. Wenn die Übertragungsoption nicht ausgewählt wird, führt ein höherer Wert von ZEIT1 dazu, dass der ausgegebene Wert negativ ist.
time_secs_difference(TIME1, TIME2)	Ganzzahl	Ergibt den Zeitunterschied in Sekunden (als ganze Zahl) zwischen den durch ZEIT1 und ZEIT2 angegebenen Uhrzeiten oder Zeitstempeln. Bei Auswahl von Tage/Minuten übertragen im Dialogfeld "Stream-Eigenschaften" wird ein höherer Wert von ZEIT1 als Verweis auf den vorhergehenden Tag (bzw. die vorangegangene Stunde, sofern nur Minuten und Sekunden im aktuellen Format angegeben werden) interpretiert. Wenn die Übertragungsoption nicht ausgewählt wird, führt ein höherer Wert von ZEIT1 dazu, dass der ausgegebene Wert negativ ist.

Konvertieren von Datums- und Zeitwerten.

Beachten Sie, dass die Konvertierungsfunktionen (und alle anderen Funktionen, für die ein spezieller Eingabetyp, wie beispielsweise ein Wert für Datum oder Uhrzeit, erforderlich ist) von den aktuell im Dialogfeld für die Stream-Optionen angegebenen Formaten abhängen. Ein Feld mit der Bezeichnung DATUM beispielsweise, das als Zeichenkette mit den Werten *Jan 2003, Feb 2003* usw. gespeichert ist, könnte wie folgt in einen Datumsspeicher konvertiert werden:

`to_date(DATUM)`

Damit diese Konvertierung funktionieren kann, müssen Sie das entsprechende Datumsformat MON JJJJ als Standarddatumsformat für den Stream festlegen. Für weitere Informationen siehe [Thema Festlegen von allgemeinen Optionen für Streams in Kapitel 5](#) auf S. 63.

Ein Beispiel, bei dem Zeichenkettenwerte mithilfe eines Füllerknotens in Datumswerte konvertiert werden, finden Sie im Stream *broadband_create_models.str*, der im Ordner *\Demos* unter dem Unterordner *streams* installiert ist. Für weitere Informationen siehe [Thema Prognoseerstellung mit dem Zeitreihenknoten in Kapitel 14 in IBM SPSS Modeler 15- Benutzerhandbuch](#).

Als Zahlen gespeicherte Datumsangaben. Beachten Sie, dass DATUM im vorherigen Beispiel der Name eines Felds ist, während es sich bei `to_date` um eine CLEM-Funktion handelt. Wenn Datumsangaben als Zahlen gespeichert wurden, können Sie sie mithilfe der Funktion `datetime_date` konvertieren. Dabei wird die Zahl als eine Anzahl von Sekunden seit dem Basisdatum (Epochen) interpretiert.

`datetime_date(DATUM)`

Indem Sie ein Datum in eine Sekundenzahl (und zurück) konvertieren, können Sie Berechnungen wie die Addition oder Subtraktion einer Anzahl von Tagen zum bzw. vom aktuellen Datum durchführen. Beispiel:

```
datetime_date((date_in_days(DATUM)-7)*60*60*24)
```

Sequenzfunktionen

Bei einigen Operationen ist die Abfolge (Sequenz) der Ereignisse von Bedeutung. Die Anwendung ermöglicht die Arbeit mit folgenden Datensatzsequenzen:

- Sequenzen und Zeitreihen
- Sequenzfunktionen
- Datensatzindizierung
- Durchschnittsbildung, Summierung und Vergleich von Werten
- Überwachen von Änderungen – Differenzierung
- @SINCE
- Offest-Werte
- Zusätzliche Sequenzfunktionen

Bei vielen Anwendungen kann jeder Datensatz, der einen Stream durchläuft, als Einzelfall betrachtet werden, der von allen anderen Fällen unabhängig ist. In solchen Situationen ist die Reihenfolge der Datensätze normalerweise nicht von Bedeutung.

Bei einigen Problemklassen jedoch ist die Datensatzsequenz höchst wichtig. Dies ist üblicherweise bei Zeitreihen der Fall, bei denen die Sequenz der Datensätze eine geordnete Sequenz von Ereignissen oder Vorkommen darstellt. Jeder Datensatz stellt einen "Schnappschuss" zu einem bestimmten Zeitpunkt dar; ein Großteil der aussagekräftigsten Informationen sind jedoch möglicherweise nicht Augenblickswerten enthalten, sondern in der Art und Weise, wie sich diese Werte im Laufe der Zeit ändern und verhalten.

Der entscheidende Parameter muss natürlich nicht die Zeit sein. So könnten die Datensätze Analysen darstellen, die in Abständen entlang einer Linie durchgeführt werden. Doch auch hier gelten dieselben Prinzipien.

Sequenz- und Sonderfunktionen lassen sich sofort an folgenden Merkmalen erkennen:

- Sie enthalten alle das Präfix @.
- Die Namen dieser Funktionen in Großbuchstaben.

Sequenzfunktionen können sich auf den derzeit von einem Knoten verarbeiteten Datensatz, auf die Datensätze, die einen Knoten bereits durchlaufen haben, und sogar, in einem Fall, auf Datensätze beziehen, die einen Knoten noch durchlaufen müssen. Sequenzfunktionen können nach Belieben mit anderen Komponenten von CLEM-Ausdrücken gemischt werden; bei einigen gelten jedoch Einschränkungen in Bezug auf die verwendeten Argumente.

Beispiele

Es kann interessant sein zu wissen, wie lange es zurückliegt, dass ein bestimmtes Ereignis eintrat oder eine Bedingung wahr war. Dies kann mit der Funktion @SINCE ermittelt werden. Beispiel:

```
@SINCE(Einkommen > Ausgaben)
```

Diese Funktion ergibt das Offset des letzten Datensatzes, bei dem diese Bedingung erfüllt war – also wie viele Datensätze vor dem aktuellen Datensatz die Bedingung wahr war. Wenn die Bedingung niemals wahr war, gibt @SINCE den Wert @INDEX + 1 zurück.

Manchmal möchte man sich in dem von @SINCE verwendeten Ausdruck auf einen Wert des aktuellen Datensatzes beziehen. Dies ist mit der Funktion @THIS möglich, die angibt, dass ein Feldname immer für den aktuellen Datensatz gilt. Um das Offset des letzten Datensatzes zu ermitteln, bei dem der Wert des Felds Concentration mehr als zweimal so hoch war wie der des aktuellen Datensatzes, können Sie folgende Funktion verwenden:

```
@SINCE(Konzentration > 2 * @THIS(Konzentration))
```

In einigen Fällen ist die für @SINCE angegebene Bedingung beim aktuellen Datensatz per definitionem wahr. Beispiel:

```
@SINCE(ID == @THIS(ID))
```

Aus diesem Grund wertet @SINCE die Bedingung nicht für den aktuellen Datensatz aus. Mit einer ähnlichen Funktion, @SINCE0, können Sie die Bedingung für den aktuellen Datensatz und die vorangegangenen auswerten; wenn die Bedingung im aktuellen Datensatz wahr ist, ergibt @SINCE0 den Wert 0.

Hinweis: Die @-Funktionen können nicht aus Skripts heraus aufgerufen werden. Für weitere Informationen siehe Thema [CLEM-Ausdrücke in Skripts in Kapitel 3 in IBM SPSS Modeler 15 Handbuch für Skripterstellung und Automatisierung](#).

Funktion	Ergebnis	Beschreibung
MEAN(FIELD)	Reelle Zahl	Ergibt den mittleren Durchschnitt der Werte für das angegebene FELD bzw. die angegebenen FELDER.
@MEAN(FIELD, EXPR)	Reelle Zahl	Ergibt den mittleren Durchschnitt der Werte für FELD in den letzten AUSDR Datensätzen, die vom aktuellen Knoten erhalten wurden, einschließlich des aktuellen Datensatzes. FELD muss der Name eines numerischen Felds sein. AUSDR kann jeder Ausdruck sein, dessen Auswertung eine ganze Zahl größer als 0 ergibt. Wenn AUSDR weggelassen wird oder die Anzahl der bisher erhaltenen Datensätze übersteigt, ergibt sich der Durchschnitt aus den bisher erhaltenen Datensätzen. Beachten Sie, dass diese Funktion nicht über ein Skript aufgerufen werden kann. Für weitere Informationen siehe Thema CLEM-Ausdrücke in Skripts in Kapitel 3 in IBM SPSS Modeler 15 Handbuch für Skripterstellung und Automatisierung .

Funktion	Ergebnis	Beschreibung
@MEAN(FIELD, EXPR, INT)	Reelle Zahl	Ergibt den mittleren Durchschnitt der Werte für <i>FELD</i> in den letzten <i>AUSDR</i> Datensätzen, die vom aktuellen Knoten erhalten wurden, einschließlich des aktuellen Datensatzes. <i>FELD</i> muss der Name eines numerischen Felds sein. <i>AUSDR</i> kann jeder Ausdruck sein, dessen Auswertung eine ganze Zahl größer als 0 ergibt. Wenn <i>AUSDR</i> weggelassen wird oder die Anzahl der bisher erhaltenen Datensätze übersteigt, ergibt sich der Durchschnitt aus den bisher erhaltenen Datensätzen. <i>GANZZ</i> gibt die maximale Anzahl von zurückliegenden Werten an, die betrachtet werden sollen. Dies ist wesentlich effizienter als die Verwendung von nur zwei Argumenten.
@DIFF1(FIELD)	Reelle Zahl	Ergibt das erste Differenzial von <i>FELD1</i> . Das Format mit einem einzelnen Argument gibt also einfach die Differenz zwischen dem aktuellen Wert und dem früheren Wert des Felds zurück. Ergibt 0, wenn die relevanten vorhergehenden Datensätze nicht vorhanden sind.
@DIFF1(FIELD1, FIELD2)	Reelle Zahl	Das Format mit einem zwei Argumenten ergibt das erste Differenzial von <i>FELD1</i> in Bezug auf <i>FELD2</i> . Ergibt 0, wenn die relevanten vorhergehenden Datensätze nicht vorhanden sind.
@DIFF2(FIELD)	Reelle Zahl	Ergibt das zweite Differenzial von <i>FELD1</i> . Das Format mit einem einzelnen Argument gibt also einfach die Differenz zwischen dem aktuellen Wert und dem früheren Wert des Felds zurück. Ergibt 0, wenn die relevanten vorhergehenden Datensätze nicht vorhanden sind.
@DIFF2(FIELD1, FIELD2)	Reelle Zahl	Das Format mit einem zwei Argumenten ergibt das erste Differenzial von <i>FELD1</i> in Bezug auf <i>FELD2</i> . Ergibt 0, wenn die relevanten vorhergehenden Datensätze nicht vorhanden sind.
@INDEX	Ganzzahl	Ergibt den Index des aktuellen Datensatzes. Indizes werden den Datensätzen zugewiesen, wenn sie am aktuellen Knoten ankommen. Der erste Datensatz erhält den Index 1. Dieser Index wird für jeden nachfolgenden Datensatz um den Wert 1 erhöht.
@LAST_NON_BLANK(FIELD)	Jede	Ergibt den letzten nichtleeren Wert für <i>FELD</i> , wie in einem aufwärts liegenden Quellen- oder Typknoten definiert. Wenn in den bisher gelesenen Datensätzen keine nichtleeren Werte für <i>FELD</i> vorhanden sind, wird \$null\$ zurückgegeben. Leerwerte (auch als benutzerdefiniert fehlende Werte bezeichnet) können für jedes Feld einzeln definiert werden.
@MAX(FIELD)	Number	Ergibt den Höchstwert für das angegebene <i>FELD</i> .
@MAX(FIELD, EXPR)	Number	Ergibt den Höchstwert für <i>FELD</i> in den letzten <i>AUSDR</i> Datensätzen, die bisher erhalten wurden, einschließlich des aktuellen Datensatzes. <i>FELD</i> muss der Name eines numerischen Felds sein. <i>AUSDR</i> kann jeder Ausdruck sein, dessen Auswertung eine ganze Zahl größer als 0 ergibt.

Funktion	Ergebnis	Beschreibung
<code>@MAX(FIELD, EXPR, INT)</code>	<i>Number</i>	Ergibt den Höchstwert für <i>FELD</i> in den letzten <i>AUSDR</i> Datensätzen, die bisher erhalten wurden, einschließlich des aktuellen Datensatzes. <i>FELD</i> muss der Name eines numerischen Felds sein. <i>AUSDR</i> kann jeder Ausdruck sein, dessen Auswertung eine ganze Zahl größer als 0 ergibt. Wenn <i>AUSDR</i> weggelassen wird oder die Anzahl der bisher erhaltenen Datensätze übersteigt, ergibt sich der Höchstwert aus den bisher erhaltenen Datensätzen. <i>GANZZ</i> gibt die maximale Anzahl von zurückliegenden Werten an, die betrachtet werden sollen. Dies ist wesentlich effizienter als die Verwendung von nur zwei Argumenten.
<code>@MIN(FIELD)</code>	<i>Number</i>	Ergibt den Mindestwert für das angegebene <i>FELD</i> .
<code>@MIN(FIELD, EXPR)</code>	<i>Number</i>	Ergibt den Mindestwert für <i>FELD</i> in den letzten <i>AUSDR</i> Datensätzen, die bisher erhalten wurden, einschließlich des aktuellen Datensatzes. <i>FELD</i> muss der Name eines numerischen Felds sein. <i>AUSDR</i> kann jeder Ausdruck sein, dessen Auswertung eine ganze Zahl größer als 0 ergibt.
<code>@MIN(FIELD, EXPR, INT)</code>	<i>Number</i>	Ergibt den Mindestwert für <i>FELD</i> in den letzten <i>AUSDR</i> Datensätzen, die bisher erhalten wurden, einschließlich des aktuellen Datensatzes. <i>FELD</i> muss der Name eines numerischen Felds sein. <i>AUSDR</i> kann jeder Ausdruck sein, dessen Auswertung eine ganze Zahl größer als 0 ergibt. Wenn <i>AUSDR</i> weggelassen wird oder die Anzahl der bisher erhaltenen Datensätze übersteigt, ergibt sich der Mindestwert aus den bisher erhaltenen Datensätzen. <i>GANZZ</i> gibt die maximale Anzahl von zurückliegenden Werten an, die betrachtet werden sollen. Dies ist wesentlich effizienter als die Verwendung von nur zwei Argumenten.
<code>@OFFSET(FIELD, EXPR)</code>	<i>Jede</i>	Ergibt den Wert von <i>FELD</i> in dem Datensatz, der gegenüber dem aktuellen Datensatz ein Offset mit dem Wert von <i>AUSDR</i> aufweist. Ein positives Offset bezieht sich auf einen Datensatz, der den Knoten bereits durchlaufen hat, ein negatives Offset gibt eine "Vorschau" auf einen Datensatz an, der noch nicht angekommen ist. Beispiel: <code>@OFFSET(Status, 1)</code> ergibt den Wert des Felds Status im vorangegangenen Datensatz, während <code>@OFFSET(Status, -4)</code> vier Datensätze nach vorn in der Sequenz blickt (also auf Datensätze, die den Knoten noch nicht durchlaufen haben), um den Wert zu ermitteln. <i>Beachten Sie, dass ein negatives Offset (Vorschau-Offset) als Konstante angegeben sein muss.</i> Bei positiven Offsets gilt Folgendes: Bei <i>AUSDR</i> kann es sich auch um einen beliebigen CLEM-Ausdruck handeln, der für den aktuellen Datensatz ausgewertet wird, um das Offset zu erhalten. In diesem Fall sollte die Version dieser Funktion mit drei Argumenten die Leistung verbessern (siehe nächste Funktion). Wenn der Ausdruck einen anderen Wert als eine nichtnegative ganze Zahl ergibt, wird ein Fehler ausgelöst. Berechnete Vorschau-Offsets sind also nicht zulässig. <i>Hinweis:</i> Eine selbstreferenzielle <code>@OFFSET</code> -Funktion kann keine Literalvorschau verwenden. In einem Füllerknoten beispielsweise kann der Wert von <i>field1</i>

Funktion	Ergebnis	Beschreibung
		nicht mithilfe eines Ausdrucks wie <code>@OFFSET(field1,-2)</code> ersetzt werden.
<code>@OFFSET(FIELD, EXPR, INT)</code>	Jede	Führt die gleiche Operation durch wie <code>@OFFSET</code> , allerdings mit einem zusätzlichen dritten Argument, <i>GANZZ</i> , das die maximale Anzahl der zurückliegenden Werte angibt, die betrachtet werden sollen. Wenn das Offset aus einem Ausdruck berechnet wird, sollte das dritte Argument zu einer Verbesserung der Leistung führen. In einem Ausdruck wie <code>@OFFSET(Foo, Month, 12)</code> beispielsweise wird das System angewiesen, nur die letzten 12 Werte von <i>Foo</i> zu behalten. Andernfalls müsste sicherheitshalber jeder Wert gespeichert werden. In Fällen, in denen es sich beim Offset-Wert um eine Konstante handelt – einschließlich der negativen "Vorschau"-Offsets, die konstant sein müssen – ist das dritte Argument sinnlos, weshalb hier die Version dieser Funktion mit zwei Argumenten verwendet werden sollte. Siehe auch den Hinweis zu selbstreferenziellen Funktionen für die Version mit zwei Argumenten weiter oben.
<code>@SDEV(FIELD)</code>	Reelle Zahl	Ergibt die Standardabweichung der Werte für das angegebene <i>FELD</i> bzw. die angegebenen <i>FELDER</i> .
<code>@SDEV(FIELD, EXPR)</code>	Reelle Zahl	Gibt die Standardabweichung der Werte für <i>FELD</i> in den letzten <i>AUSDR</i> Datensätzen zurück, die vom aktuellen Knoten erhalten wurden, einschließlich des aktuellen Datensatzes. <i>FELD</i> muss der Name eines numerischen Felds sein. <i>AUSDR</i> kann jeder Ausdruck sein, dessen Auswertung eine ganze Zahl größer als 0 ergibt. Wenn <i>AUSDR</i> weggelassen wird oder die Anzahl der bisher erhaltenen Datensätze übersteigt, ergibt sich die Standardabweichung aus allen bisher erhaltenen Datensätzen.
<code>@SDEV(FIELD, EXPR, INT)</code>	Reelle Zahl	Gibt die Standardabweichung der Werte für <i>FELD</i> in den letzten <i>AUSDR</i> Datensätzen zurück, die vom aktuellen Knoten erhalten wurden, einschließlich des aktuellen Datensatzes. <i>FELD</i> muss der Name eines numerischen Felds sein. <i>AUSDR</i> kann jeder Ausdruck sein, dessen Auswertung eine ganze Zahl größer als 0 ergibt. Wenn <i>AUSDR</i> weggelassen wird oder die Anzahl der bisher erhaltenen Datensätze übersteigt, ergibt sich die Standardabweichung aus allen bisher erhaltenen Datensätzen. <i>GANZZ</i> gibt die maximale Anzahl von zurückliegenden Werten an, die betrachtet werden sollen. Dies ist wesentlich effizienter als die Verwendung von nur zwei Argumenten.
<code>@SINCE(EXPR)</code>	Jede	Gibt an, wie viele Datensätze den Knoten durchlaufen haben, seit <i>AUSDR</i> , ein beliebiger CLEM-Ausdruck, wahr war.
<code>@SINCE(EXPR, INT)</code>	Jede	Das zweite Argument, <i>GANZZ</i> , gibt die maximale Anzahl von zurückliegenden Werten an, die betrachtet werden sollen. Wenn <i>AUSDR</i> niemals wahr war, ist <i>GANZZ</i> gleich <code>@INDEX+1</code> .
<code>@SINCE0(EXPR)</code>	Jede	Berücksichtigt den aktuellen Datensatz, was bei <code>@SINCE</code> nicht der Fall ist; <code>@SINCE0</code> ergibt 0, wenn <i>AUSDR</i> für den aktuellen Datensatz wahr ist.

Funktion	Ergebnis	Beschreibung
@SINCE0(EXPR, INT)	Jede	Das zweite Argument, <i>GANZZ</i> , gibt die maximale Anzahl von zurückliegenden Werten an, die betrachtet werden sollen.
@SUM(FIELD)	Number	Ergibt die Summe der Werte für das angegebene <i>FELD</i> bzw. die angegeben <i>FELDER</i> .
@SUM(FIELD, EXPR)	Number	Ergibt die Summe der Werte für <i>FELD</i> in den letzten <i>AUSDR</i> Datensätzen, die vom aktuellen Knoten erhalten wurden, einschließlich des aktuellen Datensatzes. <i>FELD</i> muss der Name eines numerischen Felds sein. <i>AUSDR</i> kann jeder Ausdruck sein, dessen Auswertung eine ganze Zahl größer als 0 ergibt. Wenn <i>AUSDR</i> weggelassen wird oder die Anzahl der bisher erhaltenen Datensätze übersteigt, ergibt sich die Summe aus den bisher erhaltenen Datensätzen.
@SUM(FIELD, EXPR, INT)	Number	Ergibt die Summe der Werte für <i>FELD</i> in den letzten <i>AUSDR</i> Datensätzen, die vom aktuellen Knoten erhalten wurden, einschließlich des aktuellen Datensatzes. <i>FELD</i> muss der Name eines numerischen Felds sein. <i>AUSDR</i> kann jeder Ausdruck sein, dessen Auswertung eine ganze Zahl größer als 0 ergibt. Wenn <i>AUSDR</i> weggelassen wird oder die Anzahl der bisher erhaltenen Datensätze übersteigt, ergibt sich die Summe aus den bisher erhaltenen Datensätzen. <i>GANZZ</i> gibt die maximale Anzahl von zurückliegenden Werten an, die betrachtet werden sollen. Dies ist wesentlich effizienter als die Verwendung von nur zwei Argumenten.
@THIS(FIELD)	Jede	Ergibt den Wert des Felds namens <i>FELD</i> im aktuellen Datensatz. Wird nur bei @SINCE-Ausdrücken verwendet.

Globale Funktionen

Die Funktionen @MEAN, @SUM, @MIN, @MAX und @SDEV verarbeiten maximal alle gelesenen Datensätzen bis einschließlich dem aktuellen. In einigen Fällen ist es jedoch von Vorteil, in der Lage zu sein, die Daten im aktuellen Datensatz mit Werten im gesamten Daten-Set zu vergleichen. Mit einem Globalwerteknoten können Sie Werte im gesamten Daten-Set generieren, auf die Sie dann mit den globalen Funktionen in einem CLEM-Ausdruck zugreifen können.

Beispiel:

`@GLOBAL_MAX(Alter)`

ergibt den höchsten Wert von *Age* im Daten-Set, während der Ausdruck

`(Wert - @GLOBAL_MEAN(Wert)) / @GLOBAL_SDEV(Wert)`

die Differenz zwischen dem *Value* dieses Datensatzes und dem globalen Mittelwert als Anzahl der Standardabweichungen angibt. Globale Werte können nur verwendet werden, nachdem sie von einem Globalwerteknoten berechnet wurden. Alle aktuellen Globalwerte können durch Klicken auf die Schaltfläche zum Löschen der Globalwerte auf der Registerkarte "Globalwerte" im Dialogfeld "Stream-Eigenschaften" entfernt werden.

Hinweis: Die @-Funktionen können nicht aus Skripts heraus aufgerufen werden. Für weitere Informationen siehe Thema CLEM-Ausdrücke in Skripts in Kapitel 3 in IBM SPSS Modeler 15 Handbuch für Skripterstellung und Automatisierung.

Funktion	Ergebnis	Beschreibung
@GLOBAL_MAX(FIELD)	Number	Ergibt den Höchstwert für <i>FELD</i> im gesamten Daten-Set, das zuvor durch einen Globalwerteknoten erstellt wurde. <i>FELD</i> muss der Name eines numerischen Felds sein. Wenn der entsprechende Globalwert nicht festgelegt wurde, wird ein Fehler ausgegeben. Beachten Sie, dass diese Funktion nicht über ein Skript aufgerufen werden kann. Für weitere Informationen siehe Thema CLEM-Ausdrücke in Skripts in Kapitel 3 in IBM SPSS Modeler 15 Handbuch für Skripterstellung und Automatisierung.
@GLOBAL_MIN(FIELD)	Number	Ergibt den Höchstwert für <i>FELD</i> im gesamten Daten-Set, das zuvor durch einen Globalwerteknoten generiert wurde. <i>FELD</i> muss der Name eines numerischen Felds sein. Wenn der entsprechende Globalwert nicht festgelegt wurde, wird ein Fehler ausgegeben.
@GLOBAL_SDEV(FIELD)	Number	Ergibt die Standardabweichung der Werte für <i>FELD</i> im gesamten Daten-Set, das zuvor durch einen Globalwerteknoten generiert wurde. <i>FELD</i> muss der Name eines numerischen Felds sein. Wenn der entsprechende Globalwert nicht festgelegt wurde, wird ein Fehler ausgegeben.
@GLOBAL_MEAN(FIELD)	Number	Ergibt den mittleren Durchschnitt der Werte für <i>FELD</i> im gesamten Daten-Set, das zuvor durch einen Globalwerteknoten generiert wurde. <i>FELD</i> muss der Name eines numerischen Felds sein. Wenn der entsprechende Globalwert nicht festgelegt wurde, wird ein Fehler ausgegeben.
@GLOBAL_SUM(FIELD)	Number	Ergibt die Summe der Werte für <i>FELD</i> im gesamten Daten-Set, das zuvor durch einen Globalwerteknoten generiert wurde. <i>FELD</i> muss der Name eines numerischen Felds sein. Wenn der entsprechende Globalwert nicht festgelegt wurde, wird ein Fehler ausgegeben.

Funktionen zum Umgang mit Leerstellen und Nullwerten

Mit CLEM können Sie angeben, dass bestimmte Werte in einem Feld als “Leerstellen” oder fehlende Werte betrachtet werden sollen. Die nachstehenden Funktionen können auf Leerstellen angewendet werden.

Hinweis: Die @-Funktionen können nicht aus Skripts heraus aufgerufen werden. Für weitere Informationen siehe Thema CLEM-Ausdrücke in Skripts in Kapitel 3 in IBM SPSS Modeler 15 Handbuch für Skripterstellung und Automatisierung.

Funktion	Ergebnis	Beschreibung
@BLANK(FIELD)	Boolesch	Ergibt den Wert "Wahr" für alle Datensätze, deren Werte gemäß den Regeln zum Umgang mit Leerstellen, die in einem weiter oben im Stream gelegenen Typknoten oder Quellenknoten (Registerkarte "Typen") festgelegt wurden, Leerstellen sind. Beachten Sie, dass diese Funktion nicht über ein Skript aufgerufen werden kann. Für weitere Informationen siehe Thema CLEM-Ausdrücke in Skripts in Kapitel 3 in IBM SPSS Modeler 15 Handbuch für Skripterstellung und Automatisierung.
@LAST_NON_BLANK(FIELD)	Jede	Ergibt den letzten nichtleeren Wert für <i>FELD</i> , wie in einem aufwärts liegenden Quellen- oder Typknoten definiert. Wenn in den bisher gelesenen Datensätzen keine nichtleeren Werte für <i>FELD</i> vorhanden sind, wird \$null\$ zurückgegeben. Leerwerte (auch als benutzerdefiniert fehlende Werte bezeichnet) können für jedes Feld einzeln definiert werden.
@NULL(FIELD)	Boolesch	Ergibt den Wert "Wahr", wenn der Wert von <i>FELD</i> der systemdefiniert fehlende Wert \$null\$ ist. Ergibt den Wert "Falsch" für alle anderen Werte, einschließlich benutzerdefinierten fehlenden Werten. Wenn Sie nach beiden suchen möchten, verwenden Sie @BLANK(FIELD) und @NULL(FIELD).
undef	Jede	Wird allgemein in CLEM zur Eingabe eines \$null\$-Werts verwendet – beispielsweise, um Leerwerte im Füllerknoten mit Nullen aufzufüllen.

Mithilfe des Füllerknotens können leere Felder "aufgefüllt" werden. Bei Füller- und Ableitungsknoten (nur beim Mehrfachmodus) bezieht sich die CLEM-Sonderfunktion @FIELD auf die aktuell untersuchten Felder.

Sonderfelder

Sonderfunktionen dienen zur Kennzeichnung der speziellen untersuchten Felder oder zum Erzeugen einer Felderliste als Eingabe. Wenn Sie beispielsweise mehrere Felder gleichzeitig ableiten, sollten Sie mit @FIELD angeben, dass diese Ableitungsaktion für die ausgewählten Felder durchgeführt werden soll. Mit dem Ausdruck log(@FIELD) wird ein neues Log-Feld für jedes ausgewählte Feld abgeleitet.

Hinweis: Die @-Funktionen können nicht aus Skripts heraus aufgerufen werden. Für weitere Informationen siehe Thema CLEM-Ausdrücke in Skripts in Kapitel 3 in IBM SPSS Modeler 15 Handbuch für Skripterstellung und Automatisierung.

Funktion	Ergebnis	Beschreibung
@FIELD	Jede	Führt eine Aktion für alle im Ausdruckskontext angegebenen Felder durch. Beachten Sie, dass diese Funktion nicht über ein Skript aufgerufen werden kann. Für weitere Informationen siehe Thema CLEM-Ausdrücke in Skripts in Kapitel 3 in IBM SPSS Modeler 15 Handbuch für Skripterstellung und Automatisierung.
@TARGET	Jede	Wenn ein CLEM-Ausdruck in einer benutzerdefinierten Analysefunktion verwendet wird, steht @TARGET für das Zielfeld bzw. den “korrekten Wert” für das analysierte Paar aus Zielwert und vorhergesagtem Wert. Diese Funktion wird häufig in Analyseknoten verwendet.
@PREDICTED	Jede	Wenn ein CLEM-Ausdruck in einer benutzerdefinierten Analysefunktion verwendet wird, steht @PREDICTED für den vorhergesagten Wert des analysierten Paares aus Zielwert und vorhergesagtem Wert. Diese Funktion wird häufig in Analyseknoten verwendet.
@PARTITION_FIELD	Jede	Ersetzt den Namen des aktuellen Partitionsfelds.
@TRAINING_PARTITION	Jede	Ergibt den Wert der aktuellen Trainingspartition. Beispiel: Zur Auswahl von Trainingsdatensätzen mit einem Auswahlknoten verwenden Sie den CLEM-Ausdruck: @PARTITION_FIELD = @TRAINING_PARTITION Dadurch wird gewährleistet, dass der Auswahlknoten immer funktioniert, unabhängig davon, welche Werte zur Darstellung der einzelnen Partitionen in den Daten verwendet werden.
@TESTING_PARTITION	Jede	Ergibt den Wert der aktuellen Testpartition.
@VALIDATION_PARTITION	Jede	Ergibt den Wert der aktuellen Validierungspartition.
@FIELDS_BETWEEN(start, end)	Jede	Ergibt die Liste der Feldnamen, die zwischen dem angegebenen Anfangs- und Endfeld (einschließlich dieser Felder) liegen, und zwar auf der Grundlage der natürlichen Reihenfolge der Felder in den Daten, also gemäß der Einfügereihenfolge. Für weitere Informationen siehe Thema Zusammenfassen mehrerer Felder in Kapitel 7 auf S. 129.

Funktion	Ergebnis	Beschreibung
@FIELDS_MATCHING(pattern)	Jede	Ergibt eine Liste der Feldnamen, die mit dem angegebenen Muster übereinstimmen. Das Fragezeichen (?) als Platzhalterzeichen kann die Stelle von genau einem beliebigen Zeichen einnehmen, das Sternchen (*) entspricht keinem, einem oder mehreren Zeichen. Soll ein Fragezeichen oder ein Sternchen nicht als Platzhalter fungieren, sondern bei der Suche berücksichtigt werden, geben Sie den umgekehrten Schrägstrich () als Escape-Zeichen ein. Für weitere Informationen siehe Thema Zusammenfassen mehrerer Felder in Kapitel 7 auf S. 129.
@MULTI_RESPONSE_SET	Jede	Ergibt die Liste der Felder im benannten Mehrfachantworten-Set. Für weitere Informationen siehe Thema Arbeiten mit Mehrfachantwortdaten in Kapitel 7 auf S. 130.

Verwenden von IBM SPSS Modeler mit einem Repository

Informationen zu IBM SPSS Collaboration and Deployment Services Repository

Sie können IBM® SPSS® Modeler zusammen mit einem IBM SPSS Collaboration and Deployment Services-Repository verwenden und dadurch die Lebensdauer von Data Mining-Modellen und verwandten Vorhersageobjekten verwalten. Dies ermöglicht, dass diese Objekte im Zusammenhang mit Unternehmensanwendungen, Tools und Lösungen benutzt werden. SPSS Modeler-Objekte, die auf diese Weise gemeinsam benutzbar sind, umfassen Streams, Knoten, Stream-Ausgaben, Szenarien, Projekte und Modelle. Die Objekte werden im zentralen Repository gespeichert, wo sie für andere Anwendungen freigegeben und mithilfe von umfassenden Versionserstellungs-, Metadaten- und Suchfunktionen aufgezeichnet werden können.

Hinweis: Für den Zugriff auf ein IBM® SPSS® Collaboration and Deployment Services-Repository ist eine separate Lizenz erforderlich. Weitere Informationen finden Sie im Dokument <http://www.ibm.com/software/analytics/spss/products/deployment/cds/>

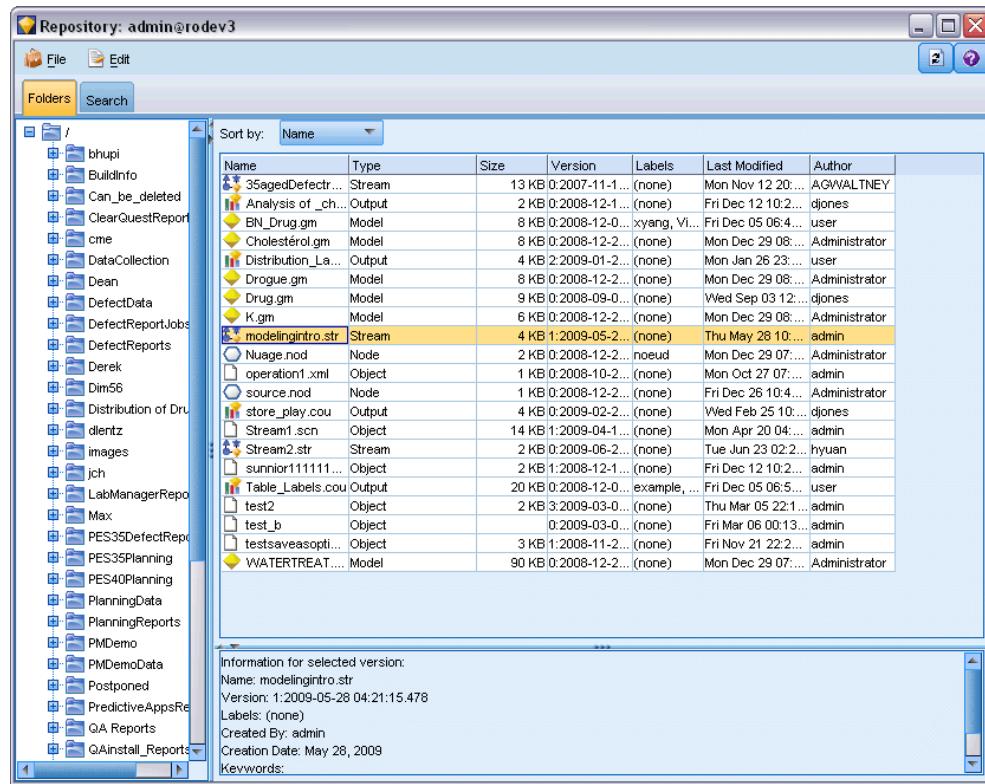
Bevor Sie SPSS Modeler mit dem Repository verwenden können, müssen Sie einen Adapter auf dem Repository-Host installieren. Ohne diesen Adapter sehen Sie eventuell folgende Meldung beim Versuch, auf Repository-Objekte von bestimmten SPSS Modeler-Knoten oder -Modellen zuzugreifen:

Das Repository benötigt evtl. Updates zur Unterstützung neuer Knoten-, Modell- und Ausgabetypen.

Eine Anleitung zur Installation des Adapters finden Sie im Handbuch *SPSS Modeler Deployment-Installation*, das unter SPSS Modeler Deployment DVD zu finden ist. Genauere Angaben dazu, wie Sie auf SPSS Modeler-Repository-Objekte von IBM® SPSS® Collaboration and Deployment Services Deployment Manager zugreifen, finden Sie im *SPSS Modeler Deployment Guide*.

Die folgenden Abschnitte enthalten Informationen zum Zugriff auf das Repository von SPSS Modeler aus.

Abbildung 9-1
Objekte im IBM SPSS Collaboration and Deployment Services Repository



Umfassende Versionserstellung und Unterstützung der Suchfunktion

Das Repository bietet umfassende Funktionen für die Objektversionserstellung und Suchmöglichkeiten. Angenommen, Sie erstellen z. B. einen Stream und speichern ihn im Repository, wo Sie ihn mit Mitarbeitern aus anderen Abteilungen gemeinsam nutzen können. Wenn Sie den Stream später in SPSS Modeler aktualisieren, können Sie die aktualisierte Version zum Repository hinzufügen, ohne die frühere Version zu überschreiben. Es können weiterhin alle Versionen aufgerufen und nach Name, Bezeichnung, verwendete Felder und anderen Attributen durchsucht werden. So können Sie z. B. alle Modellversionen suchen, die die Nettoeinkünfte als Eingabe verwenden, oder alle von einem bestimmten Autor erstellten Modelle. (Mit einem herkömmlichen Dateisystem müssten Sie jede Version unter einem anderen Dateinamen speichern, und die Beziehungen zwischen den Versionen wären der Software nicht bekannt.)

Einzelanmeldung

Mit der Einzelanmeldungsfunction können sich Benutzer mit dem Repository verbinden, ohne jedes Mal Benutzername und Passwort eingeben zu müssen. Die bestehenden Anmeldedaten für das lokale Netzwerk sind als Authentifizierung für IBM SPSS Collaboration and Deployment Services ausreichend. Diese Funktion hängt von folgenden Bedingungen ab:

- IBM SPSS Collaboration and Deployment Services muss für die Verwendung eines Einzelanmeldungsanbieters konfiguriert sein.
- Der Benutzer muss in einem Host angemeldet sein, der mit dem Anbieter kompatibel ist.

Für weitere Informationen siehe Thema [Verbinden mit dem Repository auf S. 183](#).

Speichern und Bereitstellen von Repository-Objekten

Streams, die in IBM® SPSS® Modeler erstellt werden, lassen sich so, wie sie sind, als Dateien mit der Erweiterung **.str** im Repository **speichern**. Auf diese Weise können mehrere Benutzer im gesamten Unternehmen auf einen einzelnen Stream zugreifen. Für weitere Informationen siehe Thema [Speichern von Objekten im Repository auf S. 186](#).

Es ist auch möglich, einen Stream im Repository **bereitzustellen**. Bereitgestellte Streams werden als Datei mit zusätzlichen Metadaten gespeichert. Bereitgestellte Streams können die Unternehmensfunktionen von IBM SPSS Collaboration and Deployment Services, wie automatisiertes Scoring und Modellaktualisierung, vollständig nutzen. So kann beispielsweise ein Modell automatisch in regelmäßigen Abständen aktualisiert werden, wenn neue Daten verfügbar werden. Alternativ kann ein Set von Streams zum Zweck der Champion-Challenger-Analyse verwendet werden, in der Streams verglichen werden, um zu ermitteln, welcher das wirksamste Vorhersagemodell enthält.

Streams können auf zwei verschiedene Weisen bereitgestellt werden: als Stream (mit der Erweiterung **.str**) oder als Szenario (mit der Erweiterung **.scn**). Bei der Bereitstellung als Stream kann der Stream durch die Thin-Client-Anwendung IBM® SPSS® Modeler Advantage aktiviert werden. Für weitere Informationen siehe Thema [Öffnen eines Streams in IBM SPSS Modeler Advantage in Kapitel 10 auf S. 219](#). Durch die Bereitstellung als Szenario kann der Stream von Predictive Applications Version 5, dem Vorgänger von IBM SPSS Modeler Advantage, verwendet werden.

Weitere Informationen finden Sie unter [Deployment-Optionen für den Stream auf S. 208](#).

Anforderungen für Streams, die als Szenarien bereitgestellt werden

- Um einen konsistenten Zugriff auf Unternehmensdaten zu gewährleisten, muss der Zugriff auf die als Szenarien bereitgestellten Streams über die Enterprise View-Komponente von IBM® SPSS® Collaboration and Deployment Services erfolgen. Das bedeutet, dass in SPSS Modeler für jede angegebene Scoring- bzw. Modellierungsverzweigung mindestens ein Quellenknoten vom Typ “Enterprise-Ansicht” vorhanden sein muss. Für weitere Informationen siehe Thema [Enterprise-Ansichts-Knoten in Kapitel 2 in IBM SPSS Modeler 15 Quellen-, Prozess- und Ausgabeknoten](#).

- Um den Enterprise-Ansichts-Knoten verwenden zu können, muss IBM SPSS Collaboration and Deployment Services an Ihrem Standort installiert, konfiguriert und zugreifbar sein. Dabei müssen bereits eine Enterprise-Ansicht, Application-Ansichten und DPDs (Data Provider Definitions) definiert sein. Weitere Informationen erhalten Sie bei Ihrem lokalen Administrator oder auf der Unternehmens-Website unter <http://www.ibm.com/software/analytics/spss/products/deployment/cds/>.
- Eine DPD wird anhand einer bestimmten ODBC-Datenquelle definiert. Um eine DPD aus SPSS Modeler zu verwenden, muss eine ODBC-Datenquelle auf dem SPSS Modeler Server-Host definiert sein, der denselben Namen trägt und der eine Verbindung zum selben Datenspeicher herstellt wie dem in der DPD referenzierten Datenspeicher.
- Außerdem muss auf jedem Computer, der zur Bearbeitung oder Ausführung des Streams verwendet wird, der IBM® SPSS® Collaboration and Deployment Services Enterprise View Driver installiert sein. Unter Windows installieren Sie den Treiber einfach auf dem Computer, auf dem SPSS Modeler bzw. SPSS Modeler Server installiert ist. Es ist keine weitere Konfiguration des Treibers erforderlich. Unter UNIX muss ein Verweis auf das Skript *pev.sh* zum Startskript hinzugefügt werden. Für weitere Informationen siehe Thema **Konfigurieren eines Treibers für den Enterprise-Ansichts-Knoten in Anhang B in IBM SPSS Modeler Server 15-Verwaltungs- und Leistungshandbuch**. Details zur Installation des IBM SPSS Collaboration and Deployment Services Enterprise View Driver erhalten Sie bei Ihrem lokalen Administrator.

Andere Bereitstellungsoptionen

IBM SPSS Collaboration and Deployment Services bietet zwar bereits extrem umfassende Funktionen zur Verwaltung von Unternehmensinhalten, es steht aber dennoch eine Reihe weiterer Mechanismen zum Deployment und Export von Streams zur Verfügung, darunter:

- Exportieren Sie den Stream zur späteren Verwendung in IBM® SPSS® Modeler Solution Publisher Runtime. Für weitere Informationen siehe Thema **IBM SPSS Modeler Solution Publisher in Kapitel 2 in IBM SPSS Modeler 15 Solution Publisher**.
- Export eines oder mehrerer Modelle in PMML, einem XML-basierten Format zur Kodierung von Modellinformationen. Für weitere Informationen siehe Thema **Importieren und Exportieren von Modellen als PMML in Kapitel 10 auf S. 220**.

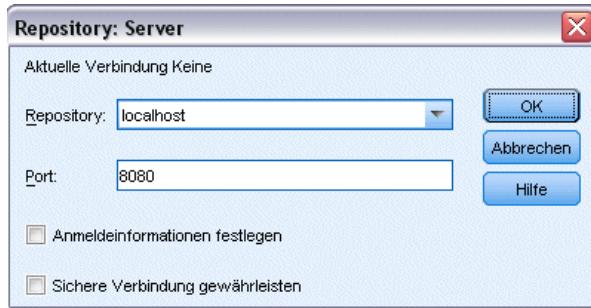
Verbinden mit dem Repository

- ▶ Um eine Verbindung mit dem Repository herzustellen, klicken Sie im IBM® SPSS® Modeler-Hauptmenü auf:
Werkzeuge > Repository > Optionen...
- ▶ Legen Sie die gewünschten Anmeldeoptionen fest.

Die Einstellungen sind für jeden Standort oder jede Installation spezifisch. Für spezielle Port- und andere Anmeldedaten wenden Sie sich an den lokalen Systemadministrator.

Hinweis: Für den Zugriff auf ein IBM® SPSS® Collaboration and Deployment Services-Repository ist eine separate Lizenz erforderlich. Weitere Informationen finden Sie im Dokument <http://www.ibm.com/software/analytics/spss/products/deployment/cds/>

Abbildung 9-2
Anmeldung bei IBM SPSS Collaboration and Deployment Services Repository



Repository. Die Repository-Installation, auf die Sie zugreifen möchten. Im Allgemeinen entspricht dies dem Namen des Host-Servers, auf dem das Repository installiert ist. Sie können jeweils nur mit einem einzigen Repository verbunden sein.

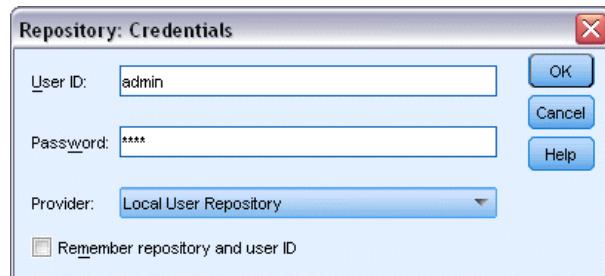
Port. Der Port, der für die Verbindung verwendet wird, in der Regel 8080.

Anmeldedaten einstellen. Belassen Sie dieses Kontrollkästchen nicht aktiviert, um die Funktion für die **Einzelanmeldung** zu aktivieren. Diese versucht, Sie mithilfe Ihres lokalen Benutzernamens und Passworts anzumelden. Falls die Einzelanmeldung nicht möglich ist oder Sie das Kontrollkästchen zur Deaktivierung der Einzelanmeldung aktivieren (z. B. zur Anmeldung in ein Administratorkonto), wird ein weiteres Fenster angezeigt, in dem Sie aufgefordert werden, Ihre Anmeldedaten einzugeben.

Sichere Verbindung gewährleisten. Legt fest, ob eine Secure Sockets Layer (SSL)-Verbindung verwendet werden soll. SSL ist ein weit verbreitetes Protokoll zum Schutz der über ein Netzwerk versendeten Daten. Um diese Funktion verwenden zu können, muss SSL auf dem Server, auf dem sich das Repository befindet, aktiviert sein. Wenden Sie sich gegebenenfalls an den lokalen Administrator, wenn Sie weitere Informationen benötigen.

Eingabe der Anmeldedaten für das Repository

Abbildung 9-3
Eingabe der IBM SPSS Collaboration and Deployment Services Repository-Anmeldedaten



Benutzername und Passwort. Geben Sie einen gültigen Benutzernamen und ein Passwort für die Anmeldung an. Bitten Sie, falls erforderlich, Ihren lokalen Administrator um weitere Informationen.

Anbieter. Wählen Sie zur Authentifizierung einen Sicherheitsanbieter. Das Repository kann für die Verwendung verschiedener Sicherheitsanbieter konfiguriert werden; wenn Sie mehr Hilfe benötigen, wenden Sie sich an Ihren lokalen Administrator.

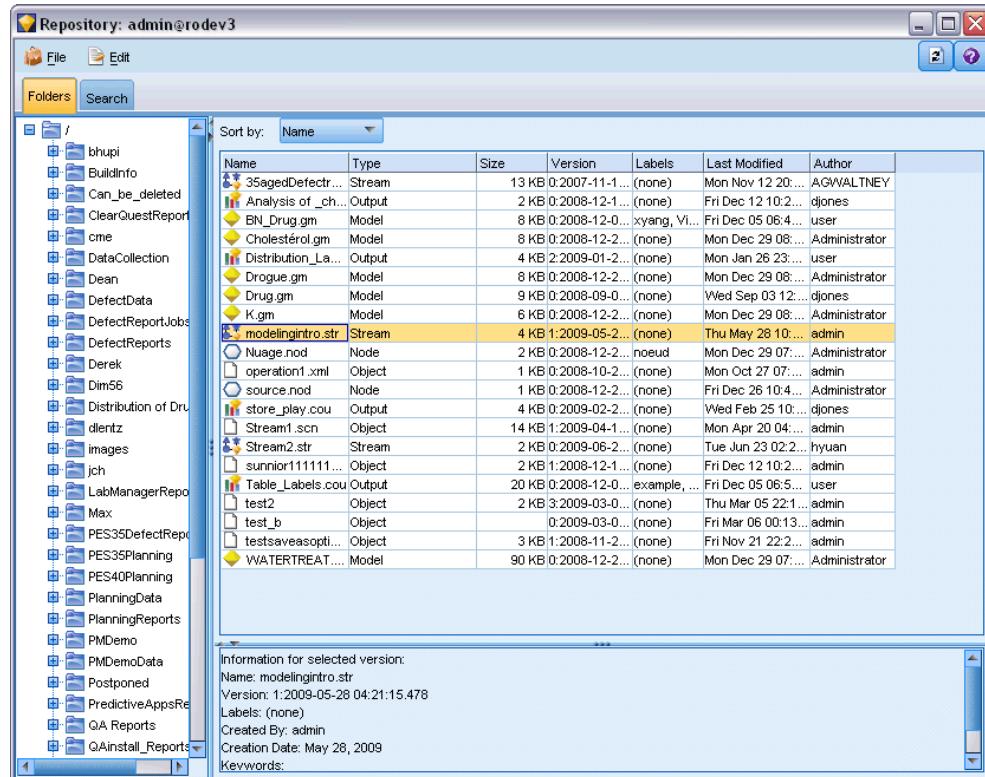
Repository und Benutzername merken. Speichert die aktuellen Einstellungen als Standard, sodass Sie diese nicht bei jeder neuen Verbindung erneut eingeben müssen.

Durchsuchen des Repository-Inhalts

Mit dem Repository können Sie die gespeicherten Inhalte ähnlich wie mit Windows Explorer durchsuchen; außerdem können Sie verschiedene *Versionen* von jedem gespeicherten Objekt durchsuchen.

- ▶ Um das IBM® SPSS® Collaboration and Deployment Services Repository-Fenster zu öffnen, klicken Sie in den SPSS Modeler-Menüs auf:
Werkzeuge > Repository > Explorative Datenanalyse...
- ▶ Geben Sie gegebenenfalls Verbindungseinstellungen zum Repository ein. [Für weitere Informationen siehe Thema Verbinden mit dem Repository auf S. 183](#). Für spezielle Port-, Passwort- und andere Verbindungsdaten wenden Sie sich an den lokalen Systemadministrator.

Abbildung 9-4
Durchsuchen des IBM SPSS Collaboration and Deployment Services Repository-Inhalts



Das Explorer-Fenster zeigt anfangs eine Baumansicht der Ordnerhierarchie. Klicken Sie auf einen Ordnernamen, um den Inhalt des Ordners anzuzeigen.

Objekte, die der aktuellen Auswahl oder Suchkriterien entsprechen, werden zusammen mit detaillierten Informationen über die ausgewählte Version, die im unteren rechten Fenster angezeigt wird, im rechten Fenster aufgelistet. Die angezeigten Attribute gelten für die aktuellste Version.

Speichern von Objekten im Repository

Abbildung 9-5
Speichern von Modellen



Sie können Streams, Knoten, Modelle, Modellpaletten, Projekte und Ausgabeobjekte im Repository speichern, von wo andere Benutzer und Anwendungen darauf zugreifen können.

Hinweis: Für den Zugriff auf ein IBM® SPSS® Collaboration and Deployment Services-Repository ist eine separate Lizenz erforderlich. Weitere Informationen finden Sie im Dokument <http://www.ibm.com/software/analytics/spss/products/deployment/cds/>

Sie können auch Stream-Ausgaben im Repository in einem Format veröffentlichen, in dem andere Benutzer diese mithilfe von IBM® SPSS® Collaboration and Deployment Services Deployment Portal über das Internet betrachten können. Für weitere Informationen siehe Thema Im Web veröffentlichen in Kapitel 6 in *IBM SPSS Modeler 15 Quellen-, Prozess- und Ausgabeknoten*.

Festlegen der Objekteigenschaften

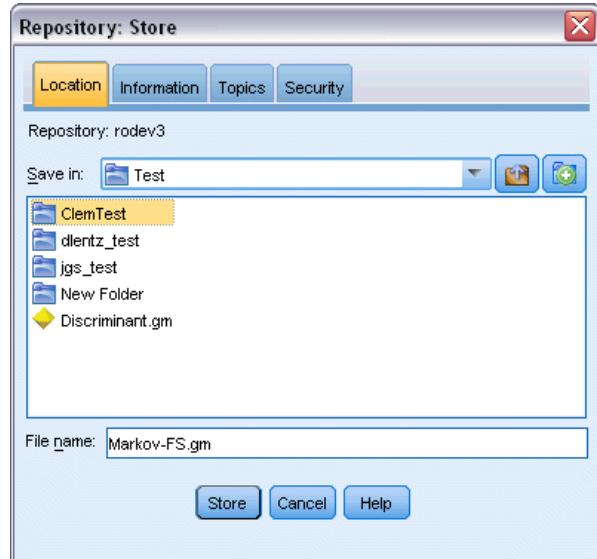
Beim Speichern eines Objekts wird das Dialogfeld “Repository: Speichern” angezeigt, in dem Sie die Werte von mehreren Eigenschaften für das Objekt einstellen können. Sie verfügen über folgende Möglichkeiten:

- Namen und Repository-Ordner wählen, unter dem das Objekt gespeichert werden soll
- Informationen über das Objekt hinzufügen, z. B. Versionslabel und andere Eigenschaften, nach denen gesucht werden kann
- Dem Objekt ein oder mehrere Klassifizierungsthemen hinzufügen
- Sicherheitsoptionen für das Objekt festlegen

In den folgenden Abschnitten werden die einstellbaren Eigenschaften beschrieben.

Wählen des Speicherorts für Objekte

Abbildung 9-6
Wählen des Speicherorts für ein Objekt



Speichern in. Zeigt den aktuellen Ordner an, in dem das Objekt gespeichert wird. Doppelklicken Sie auf einen Ordnernamen in der Liste, um diesen Ordner als aktuellen Ordner festzulegen. Mit der Schaltfläche “Ordner nach oben” navigieren Sie zum übergeordneten Ordner. Mit der Schaltfläche “Neuer Ordner” erstellen Sie einen Ordner auf der aktuellen Ebene.

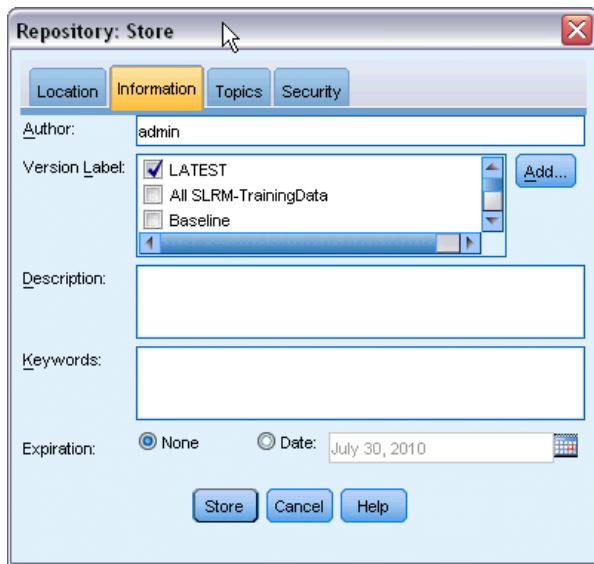
Dateiname. Der Name, unter dem das Objekt gespeichert wird.

Speichern. Speichert das Objekt am aktuellen Speicherort.

Hinzufügen von Informationen über gespeicherte Objekte

Alle Felder auf dieser Registerkarte sind optional.

Abbildung 9-7
Hinzufügen von Informationen über das Objekt



Autor. Der Benutzername des Benutzers, der das Objekt im Repository erstellt. Standardmäßig wird der Benutzername angezeigt, der für die Repository-Verbindung verwendet wurde, Sie können diesen Namen aber hier ändern.

Versionsbezeichnung. Wählen Sie eine Bezeichnung aus der Liste, um die Objektversion anzugeben, oder klicken Sie auf Hinzufügen, um eine neue Bezeichnung zu erstellen. Verwenden Sie in der Bezeichnung nicht das Zeichen “[“. Stellen Sie sicher, dass keine Kontrollkästchen aktiviert sind, wenn Sie dieser Objektversion keine Bezeichnung zuweisen möchten. Für weitere Informationen siehe Thema Anzeigen und Bearbeiten von Objekteigenschaften auf S. 204.

Beschreibung. Eine Beschreibung des Objekts. Benutzer können Objekte nach Beschreibung suchen (siehe Hinweis).

Schlüsselwörter. Ein oder mehrere Stichwörter, die sich auf das Objekt beziehen und die zum Suchen verwendet werden können (siehe Hinweis).

Ablaufdatum. Ein Datum, nach dem das Objekt nicht mehr für allgemeine Benutzer sichtbar ist, obwohl es sein Eigentümer und der Repository-Administrator noch sehen können. Zum Festlegen des Ablaufdatums wählen Sie die Option Datum und geben Sie das Datum ein oder wählen Sie ein Datum mithilfe der Kalender-Schaltfläche.

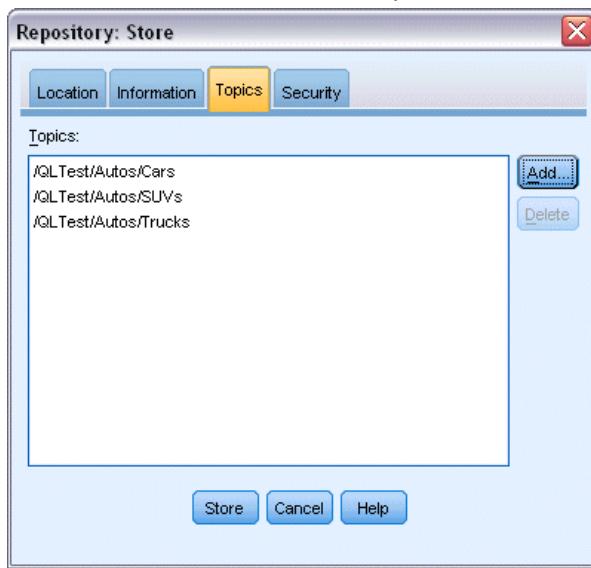
Speichern. Speichert das Objekt am aktuellen Speicherort.

Hinweis: Die Informationen in den Feldern Beschreibung und Schlüsselwörter werden getrennt von den Angaben in SPSS Modeler auf der Registerkarte “Anmerkungen” des Objekts behandelt. Eine Repository-Suche nach Beschreibung oder Schlüsselwort liefert keine Informationen von der Registerkarte “Anmerkungen”. Für weitere Informationen siehe Thema Suchen nach Objekten im Repository auf S. 197.

Zuweisen von Themen zu einem gespeicherten Objekt

Themen sind ein hierarchisches Klassifizierungssystem für den Inhalt, der im Repository gespeichert ist. Sie können beim Speichern von Objekten aus den verfügbaren Themen wählen, und Benutzer können auch Objekte nach Thema suchen. Die Liste verfügbarer Themen wird von Repository-Benutzern mit den geeigneten Berechtigungen festgelegt. (Weitere Informationen finden Sie im *Deployment Manager User's Guide*).

Abbildung 9-8
Zuweisen von Themen zu einem Objekt



So weisen Sie dem Objekt ein Thema zu:

- ▶ Klicken Sie auf die Schaltfläche Hinzufügen.
- ▶ Klicken Sie in der Liste verfügbarer Themen auf einen Themennamen.
- ▶ Klicken Sie auf OK.

So entfernen Sie eine Themenzuweisung:

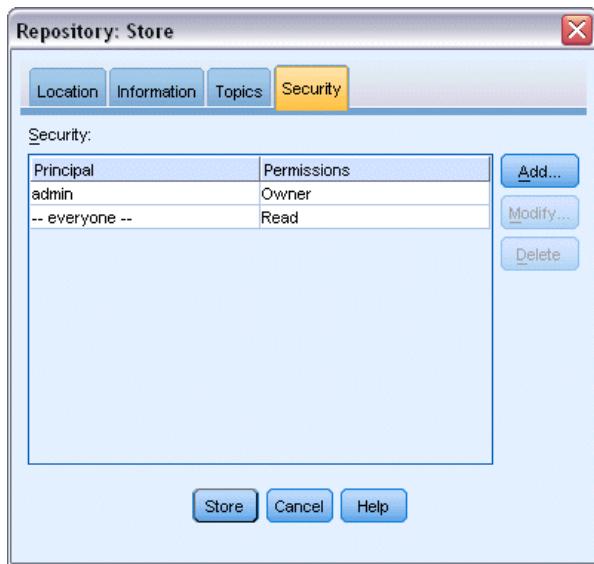
- ▶ Wählen Sie das Thema in der Liste der zugewiesenen Themen aus.
- ▶ Klicken Sie auf Löschen.

Festlegen der Sicherheitsoptionen für gespeicherte Objekte

Sie können eine Reihe von Sicherheitsoptionen für ein gespeichertes Objekt festlegen. Für ein oder mehrere **Prinzipale** (d. h. Benutzer oder Benutzergruppen) können Sie:

- dem Objekt Zugriffsrechte zuweisen
- Zugriffsrechte auf das Objekt ändern
- Zugriffsrechte auf das Objekt entfernen

Abbildung 9-9
Festlegen von Sicherheitsoptionen für ein Objekt



Prinzipal. Der Repository-Benutzername des Benutzers oder der Gruppe mit Zugriffsrechten auf dieses Objekt.

Berechtigungen. Die Zugriffsrechte dieses Benutzers bzw. dieser Gruppe auf das Objekt.

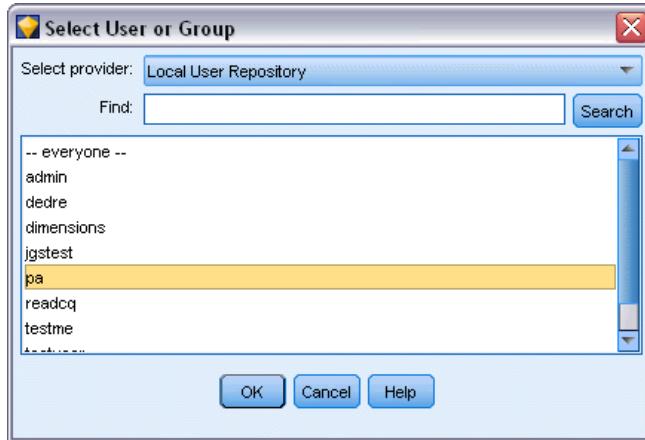
Hinzufügen. Ermöglicht Ihnen, einen oder mehrere Benutzer zu der Liste hinzuzufügen, in der sich diejenigen mit Zugriffsrechten auf dieses Objekt befinden. Für weitere Informationen siehe [Thema Hinzufügen eines Benutzers zur Berechtigungsliste auf S. 191](#).

Ändern. Ermöglicht Ihnen, die Zugriffsrechte des ausgewählten Benutzers bzw. der ausgewählten Gruppe auf dieses Objekt zu ändern. Lesezugriff wird standardmäßig gewährt. Mithilfe dieser Option können Sie zusätzliche Zugriffsrechte gewähren, nämlich "Eigentümer", "Schreiben", "Löschen" und "Ändern".

Löschen. Löscht den ausgewählten Benutzer bzw. die ausgewählte Gruppe aus der Berechtigungsliste für dieses Objekt.

Hinzufügen eines Benutzers zur Berechtigungsliste

Abbildung 9-10
Hinzufügen eines Benutzers zur Berechtigungsliste für ein Objekt



Anbieter auswählen. Wählen Sie zur Authentifizierung einen Sicherheitsanbieter. Das Repository kann für die Verwendung verschiedener Sicherheitsanbieter konfiguriert werden; wenn Sie mehr Hilfe benötigen, wenden Sie sich an Ihren lokalen Administrator.

Suchen. Geben Sie den Repository-Benutzernamen des Benutzers bzw. der Gruppe ein, die Sie hinzufügen möchten, und klicken Sie auf Suchen, um diesen Namen in der Benutzerliste anzuzeigen. Um mehrere Benutzernamen gleichzeitig hinzuzufügen, lassen Sie dieses Feld leer und klicken Sie einfach auf Suchen, um eine Liste mit allen Repository-Benutzernamen anzuzeigen.

Benutzerliste. Wählen Sie einen oder mehrere Benutzernamen aus der Liste und klicken Sie auf OK, um sie der Berechtigungsliste hinzuzufügen.

Ändern der Zugriffsrechte für ein Objekt

Abbildung 9-11
Ändern der Zugriffsrechte für ein Objekt



Eigentümer. Wählen Sie diese Option, um diesem Benutzer bzw. dieser Gruppe Zugriffsrechte auf das Objekt zu geben. Der Eigentümer verfügt über die kompletten Zugriffsrechte auf das Objekt einschließlich Lösch- und Änderungsrechten.

Lesen. Standardmäßig hat ein Benutzer bzw. eine Gruppe, die nicht Objekteigentümer ist, nur Leserechte auf das Objekt. Markieren Sie die geeigneten Kontrollkästchen für Schreiben, Löschen und Ändern, um die entsprechenden Zugriffsrechte für diesen Benutzer bzw. diese Gruppe hinzuzufügen.

Speichern von Streams

Sie können einen Stream als *.str*-Datei im Repository speichern, von wo andere Benutzer darauf zugreifen können.

Hinweis: Weitere Informationen über das Bereitstellen eines Streams, um zusätzliche Repository-Funktionen zu nutzen, finden Sie unter Bereitstellen von Streams auf S. 207.

So speichern Sie den aktuellen Stream:

- ▶ Klicken Sie im Hauptmenü auf Folgendes:
Datei > Speichern > Als Stream speichern...
- ▶ Geben Sie gegebenenfalls Verbindungseinstellungen zum Repository ein. [Für weitere Informationen siehe Thema Verbinden mit dem Repository auf S. 183](#). Für spezielle Port-, Passwort- und andere Verbindungsdaten wenden Sie sich an den lokalen Systemadministrator.
- ▶ Wählen Sie im Dialogfeld “Repository: Speichern” den Ordner aus, in dem Sie das Objekt speichern möchten, geben Sie alle weiteren Informationen an, die Sie aufzeichnen möchten, und klicken Sie auf die Schaltfläche Speichern. [Für weitere Informationen siehe Thema Festlegen der Objekteigenschaften auf S. 186](#).

Temporäres Speichern von Projekten

Sie können ein vollständiges IBM® SPSS® Modeler-Projekt als *.cpj*-Datei im Repository speichern, damit andere Benutzer darauf zugreifen können.

Da eine Projektdatei ein Container für andere SPSS Modeler-Objekte ist, müssen Sie angeben, wo SPSS Modeler die Objekte des Projekts im Repository speichern soll. Dies geschieht über eine Einstellung im Dialogfeld Projekteigenschaften. [Für weitere Informationen siehe Thema Festlegen der Projekteigenschaften in Kapitel 11 auf S. 231](#).

Sobald Sie ein Projekt zum Speichern von Objekten im Repository konfiguriert haben, werden Sie immer, wenn Sie ein neues Objekt zum Projekt hinzufügen, automatisch von SPSS Modeler aufgefordert, das Objekt zu speichern.

Nach Abschluss der SPSS Modeler-Sitzung müssen Sie eine neue Version der Projektdatei speichern, damit Ihre Änderungen erhalten bleiben. Die Projektdatei enthält automatisch die aktuellsten Versionen ihrer Objekte (und ruft sie ab). Wenn Sie einem Projekt während einer SPSS Modeler-Sitzung keine Objekte hinzugefügt haben, brauchen Sie die Projektdatei nicht wiederherzustellen. Sie müssen jedoch neuere Versionen der Projektobjekte (Streams, Ausgaben usw.), die Sie geändert haben, temporär speichern.

So speichern Sie ein Projekt:

- ▶ Wählen Sie das Projekt in der Registerkarte “CRISP-DM” oder “Klassen” im Manager-Bereich in SPSS Modeler aus. Klicken Sie dann im Hauptmenü auf:
Datei > Projekt > Projekt temporär speichern...
- ▶ Geben Sie gegebenenfalls Verbindungseinstellungen zum Repository ein. [Für weitere Informationen siehe Thema Verbinden mit dem Repository auf S. 183.](#) Für spezielle Port-, Passwort- und andere Verbindungsdaten wenden Sie sich an den lokalen Systemadministrator.
- ▶ Wählen Sie im Dialogfeld “Repository: Speichern” den Ordner aus, in dem Sie das Objekt speichern möchten, geben Sie alle weiteren Informationen an, die Sie aufzeichnen möchten, und klicken Sie auf die Schaltfläche Speichern. [Für weitere Informationen siehe Thema Festlegen der Objekteigenschaften auf S. 186.](#)

Speichern von Knoten

Sie können eine individuelle Knotendefinition aus dem aktuellen Stream als *.nod*-Datei im Repository speichern, von wo andere Benutzer darauf zugreifen können.

So speichern Sie einen Knoten:

- ▶ Klicken Sie im Stream-Zeichenbereich mit der rechten Maustaste auf den Knoten und klicken Sie auf Knoten speichern.
- ▶ Geben Sie gegebenenfalls Verbindungseinstellungen zum Repository ein. [Für weitere Informationen siehe Thema Verbinden mit dem Repository auf S. 183.](#) Für spezielle Port-, Passwort- und andere Verbindungsdaten wenden Sie sich an den lokalen Systemadministrator.
- ▶ Wählen Sie im Dialogfeld “Repository: Speichern” den Ordner aus, in dem Sie das Objekt speichern möchten, geben Sie alle weiteren Informationen an, die Sie aufzeichnen möchten, und klicken Sie auf die Schaltfläche Speichern. [Für weitere Informationen siehe Thema Festlegen der Objekteigenschaften auf S. 186.](#)

Speichern von Ausgabeobjekten

Sie können ein Ausgabeobjekt aus dem aktuellen Stream als *.cou*-Datei im Repository speichern, von wo andere Benutzer darauf zugreifen können.

So speichern Sie ein Ausgabeobjekt:

- ▶ Klicken Sie in der Registerkarte “Ausgaben” im Manager-Bereich in SPSS Modeler auf das Objekt und klicken Sie dann im Hauptmenü auf:
Datei > Ausgaben > Ausgabe speichern...
- ▶ Klicken Sie als Alternative mit der rechten Maustaste auf ein Objekt in der Registerkarte “Ausgaben” und klicken Sie auf Speichern.
- ▶ Geben Sie gegebenenfalls Verbindungseinstellungen zum Repository ein. [Für weitere Informationen siehe Thema Verbinden mit dem Repository auf S. 183.](#) Für spezielle Port-, Passwort- und andere Verbindungsdaten wenden Sie sich an den lokalen Systemadministrator.

- ▶ Wählen Sie im Dialogfeld “Repository: Speichern” den Ordner aus, in dem Sie das Objekt speichern möchten, geben Sie alle weiteren Informationen an, die Sie aufzeichnen möchten, und klicken Sie auf die Schaltfläche Speichern. [Für weitere Informationen siehe Thema Festlegen der Objekteigenschaften auf S. 186.](#)

Speichern von Modellen und Modellpaletten

Sie können ein einzelnes Modell als .gm-Datei im Repository speichern, von wo andere Benutzer darauf zugreifen können. Sie können auch den vollständigen Inhalt der Modellpalette als .gen-Datei im Repository speichern.

Speichern von Modellen

- ▶ Klicken Sie in der Modellpalette in SPSS Modeler auf das Objekt und klicken Sie dann im Hauptmenü auf:
Datei > Models > Modell speichern...
- ▶ Klicken Sie als Alternative mit der rechten Maustaste auf ein Objekt in der Modellpalette und klicken Sie dann auf Modell speichern.
- ▶ Fahren Sie mit “Abschließen des Speichervorgangs” fort.

Speichern einer Modellpalette

- ▶ Klicken Sie mit der rechten Maustaste in den Hintergrund der Modellpalette.
- ▶ Klicken Sie im Popup-Menü auf Palette speichern.
- ▶ Fahren Sie mit “Abschließen des Speichervorgangs” fort.

Abschließen des Speichervorgangs

- ▶ Geben Sie gegebenenfalls Verbindungseinstellungen zum Repository ein. [Für weitere Informationen siehe Thema Verbinden mit dem Repository auf S. 183.](#) Für spezielle Port-, Passwort- und andere Verbindungsdaten wenden Sie sich an den lokalen Systemadministrator.
- ▶ Wählen Sie im Dialogfeld “Repository: Speichern” den Ordner aus, in dem Sie das Objekt speichern möchten, geben Sie alle weiteren Informationen an, die Sie aufzeichnen möchten, und klicken Sie auf die Schaltfläche Speichern. [Für weitere Informationen siehe Thema Festlegen der Objekteigenschaften auf S. 186.](#)

Abrufen von Objekten aus dem Repository

Sie können Streams, Modelle, Modellpaletten, Knoten, Projekte und Ausgabeobjekte abrufen, die temporär im Repository gespeichert wurden.

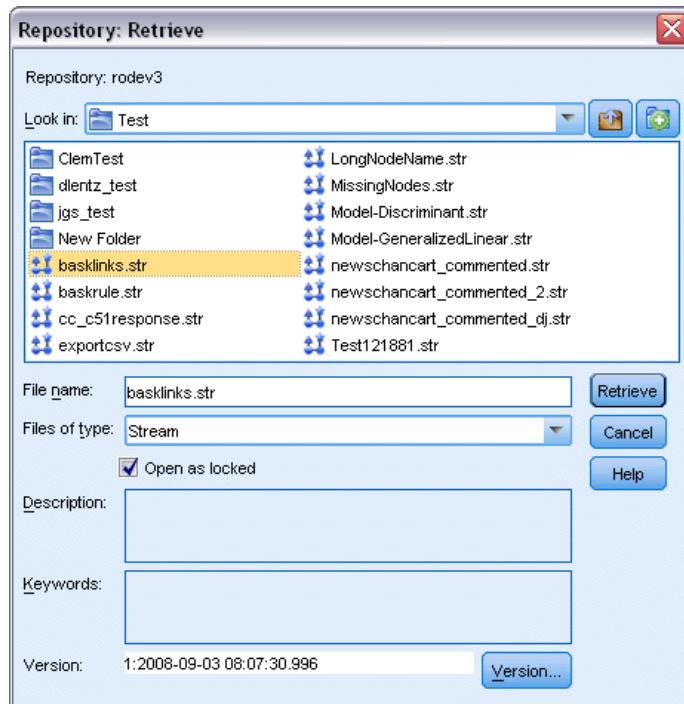
Hinweis: Sie können neben der hier beschriebenen Verwendung der Menüoptionen diese auch nutzen, um Streams, Ausgabeobjekte, Modelle und Modellpaletten abzurufen, indem Sie mit der rechten Maustaste in die jeweilige Registerkarte des Manager-Bereichs oben rechts im SPSS Modeler-Fenster klicken.

- ▶ Zum Abrufen eines Streams klicken Sie im IBM® SPSS® Modeler-Hauptmenü auf:
Datei > Stream abrufen...
- ▶ Um ein Modell, eine Modellpalette, ein Projekt oder ein Ausgabeobjekt abzurufen, klicken Sie im SPSS Modeler-Hauptmenü auf:
Datei > Models > Modell abrufen...
oder
Datei > Models > Modellpalette abrufen...
- oder*
Datei > Projekte > Projekt abrufen...
- oder*
Datei > Ausgaben > Ausgabe abrufen...
- ▶ Klicken Sie als Alternative mit der rechten Maustaste in den Manager- oder Projektbereich und klicken Sie im Popup-Menü auf Abrufen.
- ▶ Zum Abrufen eines Knotens klicken Sie im SPSS Modeler-Hauptmenü auf:
Einfügen > Knoten (oder Superknoten) aus Repository...
- ▶ Geben Sie gegebenenfalls Verbindungseinstellungen zum Repository ein. [Für weitere Informationen siehe Thema Verbinden mit dem Repository auf S. 183](#). Für spezielle Port-, Passwort- und andere Verbindungsdaten wenden Sie sich an den lokalen Systemadministrator.
- ▶ Navigieren Sie im Dialogfeld “Repository: Abrufen” zum Objekt, wählen Sie es aus und klicken Sie auf die Schaltfläche Abrufen.

Wählen eines abzurufenden Objekts

Abbildung 9-12

Abrufen eines Objekts aus dem IBM SPSS Collaboration and Deployment Services Repository



Suchen in: Zeigt die Ordnerhierarchie für den aktuellen Ordner. Um zu einem anderen Ordner zu navigieren, wählen Sie einen aus dieser Liste, um direkt zu dem Ordner zu wechseln, oder navigieren Sie mithilfe der Objektliste unter diesem Feld.

Schaltfläche "Ordner nach oben". Wechselt in der Hierarchie auf die Ebene über dem aktuellen Ordner.

Schaltfläche "Neuer Ordner". Erstellt auf der aktuellen Ebene in der Hierarchie einen neuen Ordner.

Dateiname. Der Repository-Dateiname des ausgewählten Objekts. Klicken Sie zum Abrufen dieses Objekts auf Abrufen.

Dateien des Typs. Der Typ des Objekts, das Sie abrufen möchten. Nur Objekte dieses Typs werden zusammen mit Ordnern in der Objektliste angezeigt. Um Objekte eines anderen Typs zum Abrufen anzuzeigen, wählen Sie den Objekttyp aus der Liste aus.

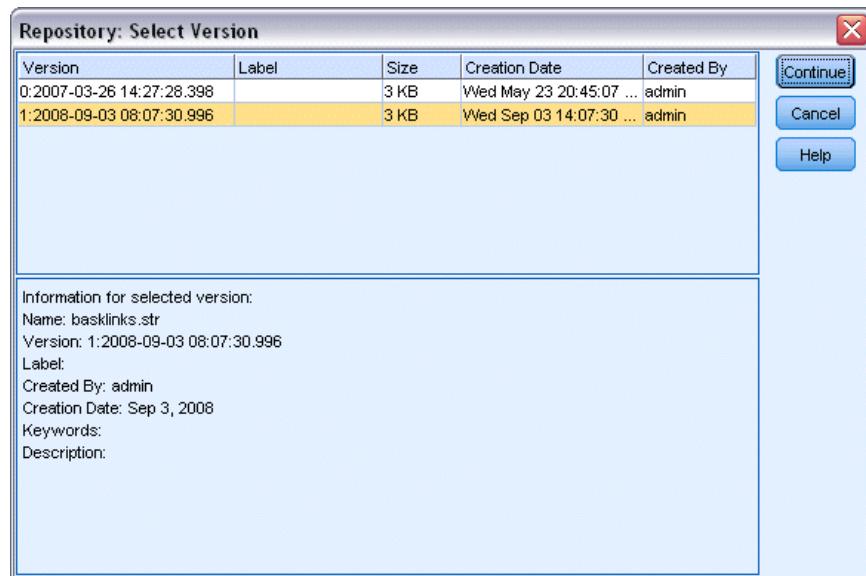
Als gesperrt öffnen. Ein Objekt wird beim Abrufen standardmäßig im Repository gesperrt, damit andere es nicht aktualisieren können. Wenn das Objekt beim Abrufen nicht gesperrt werden soll, deaktivieren Sie dieses Kontrollkästchen.

Beschreibung, Schlüsselwörter. Wenn beim Speichern des Objekts zusätzliche Details über das Objekt definiert wurden, werden diese Details hier angezeigt. Für weitere Informationen siehe Thema [Hinzufügen von Informationen über gespeicherte Objekte](#) auf S. 187.

Version. Um eine andere Objektversion als die aktuellste abzurufen, klicken Sie auf diese Schaltfläche. Es werden detaillierte Informationen für alle Versionen angezeigt, damit Sie die gewünschte Version auswählen können.

Auswählen einer Objektversion

Abbildung 9-13
Auswählen einer Version eines Objekts



So wählen Sie eine bestimmte Version eines Repository-Objekts aus:

- ▶ (Optional) Sortieren Sie die Liste nach Version, Beschriftung, Größe, Erstellungsdatum oder Ersteller, indem Sie auf den Titel der entsprechenden Spalte doppelklicken.
- ▶ Wählen Sie die Objektversion aus, mit der Sie arbeiten möchten.
- ▶ Klicken Sie auf “Weiter”.

Suchen nach Objekten im Repository

Sie können Objekten anhand von Namen, Ordner, Typ, Beschriftung, Datum oder anderen Kriterien suchen.

Suchen nach Name

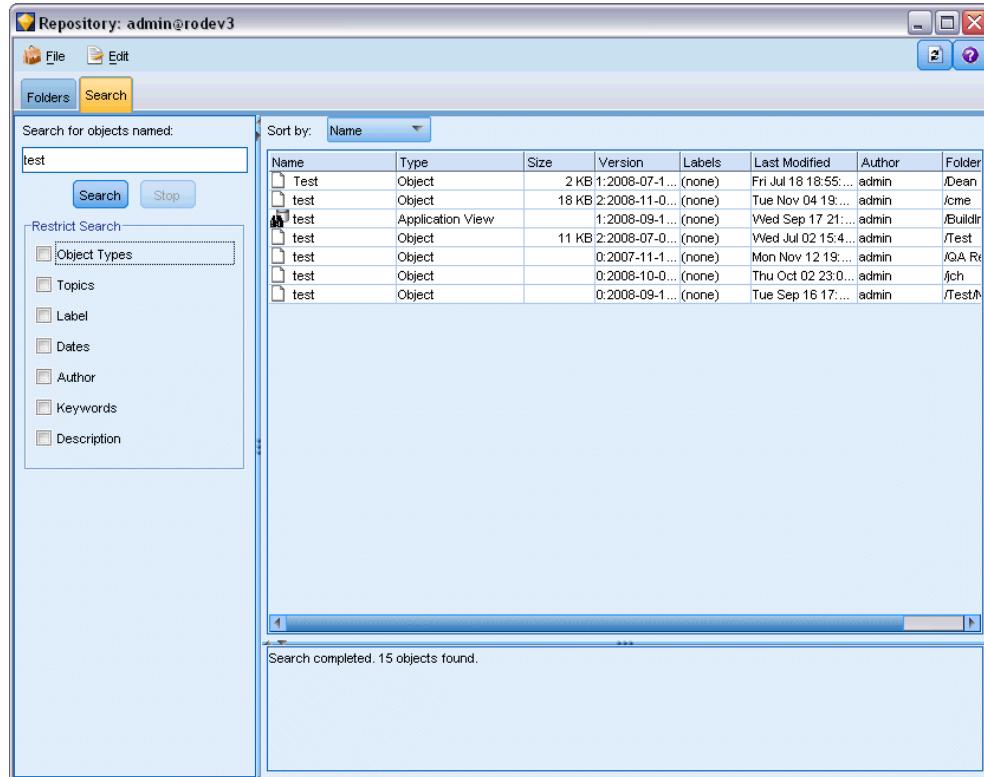
So suchen Sie ein Objekt nach Name:

- ▶ Klicken Sie im IBM® SPSS® Modeler-Hauptmenü auf:
Werkzeuge > Repository > Explorative Datenanalyse...

- ▶ Geben Sie gegebenenfalls Verbindungseinstellungen zum Repository ein. Für weitere Informationen siehe Thema [Verbinden mit dem Repository](#) auf S. 183. Für spezielle Port-, Passwort- und andere Verbindungsdaten wenden Sie sich an den lokalen Systemadministrator.
 - ▶ Klicken Sie auf die Registerkarte Suche.
 - ▶ Geben Sie im Feld Suchen nach Objekten mit Namen den Namen des zu suchenden Objekts an.

Abbildung 9-14

Suchen nach Objekten nach Name



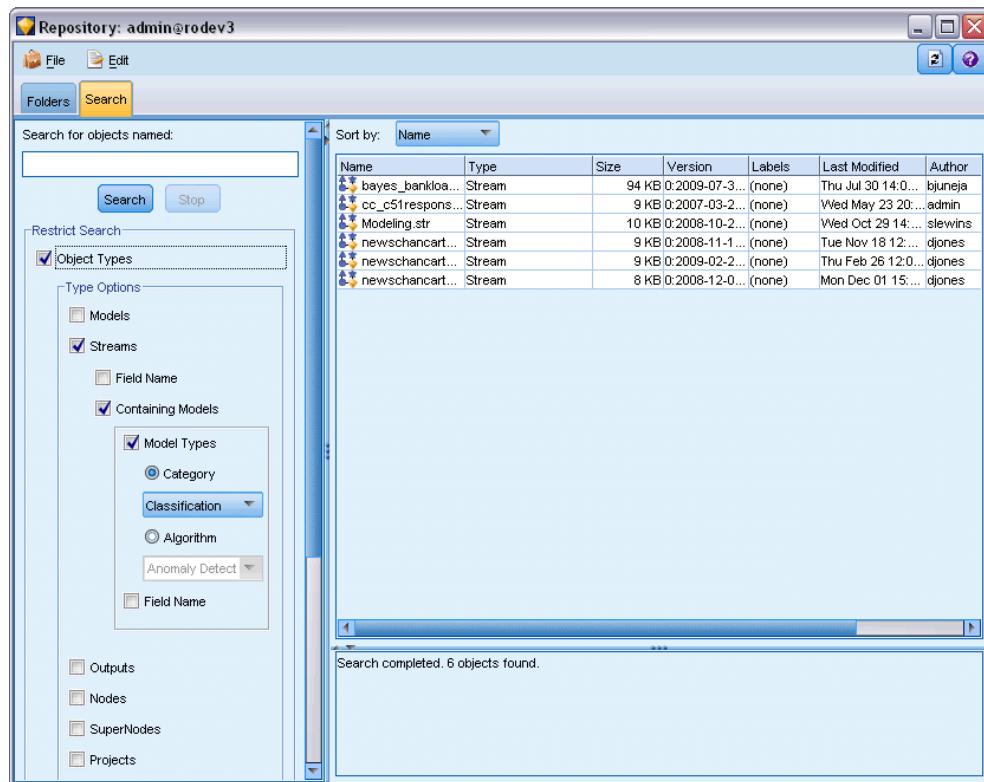
Wenn Sie anhand des Namens nach Objekten suchen, kann ein Sternchen (*) als Platzhalterzeichen für eine beliebige Zeichenkette oder ein Fragezeichen (?) als Platzhalterzeichen für ein einzelnes Zeichen verwendet werden. So werden beispielsweise mit **cluster** alle Objekte gefunden, die irgendwo im Namen die Zeichenfolge *cluster* enthalten. Mit der Suchzeichenkette *m0?_** werden beispielsweise die Namen *M01_cluster.str* und *M02_cluster.str*, nicht jedoch *M01a_cluster.str* gefunden. Bei der Suche wird nicht zwischen Groß- und Kleinschreibung unterschieden (*cluster* entspricht *Cluster* und *CLUSTER*).

Hinweis: Bei einer großen Anzahl von Objekten kann der Suchvorgang einige Zeit in Anspruch nehmen.

Suchen nach anderen Kriterien

Sie können die Suche auf der Basis von Titel, Beschriftung, Datum, Autor, Schlüsselwort, Indexinhalt oder Beschreibung ausführen. Es werden lediglich Objekte gefunden, die *allen* angegebenen Suchkriterien entsprechen. Sie könnten beispielsweise alle Streams suchen, die mindestens ein Clustering-Modell enthalten und auf die außerdem eine bestimmte Beschriftung angewendet wurde und die nach einem bestimmten Datum bearbeitet wurden.

Abbildung 9-15
Suchen nach Streams, die einen bestimmten Modelltyp enthalten



Objekttypen. Sie können die Suche auf Modelle, Streams, Ausgaben, Knoten, Superknoten, Projekte, Modellpaletten, Szenarios oder andere Objekttypen einschränken.

- **Modelle.** Sie können anhand der Kategorie (Klassifizierung, Näherung, Clusterbildung usw.) eines bestimmten Modellierungsalgorithmus, wie beispielsweise Kohonen, nach Modellen suchen.

Sie können die Suche auch nach verwendeten Feldern durchführen – z. B. nach allen Modellen, die ein Feld namens *Einkommen* als Eingabe oder Ausgabe (Ziel) verwenden.

- **Streams.** Für Streams können Sie die Suche nach verwendeten Feldern oder Modelltyp (Kategorie oder Algorithmus) im Stream beschränken.

Themen. Sie können nach Modellen suchen, die mit bestimmten Themen aus einer Liste verknüpft sind und von Repository-Benutzern mit den geeigneten Berechtigungen festgelegt wurden. (Weitere Informationen finden Sie im *Deployment Manager User's Guide*). Um die Liste abzurufen, aktivieren Sie dieses Kontrollkästchen. Klicken Sie dann auf die Schaltfläche "Themen"

hinzufügen”, die eingeblendet wird, wählen Sie ein oder mehrere Themen in der Liste aus und klicken Sie auf “OK”.

Bezeichnung. Beschränkt die Suche auf bestimmte Objektversionsbeschriftungen.

Datumsangaben. Sie können ein Erstellungs- oder Änderungsdatum angeben und nach Objekten vor, nach oder innerhalb des angegebenen Datumsbereichs suchen.

Autor. Beschränkt die Suche auf Objekte, die von einem bestimmten Benutzer erstellt wurden.

Schlüsselwörter. Suche nach bestimmten Schlüsselwörtern. In SPSS Modeler werden Schlüsselwörter auf der Registerkarte “Anmerkung” für einen Stream, ein Modell oder ein Ausgabeobjekt angegeben.

Beschreibung. Suche nach bestimmten Begriffen im Beschreibungsfeld. In SPSS Modeler wird die Beschreibung auf der Registerkarte “Anmerkung” für einen Stream, ein Modell oder ein Ausgabeobjekt angegeben. Mehrere Suchausdrücke können durch Strichpunkte voneinander getrennt werden, beispielsweise Einkommen, Art Feldfruchtyp, Höhe der Subventionen. (Beachten Sie, dass innerhalb von Suchausdrücken, Leerzeichen von Bedeutung sind. Zum Beispiel sind Höhe der Subventionen mit einem Leerzeichen vor “der” und Höhe der Subventionen mit zwei Leerzeichen vor “der” nicht dasselbe.)

Ändern von Repository-Objekten

Sie können bestehende Objekte im Repository direkt von SPSS Modeler aus ändern. Sie verfügen über folgende Möglichkeiten:

- Ordner erstellen, umbenennen oder löschen
- Objekten sperren oder entsperren
- Objekte löschen

Erstellen, Umbenennen und Löschen von Ordnern

- ▶ Um Operationen an Ordner im Repository auszuführen, klicken Sie im SPSS Modeler-Hauptmenü auf:
Werkzeuge > Repository > Explorative Datenanalyse...
- ▶ Geben Sie gegebenenfalls Verbindungseinstellungen zum Repository ein. [Für weitere Informationen siehe Thema Verbinden mit dem Repository auf S. 183](#). Für spezielle Port-, Passwort- und andere Verbindungsdaten wenden Sie sich an den lokalen Systemadministrator.
- ▶ Vergewissern Sie sich, dass die Registerkarte Ordner aktiv ist.
- ▶ Um einen neuen Ordner zu erstellen, klicken Sie mit der rechten Maustaste auf den übergeordneten Ordner und klicken Sie dann auf Neuer Ordner.
- ▶ Um einen Ordner umzubenennen, klicken Sie mit der rechten Maustaste darauf und klicken Sie dann auf Ordner umbenennen.
- ▶ Um einen Ordner zu löschen, klicken Sie mit der rechten Maustaste darauf und klicken Sie dann auf Ordner löschen.

Sperren und Ent sperren von Repository-Objekten

Sie können ein Objekt sperren, um zu verhindern, dass andere Benutzer seine bestehenden Versionen aktualisieren oder neue Versionen erstellen. Ein gesperrtes Objekt wird durch ein Vorhängeschloss-Symbol auf dem Objektsymbol gekennzeichnet.

Abbildung 9-16
Gesperrtes Objekt



So sperren Sie ein Objekt:

- ▶ Klicken Sie im Repository-Explorer-Fenster mit der rechten Maustaste auf das gewünschte Objekt.
- ▶ Klicken Sie auf Sperren.

So entsperren Sie ein Objekt:

- ▶ Klicken Sie im Repository-Explorer-Fenster mit der rechten Maustaste auf das gewünschte Objekt.
- ▶ Klicken Sie auf Ent sperren.

Löschen von Repository-Objekten

Bevor Sie ein Objekt aus dem Repository löschen, müssen Sie entscheiden, ob Sie alle Versionen oder nur eine bestimmte Version des Objekts löschen möchten.

So löschen Sie alle Versionen eines Objekts:

- ▶ Klicken Sie im Repository-Explorer-Fenster mit der rechten Maustaste auf das gewünschte Objekt.
- ▶ Klicken Sie auf Objekte löschen.

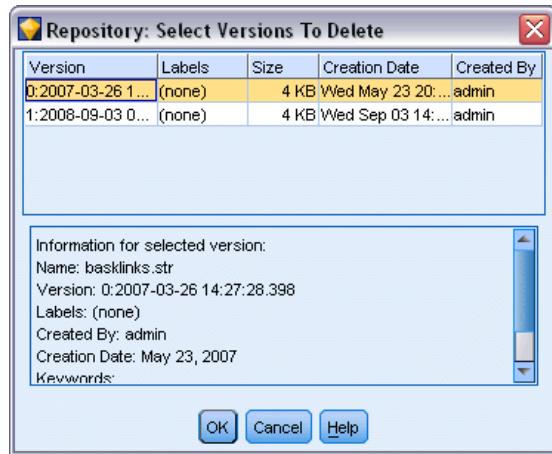
So löschen Sie die aktuellste Version eines Objekts:

- ▶ Klicken Sie im Repository-Explorer-Fenster mit der rechten Maustaste auf das gewünschte Objekt.
- ▶ Klicken Sie auf Löschen.

So löschen Sie eine frühere Version eines Objekts:

- ▶ Klicken Sie im Repository-Explorer-Fenster mit der rechten Maustaste auf das gewünschte Objekt.
- ▶ Klicken Sie auf Versionen löschen.
- ▶ Wählen Sie die zu löschen den Versionen aus und klicken Sie auf OK.

Abbildung 9-17
Auswählen der zu löschen Versionen



Verwalten der Eigenschaften von Repository-Objekten

Sie können zahlreiche Objekteigenschaften von SPSS Modeler aus steuern. Sie verfügen über folgende Möglichkeiten:

- Eigenschaften eines Ordners anzeigen
- Eigenschaften eines Objekts anzeigen und bearbeiten
- Versionsbeschriftungen für ein Objekt erstellen, zuweisen und löschen

Anzeigen von Ordnereigenschaften

Um Eigenschaften für einen beliebigen Ordner im Repository-Fenster anzuzeigen, klicken Sie mit der rechten Maustaste auf den gewünschten Ordner. Klicken Sie auf Ordnereigenschaften.

Registerkarte "Allgemein"

Abbildung 9-18
Ordnereigenschaften

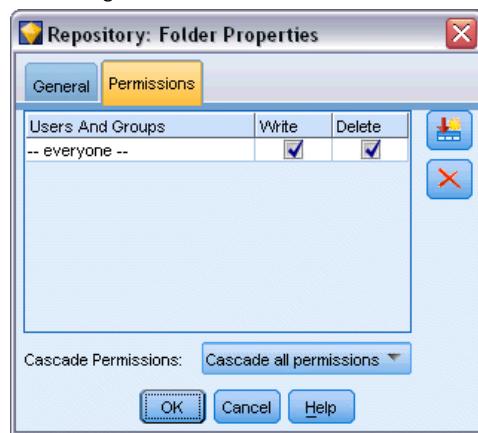


Zeigt den Ordnernamen, das Erstellungs- und Änderungsdatum an.

Registerkarte "Berechtigungen"

Legt die Lese- und Schreibberechtigungen für den Ordner fest. Es werden alle Benutzer und Gruppen mit Zugriff auf den übergeordneten Ordner aufgelistet. Es gibt eine Hierarchie der Berechtigungen. Wenn Sie beispielsweise nicht über eine Leseberechtigung verfügen, können Sie sie auch nicht über eine Schreibberechtigung verfügen. Wenn Sie nicht über eine Schreibberechtigung verfügen, können Sie auch nicht über eine Löschberechtigung verfügen.

Abbildung 9-19
Ordnereigenschaften



Benutzer und Gruppen. Listet die Repository-Benutzer und -Gruppen auf, die mindestens über Leserechte auf diesen Ordner verfügen. Aktivieren Sie die Kontrollkästchen für Schreiben und Löschen, um die entsprechenden Zugriffsrechte für einen bestimmten Benutzer bzw. eine bestimmte Gruppe hinzuzufügen. Klicken Sie auf das Symbol Benutzer/Gruppen hinzufügen rechts auf der Registerkarte "Berechtigungen", um weiteren Benutzern und Gruppen den Zugriff zuzuweisen. Die Liste mit den verfügbaren Benutzern und Gruppen wird vom Administrator gesteuert.

Berechtigungen weiterleiten. Wählen Sie eine Option, um zu steuern, wie Änderungen am aktuellen Ordner seinen untergeordneten Ordnern (falls vorhanden) zugewiesen werden.

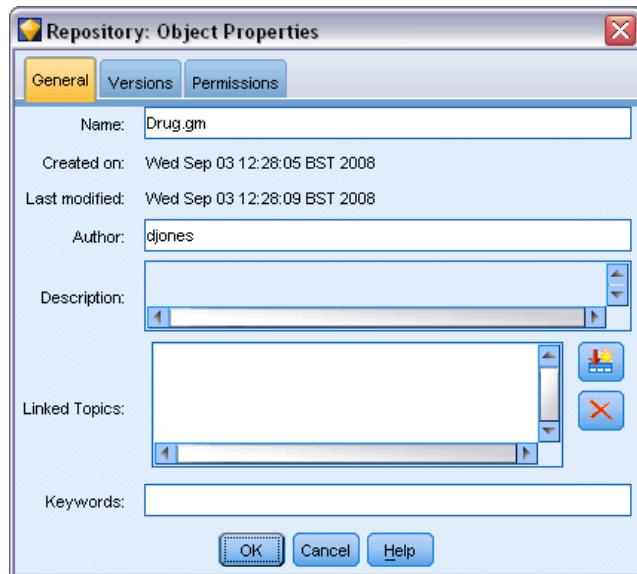
- **Alle Berechtigungen weiterleiten.** Leitet die Berechtigungseinstellungen vom aktuellen Ordner zu allen untergeordneten Ordnern weiter. Dies stellt eine schnelle Möglichkeit dar, Berechtigungen für mehrere Ordner gleichzeitig festzulegen. Legen Sie die Berechtigungen für den übergeordneten Ordner wie gewünscht fest und leiten Sie diese dann nach Bedarf weiter.
- **Nur Änderungen weiterleiten.** Leitet nur Änderungen weiter, die seit den letzten Änderungen vorgenommen wurden. Wenn z. B. eine neue Gruppe hinzugefügt wurde und sie dieser Gruppe Zugriff auf alle Ordner unter "Umsatz" zuweisen möchten, können Sie der Gruppe Zugriff auf den Stammordner "Umsatz" geben und die Änderung an alle Unterordner weiterleiten. Alle anderen Berechtigungen für bereits vorhandene Unterordner bleiben wie zuvor bestehen.
- **Nicht weiterleiten.** Alle vorgenommenen Änderungen gelten lediglich für den aktuellen Ordner und werden nicht an die untergeordneten Ordner weitergeleitet.

Anzeigen und Bearbeiten von Objekteigenschaften

Im Dialogfeld “Objekteigenschaften” können Sie Eigenschaften anzeigen und bearbeiten. Zwar können einige Eigenschaften nicht geändert werden, Sie können jedoch immer ein Objekt aktualisieren, indem Sie eine neue Version hinzufügen.

- ▶ Klicken Sie im Repository-Fenster mit der rechten Maustaste auf das gewünschte Objekt.
- ▶ Klicken Sie auf Objekteigenschaften.

Abbildung 9-20
Objekteigenschaften



Registerkarte “Allgemein”

Name. Der Name des Objekts, wie im Repository angezeigt.

Erstellt am. Datum, an dem das Objekt (nicht die Version) erstellt wurde.

Zuletzt geändert. Datum, an dem die aktuellste Version geändert wurde.

Autor. Der Anmeldename des Benutzers.

Beschreibung. Standardmäßig enthält diese Option die Beschreibung, die in SPSS Modeler auf der Registerkarte “Anmerkung” angegeben ist.

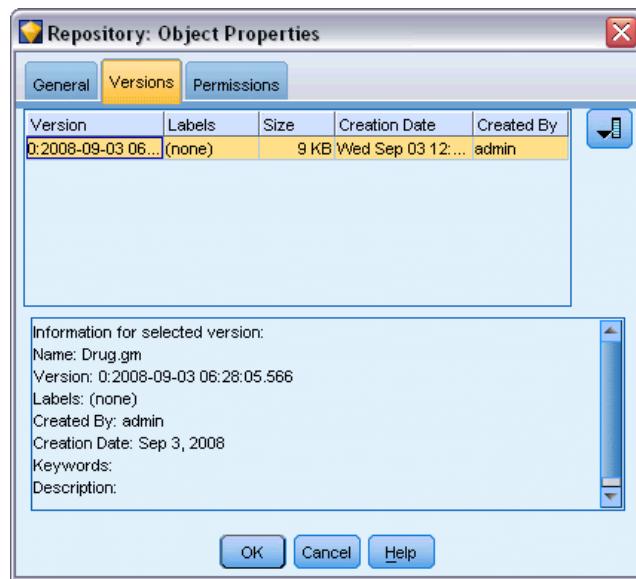
Verknüpfte Themen. Mit dem Repository können Modelle und verwandte Objekte nach Themen geordnet werden. Die Liste verfügbarer Themen wird von Repository-Benutzern mit den geeigneten Berechtigungen festgelegt. (Weitere Informationen finden Sie im *Deployment Manager User's Guide*).

Schlüsselwörter. Schlüsselwörter werden auf der Registerkarte “Anmerkung” für einen Stream, ein Modell oder ein Ausgabeobjekt angegeben. Mehrere Schlüsselwörter sollten durch Leerzeichen voneinander getrennt werden, insgesamt sind bis zu 255 Zeichen zulässig. (Wenn Schlüsselwörter Leerzeichen enthalten, müssen Sie zur Trennung Ausrufezeichen verwenden.)

Registerkarte "Versionen"

Die im Repository gespeicherten Objekte weisen möglicherweise mehrere Versionen auf. Auf der Registerkarte "Versionen" werden Informationen über jede Version angezeigt.

Abbildung 9-21
Versionseigenschaften



Die folgenden Eigenschaften können für bestimmte Versionen eines gespeicherten Objekts angegeben oder geändert werden:

Version. Eindeutiger Bezeichner für die Version, generiert basierend auf der Uhrzeit, zu der die Version gespeichert wurde.

Bezeichnung. Aktuelle Beschriftung für die Version, falls vorhanden. Im Gegensatz zum Versionsbezeichner können die Beschriftungen von einer Version eines Objekts zu einer anderen verschoben werden.

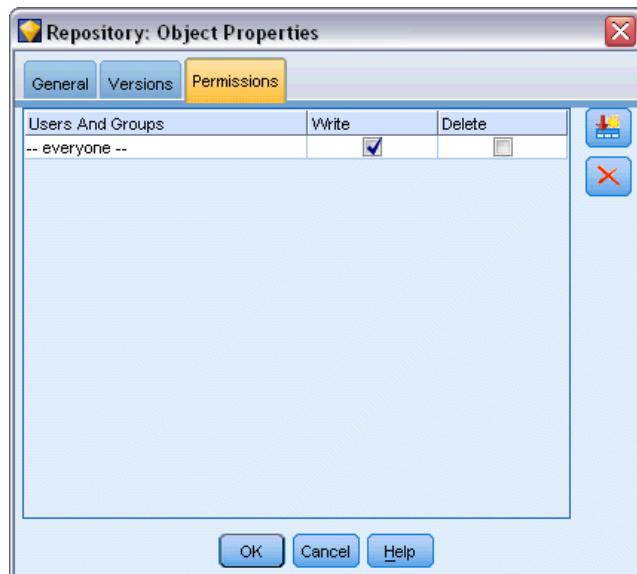
Die Dateigröße, das Erstellungsdatum und der Autor werden ebenfalls für jede Version angezeigt.

Labels bearbeiten. Klicken Sie auf das Symbol Labels bearbeiten rechts oben in der Registerkarte "Versionen", um Beschriftungen für gespeicherte Objekte zu definieren, anzuwenden oder zu entfernen. Für weitere Informationen siehe Thema [Verwalten von Objektversionsbeschriftungen auf S. 206](#).

Registerkarte "Berechtigungen"

Auf der Registerkarte Berechtigungen können Sie Lese- und Schreibberechtigungen für das Objekt festlegen. Es werden alle Benutzer und Gruppen mit Zugriff auf das aktuelle Objekt aufgelistet. Es gibt eine Hierarchie der Berechtigungen. Wenn Sie beispielsweise nicht über eine Leseberechtigung verfügen, können Sie sie auch nicht über eine Schreibberechtigung verfügen. Wenn Sie nicht über eine Schreibberechtigung verfügen, können Sie auch nicht über eine Löschberechtigung verfügen.

Abbildung 9-22
Objektzugriffsrechte



Benutzer und Gruppen. Listet die Repository-Benutzer und -Gruppen auf, die mindestens über Leserechte auf dieses Objekt verfügen. Aktivieren Sie die Kontrollkästchen für Schreiben und Löschen, um die entsprechenden Zugriffsrechte für einen bestimmten Benutzer bzw. eine bestimmte Gruppe hinzuzufügen. Klicken Sie auf das Symbol Benutzer/Gruppen hinzufügen rechts auf der Registerkarte „Berechtigungen“, um weiteren Benutzern und Gruppen den Zugriff zuzuweisen. Die Liste mit den verfügbaren Benutzern und Gruppen wird vom Administrator gesteuert.

Verwalten von Objektversionsbeschriftungen

Im Dialogfeld „Versionslabel bearbeiten“ haben Sie folgende Möglichkeiten:

- Anwenden von Beschriftungen auf das ausgewählte Objekt
- Entfernen von Beschriftungen vom ausgewählten Objekt
- Definieren einer neuen Beschriftung und Anwenden auf das Objekt

So wenden Sie Beschriftungen auf ein Objekt an:

- ▶ Wählen Sie in der Liste Verfügbare Labels eine oder mehrere Beschriftungen aus.
- ▶ Klicken Sie auf die Schaltfläche mit dem Rechtspfeil, um die ausgewählten Beschriftungen in die Liste Angewendete Labels zu verschieben.
- ▶ Klicken Sie auf OK.

So entfernen Sie Beschriftungen von einem Objekt:

- ▶ Wählen Sie in der Liste Angewendete Labels eines oder mehrere Labels aus.

- ▶ Klicken Sie auf die Schaltfläche mit dem Linkspfeil, um die ausgewählten Beschriftungen in die Liste Verfügbare Labels zu verschieben.
- ▶ Klicken Sie auf OK.

So definieren Sie eine neue Beschriftung und wenden sie auf das Objekt an:

- ▶ Geben Sie den Beschriftungsnamen in das Feld Neues Label ein.
- ▶ Klicken Sie auf die Schaltfläche mit dem Rechtspfeil, um die neue Beschriftung in die Liste Angewendete Labels zu verschieben.
- ▶ Klicken Sie auf OK.

Bereitstellen von Streams

Damit ein Stream mit der Thin-Client-Anwendung IBM® SPSS® Modeler Advantage benutzt werden kann, muss er als Stream (.str-Datei) im Repository bereitgestellt werden.

Unabhängig davon, ob ein Stream als Stream (.str-Datei) oder als Szenario (.scn-Datei) bereitgestellt wird, kann das Objekt die Unternehmensfunktionen von IBM® SPSS® Collaboration and Deployment Services vollständig nutzen. Für weitere Informationen siehe Thema Speichern und Bereitstellen von Repository-Objekten auf S. 182.

So stellen Sie den aktuellen Stream bereit (Verfahren mit dem Menü “Datei”):

- ▶ Klicken Sie im Hauptmenü auf Folgendes:
Datei > Speichern > Bereitstellen
- ▶ Wählen Sie den Bereitstellungstyp und füllen Sie das übrige Dialogfeld wie erforderlich aus.
- ▶ Klicken Sie auf Als Stream bereitstellen, um den Stream für die Verwendung mit IBM SPSS Modeler Advantage oder IBM SPSS Collaboration and Deployment Services bereitzustellen. Klicken Sie auf Als Szenario bereitstellen, um den Stream für die Verwendung mit IBM SPSS Collaboration and Deployment Services oder Predictive Applications Version 5 bereitzustellen.
- ▶ Klicken Sie auf Speichern. Wenn Sie Informationen wünschen, klicken Sie auf Hilfe.
- ▶ Fahren Sie mit “Abschließen des Bereitstellungsvorgangs” fort.

So stellen Sie den aktuellen Stream bereit (Verfahren mit dem Menü “Extras”):

- ▶ Klicken Sie im Hauptmenü auf Folgendes:
Werkzeuge > Stream-Eigenschaften > Bereitstellung
- ▶ Wählen Sie den Bereitstellungstyp, füllen Sie den Rest der Registerkarte “Bereitstellung” wie erforderlich aus und klicken Sie auf Speichern. Für weitere Informationen siehe Thema Deployment-Optionen für den Stream auf S. 208.

Abschließen des Bereitstellungsvorgangs

- Geben Sie gegebenenfalls Verbindungseinstellungen zum Repository ein. Für weitere Informationen siehe Thema [Verbinden mit dem Repository](#) auf S. 183. Für spezielle Port-, Passwort- und andere Verbindungsdaten wenden Sie sich an den lokalen Systemadministrator.

Abbildung 9-23
Speichern eines Streams im Repository



- Wählen Sie im Dialogfeld “Repository: Speichern” den Ordner aus, in dem Sie das Objekt speichern möchten, geben Sie alle weiteren Informationen an, die Sie aufzeichnen möchten, und klicken Sie auf die Schaltfläche Speichern. Für weitere Informationen siehe Thema [Festlegen der Objekteigenschaften](#) auf S. 186.

Deployment-Optionen für den Stream

Auf der Registerkarte “Bereitstellung” im Dialogfeld “Stream-Optionen” können Sie Optionen für die Bereitstellung des Streams festlegen. Die Bereitstellung ist als Stream oder als Szenario möglich.

Beim Bereitstellen als Stream können Sie den Stream in der Thin-Client-Anwendung IBM® SPSS® Modeler Advantage öffnen und ändern. Der Stream wird im Repository als Datei mit der Erweiterung *.str* gespeichert.

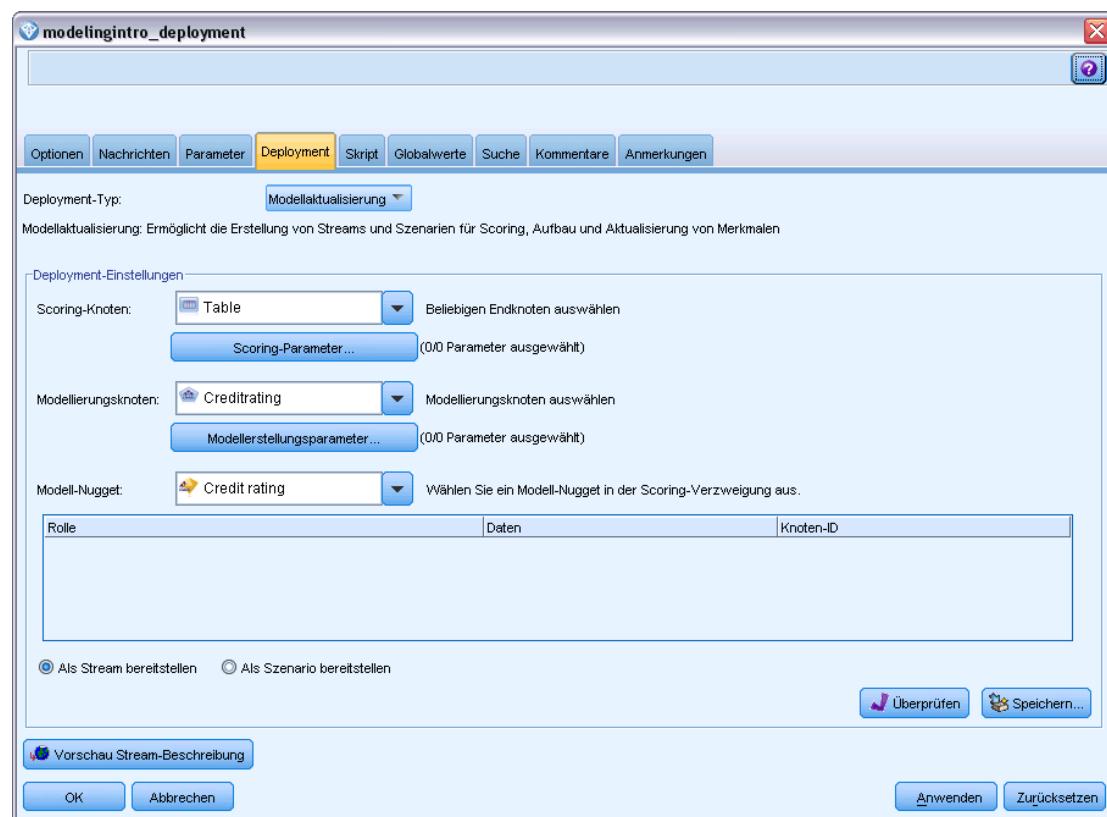
Beim Bereitstellen als Szenario wird der Stream im Repository als Datei mit der Erweiterung *.scn* gespeichert. Damit kann ein Stream auch von Predictive Applications Version 5 verwendet werden.

Unabhängig davon, ob die Bereitstellung als Stream oder als Szenario erfolgt, können Sie die zusätzliche Funktionalität nutzen, die in IBM® SPSS® Collaboration and Deployment Services zur Verfügung steht, z. B. Mehrbenutzerzugriff, automatisches Scoring, Modellaktualisierung und Champion-Challenger-Analyse.

Über die Registerkarte “Bereitstellung” können Sie auch eine Vorschau der Stream-Beschreibung anzeigen, die IBM® SPSS® Modeler für den Stream erstellt. Für weitere Informationen siehe Thema [Stream-Beschreibungen in Kapitel 5 auf S. 83](#).

Hinweis: Um konsistenten Zugriff auf Unternehmensdaten sicherzustellen, muss ein Stream, der als Szenario bereitgestellt ist, über IBM® SPSS® Collaboration and Deployment Services Enterprise View auf seine Quelldaten zugreifen. In einem solchen Fall muss der Stream in jeder angegebenen Scoring- bzw. Modellierungsverzweigung mindestens einen Quellenknoten vom Typ “Enterprise-Ansicht” enthalten. Für weitere Informationen siehe Thema [Enterprise-Ansichts-Knoten in Kapitel 2 in IBM SPSS Modeler 15 Quellen-, Prozess- und Ausgabeknoten](#).

Abbildung 9-24
Deployment-Optionen für den Stream



Bereitstellungstyp. Wählen Sie, wie der Stream bereitgestellt werden soll. Alle Streams benötigen einen speziellen Scoring-Knoten, bevor sie bereitgestellt werden können; weitere Voraussetzungen und Optionen hängen vom Bereitstellungstyp ab.

- <keine>. Der Stream wird nicht für das Repository bereitgestellt. Alle Optionen mit Ausnahme der Vorschau der Stream-Beschreibung sind deaktiviert.

- Nur Scoring. Der Stream wird dem Repository bereitgestellt, wenn Sie auf die Schaltfläche Speichern klicken. Daten können mithilfe des Knotens gescort werden, den Sie im Feld Scoring-Knoten angeben.
- Modellaktualisierung. Wie "Nur Scoring", jedoch kann das Modell zusätzlich im Repository mit Objekten aktualisiert werden, die Sie in den Feldern Modellierungsknoten und Modell-Nugget festlegen.

Scoring-Knoten. Wählen Sie ein Diagramm, einen Ausgabe- oder einen Exportknoten, um die Stream-Verzweigung zu identifizieren, die zum Scoring der Daten benutzt werden soll. Der Stream kann zwar eine beliebige Anzahl an gültigen Verzweigungen, Modellen und Endknoten enthalten, zu Deployment-Zwecken muss jedoch genau eine Scoring-Verzweigung angegeben werden. Dies ist die grundlegendste Voraussetzung für das Deployment eines Streams.

Scoring-Parameter. Dient zur Angabe von Parametern, die bei der Ausführung der Scoring-Verzweigung geändert werden können. Für weitere Informationen siehe Thema [Scoring und Modellierung von Parametern auf S. 211](#).

Modellierungsknoten. Gibt für die Modellaktualisierung den Modellierungsknoten an, der zum Generieren bzw. zur Aktualisierung des Modells im Repository verwendet wird. Dies muss ein Modellierungsknoten desselben Typs sein wie für Modell-Nugget angegeben wurde.

Modellerstellungsparameter. Dient zur Angabe von Parametern, die bei der Ausführung des Modellierungsknotens geändert werden können. Für weitere Informationen siehe Thema [Scoring und Modellierung von Parametern auf S. 211](#).

Modell-Nugget. Gibt zum Zwecke der Modellaktualisierung das Modell-Nugget an, das bei jeder Aktualisierung des Streams im Repository aktualisiert bzw. neu generiert werden soll (üblicherweise im Rahmen einer geplanten Aufgabe). Das Modell muss sich in der Scoring-Verzweigung befinden. In der Scoring-Verzweigung können mehrere Modelle vorhanden sein, es kann jedoch nur eines angegeben werden. Beachten Sie: Bei der erstmaligen Erstellung des Streams kann es sich hierbei im Grunde um ein Platzhaltermodell handeln, das immer dann aktualisiert bzw. neu generiert wird, wenn neue Daten verfügbar sind.

Als Stream bereitstellen. Klicken Sie auf diese Option, wenn Sie den Stream mit IBM SPSS Modeler Advantage oder IBM SPSS Collaboration and Deployment Services verwenden möchten (siehe auch den folgenden Hinweis).

Als Szenario bereitstellen. Klicken Sie auf diese Option, wenn Sie den Stream mit IBM SPSS Collaboration and Deployment Services oder Predictive Applications Version 5 verwenden möchten (siehe auch den folgenden Hinweis).

Überprüfen. Klicken Sie auf diese Schaltfläche, um zu prüfen, ob es sich um einen gültigen Stream zur Bereitstellung handelt. Alle Streams müssen über einen speziellen Scoring-Knoten verfügen, bevor sie bereitgestellt werden können. Wenn Sie die Bereitstellung als Szenario vornehmen, muss der Stream außerdem einen gültigen Enterprise View-Quellenknoten enthalten. Wenn diese Bedingungen nicht erfüllt sind, werden Fehlermeldungen angezeigt.

Speichern. Stellt den Stream bereit, sofern dieser gültig ist. Andernfalls wird eine Fehlermeldung angezeigt. Klicken Sie auf die Schaltfläche Korrigieren, um den Fehler zu korrigieren und es erneut zu versuchen.

Vorschau der Stream-Beschreibung anzeigen. Ermöglicht die Anzeige des Inhalts einer Stream-Beschreibung, die SPSS Modeler für den Stream erstellt. Für weitere Informationen siehe Thema Stream-Beschreibungen in Kapitel 5 auf S. 83.

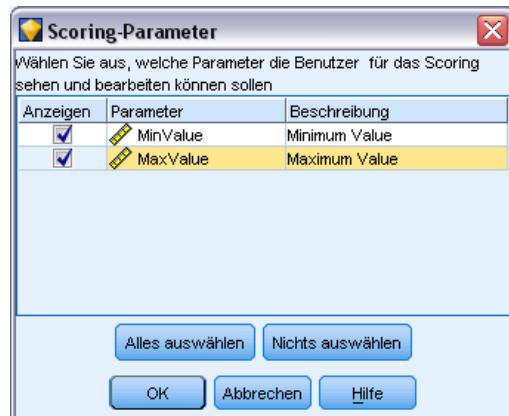
Hinweis: (Bereitstellung als Stream oder als Szenario) Innerhalb der Modellierungsverzweigung können mehrere Enterprise View-Knoten verwendet werden. Wenn ja, sollte in den meisten Fällen eine einzelne Datenverbindung für alle Enterprise-Ansichts-Knoten innerhalb der Verzweigung verwendet werden; für Champion/Challenger-Analyse ist sie sogar erforderlich.

- Wenn keine Champion/Challenger-Unterstützung benötigt wird, können innerhalb derselben Verzweigung verschiedene Enterprise View-Verbindungen verwendet werden, vorausgesetzt, die Verbindungen unterscheiden sich nur nach Daten-Provider-Definition (DPD).
- Diese Einschränkungen gelten jeweils nur innerhalb einer bestimmten Verzweigung. Zwischen den Scoring- und den Modellerstellungsverzweigungen können verschiedene Enterprise View-Verbindungen ohne derartige Einschränkungen verwendet werden.

Scoring und Modellierung von Parametern

Beim Deployment eines Streams für IBM SPSS Collaboration and Deployment Services können Sie auswählen, welche Parameter angezeigt bzw. bearbeitet werden können, wenn das Modell aktualisiert oder gescort wird. Sie können beispielsweise Mindest- und Höchstwerte oder einen anderen Wert angeben, der sich bei jeder Ausführung eines Auftrags ändern kann.

Abbildung 9-25
Dialogfeld "Scoring-Parameter"



- Um einen Parameter sichtbar zu machen, sodass er nach dem Deployment des Streams angezeigt bzw. bearbeitet werden kann, müssen Sie ihn aus der Liste im Dialogfeld auswählen.

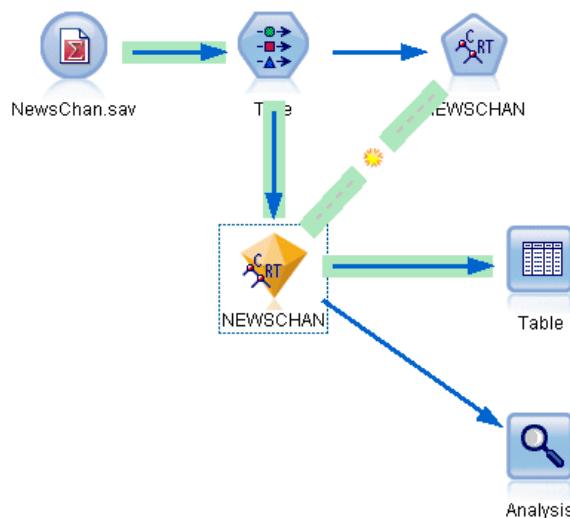
Die Liste der verfügbaren Parameter wird im Dialogfeld für die Stream-Eigenschaften auf der Registerkarte "Parameter" definiert. Für weitere Informationen siehe Thema Festlegen der Stream- und Sitzungsparameter in Kapitel 5 auf S. 76.

Scoring-Verzweigung

Wenn Sie einen Stream bereitstellen, muss eine Verzweigung des Streams als **Scoring-Verzweigung** (d. h. die Verzweigung, die den Scoring-Knoten enthält) festgelegt werden. Wenn Sie eine Verzweigung als Scoring-Verzweigung festlegen, wird in der Stream-Zeichenfläche diese Verzweigung sowie die Modellverknüpfung zum Nugget auf der Scoring-Verzweigung hervorgehoben. Die visuelle Repräsentation ist besonders nützlich in komplexen Streams mit mehreren Verzweigungen, in denen die Scoring-Verzweigung nicht sofort erkennbar ist.

Hinweis: Nur eine Stream-Verzweigung kann als Scoring-Verzweigung festgelegt werden.

Abbildung 9-26
Stream mit hervorgehobener Scoring-Verzweigung



Wenn für den Stream bereits eine Scoring-Verzweigung definiert wurde, wird diese durch die neu festgelegte Verzweigung als Scoring-Verzweigung ersetzt. Sie können die Farbe der Scoring-Verzweigung mithilfe einer Option für benutzerdefinierte Farben festlegen. Für weitere Informationen siehe Thema Festlegen von Anzeigoptionen in Kapitel 12 auf S. 246.

Sie können die Anzeige der Scoring-Verzweigung mithilfe der Symbolleistenschaltfläche zum Ein-/Ausblenden von Stream-Markups ein- und ausblenden.

Abbildung 9-27
Symbolleistenschaltfläche zum Ein-/Ausblenden von Stream-Markups



Identifizieren der Scoring-Verzweigung für die Bereitstellung

Sie können die Scoring-Verzweigung über das Popup-Menü eines Endknotens oder über das Menü “Extras” festlegen. Bei Verwendung des Popup-Menüs wird der Scoring-Knoten automatisch in der Registerkarte “Bereitstellung” der Stream-Eigenschaften festgelegt.

So legen Sie eine Verzweigung als die Scoring-Verzweigung fest (Popup-Menü):

- ▶ Verbinden Sie das Modell-Nugget mit einem Endknoten (einem Verarbeitungs- oder Ausgabeknoten unterhalb des Nuggets).
- ▶ Klicken Sie mit der rechten Maustaste auf den Endknoten.
- ▶ Klicken Sie im Menü auf die Option Als Scoring-Verzweigung verwenden.

So legen Sie eine Verzweigung als die Scoring-Verzweigung fest (Menü “Extras”):

- ▶ Verbinden Sie das Modell-Nugget mit einem Endknoten (einem Verarbeitungs- oder Ausgabeknoten unterhalb des Nuggets).
- ▶ Klicken Sie im Hauptmenü auf Folgendes:
Werkzeuge > Stream-Eigenschaften > Bereitstellung
- ▶ Klicken Sie in der Liste Bereitstellungstyp wie gewünscht auf die Option Nur Scoring oder Modellaktualisierung. [Für weitere Informationen siehe Thema Deployment-Optionen für den Stream auf S. 208.](#)
- ▶ Klicken Sie auf das Feld Scoring-Knoten und wählen Sie einen Endknoten aus der Liste.
- ▶ Klicken Sie auf OK.

Modellaktualisierung

Eine Modellaktualisierung ist der Vorgang, bei dem ein vorhandenes Modell in einem Stream anhand von neueren Daten neu aufgebaut wird. Der Stream selbst ändert sich im Repository nicht. Beispielsweise bleiben Algorithmustyp und Stream-spezifische Einstellungen gleich, aber das Modell wird mit neuen Daten neu trainiert und dann aktualisiert, wenn die neue Version des Modells besser als die alte Version funktioniert.

Nur ein Modell-Nugget in einem Stream kann auf Aktualisierung eingestellt werden – dies ist das so genannte **Aktualisierungsmodell**. Wenn Sie auf die Option Modellaktualisierung in der Registerkarte “Bereitstellung” klicken (siehe Deployment-Optionen für den Stream auf S. 208), wird das zu diesem Zeitpunkt festgelegte Modell-Nugget zu Ihrem Aktualisierungsmodell. Sie können auch aus dem Popup-Menü eines Modell-Nuggets ein Modell als Aktualisierungsmodell festlegen. Das Nugget muss sich dafür bereits auf der Scoring-Verzweigung befinden.

Wenn Sie den Modellaktualisierungsstatus eines Nuggets deaktivieren, hat dies dieselbe Wirkung wie das Einstellen des Bereitstellungstyps des Streams auf “Nur Scoring” und die Registerkarte “Bereitstellung” des Dialogfelds für die Stream-Eigenschaften wird entsprechend aktualisiert. Sie

können diesen Status mit der Option Als Aktualisierungsmodell verwenden im Popup-Menü des Nuggets auf der aktuellen Scoring-Verzweigung ein- und ausschalten.

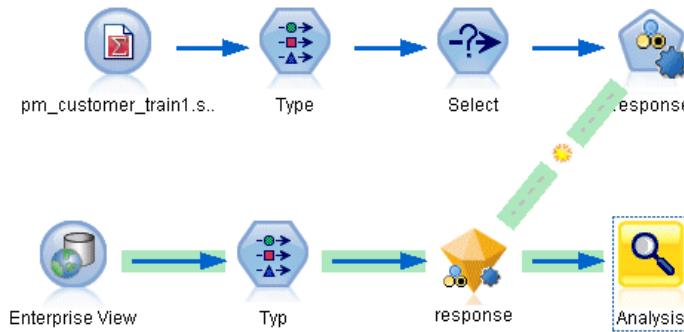
Beim Entfernen der Modellverknüpfung auf der Scoring-Verzweigung wird auch der Aktualisierungsmodell-Status des Nuggets entfernt. Sie können das Entfernen der Modellverknüpfung über das Menü „Bearbeiten“ oder die Symbolleiste widerrufen, aber damit wird auch der „Aktualisierungsmodell“-Status wieder eingestellt.

Auswählen des Aktualisierungsmodells

Wie die Scoring-Verzweigung wird auch die Verknüpfung zum Aktualisierungsmodell im Stream hervorgehoben. Das als Aktualisierungsmodell gewählte Modell-Nugget und die hervorgehobene Verknüpfung hängen davon ab, wie viele Nuggets sich im Stream befinden.

Einzelnes Modell im Stream

Abbildung 9-28
Scoring-Verzweigung mit Einzelmodell im Stream



Wenn ein einzelnes, verknüpftes Modell-Nugget auf einer Scoring-Verzweigung als solches identifiziert wird, wird dieses Nugget das Aktualisierungsmodell des Streams.

Mehrere Modelle im Stream

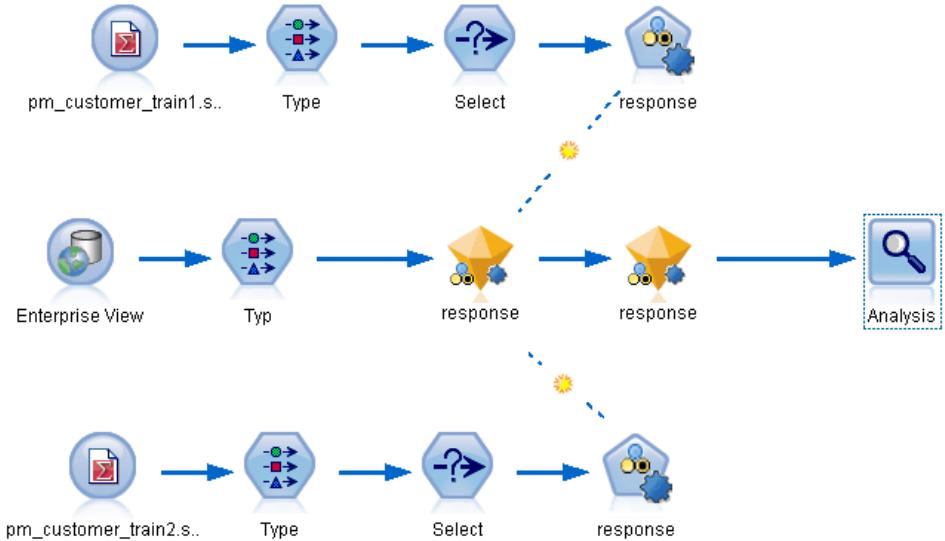
Falls der Stream mehrere verknüpfte Nuggets enthält, wird das Aktualisierungsmodell wie folgt ausgewählt.

Wenn ein Modell-Nugget in der Registerkarte „Bereitstellung“ des Dialogfelds für die Stream-Eigenschaften definiert wurde und sich im Stream befindet, wird dieses Nugget zum Aktualisierungsmodell.

Wenn in der Registerkarte „Bereitstellung“ kein Nugget definiert wurde oder wenn einer definiert wurde, sich aber nicht auf der Scoring-Verzweigung befindet, wird das Nugget, das sich am nächsten zum Endknoten befindet, zum Aktualisierungsmodell.

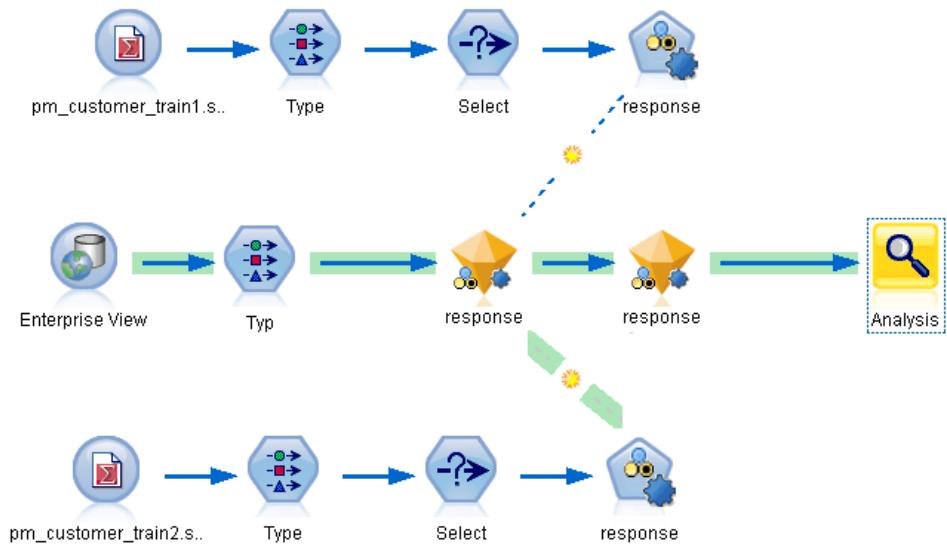
Nehmen Sie zur Illustration an, es liegt folgender Stream vor.

Abbildung 9-29
Scoring-Verzweigung mit mehreren Modellen im Stream



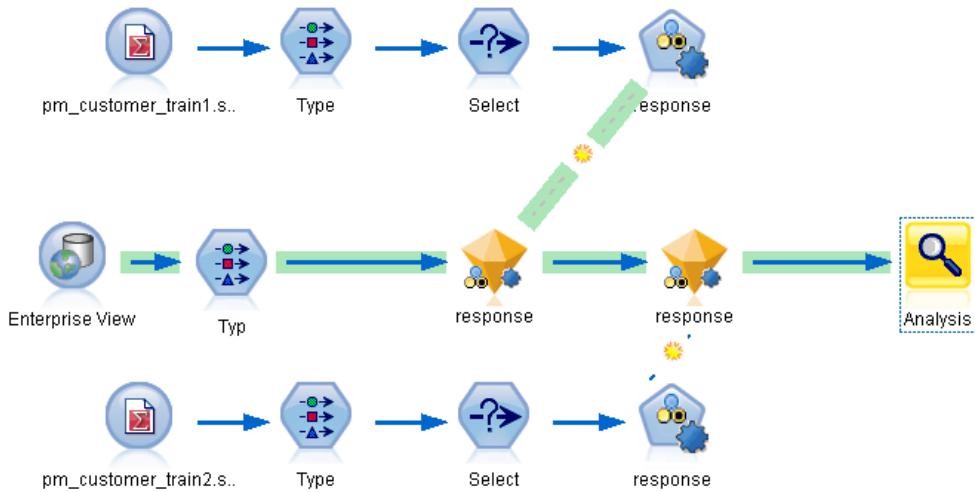
Sie klicken mit der rechten Maustaste auf den Analyseknoten und legen in seinem Menü die Scoring-Verzweigung fest, die nun hervorgehoben wird. Damit wird auch das Modell, das sich am nächsten zum Analyseknoten befindet, als Aktualisierungsmodell festgelegt, wie durch die hervorgehobene Aktualisierungsverknüpfung gezeigt.

Abbildung 9-30
Hervorgehobene Scoring-Verzweigung mit mehreren Modellen und Aktualisierungsverknüpfung



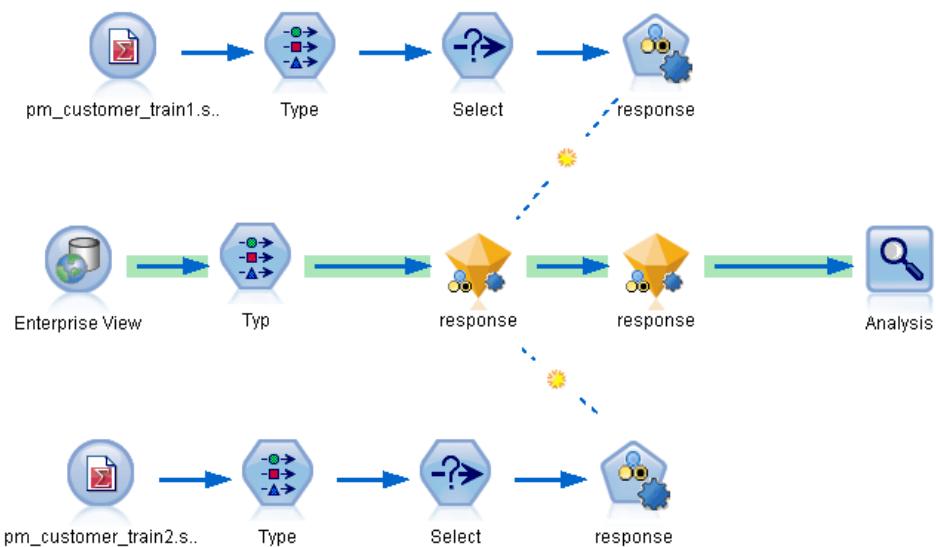
Sie beschließen jedoch, das andere Modell-Nugget im Stream als Aktualisierungsmodell zu verwenden, deshalb legen Sie über sein Menü fest, dass seine Modellverknüpfung als Aktualisierungsverknüpfung verwendet werden soll.

Abbildung 9-31
Scoring-Verzweigung mit Aktualisierungsverknüpfung auf erstes Modell-Nugget geändert



Wenn Sie anschließend die Auswahl beider Modellverknüpfungen als Aktualisierungsverknüpfungen aufheben, wird nur die Scoring-Verzweigung hervorgehoben, nicht die Verknüpfungen. Der Bereitstellungstyp wird auf "Nur Scoring" eingestellt.

Abbildung 9-32
Scoring-Verzweigung mit mehreren Modellen und ohne Aktualisierungsverknüpfungen



Hinweis: Sie können nach Wunsch eine der Verknüpfungen auf den Status “Ersetzen” einstellen, aber nicht die andere. In diesem Fall ist das als Aktualisierungsmodell gewählte Modell-Nugget dasjenige, das über die Aktualisierungsverknüpfung verfügt und sich am nächsten zum Endknoten befindet, wenn die Scoring-Verzweigung festgelegt wird.

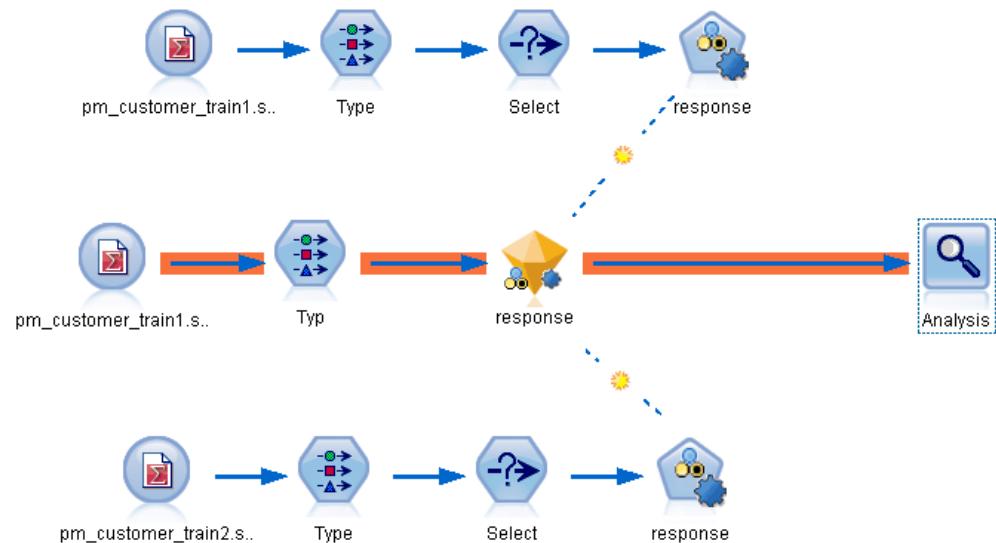
Keine Modelle im Stream

Wenn der Stream keine Modelle oder nur Modelle ohne Modellverknüpfungen enthält, ist der Bereitstellungstyp auf “Nur Scoring” eingestellt.

Überprüfen einer Scoring-Verzweigung auf Fehler

Wenn Sie die Scoring-Verzweigung festlegen, wird sie auf Fehler überprüft, z. B. bei der Bereitstellung als Szenario auf das Fehlen eines Enterprise-Ansichts-Knotens im Stream.

Abbildung 9-33
Scoring-Verzweigung mit Fehlern



Bei Auffinden eines Fehlers wird die Scoring-Verzweigung in der entsprechenden Fehlerfarbe markiert und eine Fehlermeldung angezeigt. Sie können die Fehlerfarbe mithilfe einer Option für benutzerdefinierte Farben festlegen. Für weitere Informationen siehe Thema [Festlegen von Anzeigeeoptionen in Kapitel 12 auf S. 246](#).

Gehen Sie bei Auffinden eines Fehlers wie folgt vor:

- ▶ Korrigieren Sie den Fehler gemäß dem Inhalt der Fehlermeldung.

- Führen Sie eine der folgenden Aktionen aus:
 - Klicken Sie mit der rechten Maustaste auf den Endknoten und klicken Sie im Popup-Menü auf Szenario prüfen.
 - Klicken Sie im Hauptmenü auf Folgendes:
Werkzeuge > Stream-Eigenschaften > Bereitstellung
und klicken Sie auf Überprüfen.
- Wiederholen Sie diesen Vorgang, bis keine Fehler mehr gefunden werden.

Exportieren in externe Anwendungen

IBM® SPSS® Modeler bietet eine Reihe von Mechanismen zum Exportieren des gesamten Data-Mining-Prozesses in externe Anwendungen, damit Sie die durchgeföhrten Arbeiten zur Vorbereitung von Daten und zum Erstellen von Modellen auch außerhalb von SPSS Modeler gewinnbringend einsetzen können.

Im vorherigen Abschnitt wurde gezeigt, wie Sie Streams für ein IBM SPSS Collaboration and Deployment Services-Repository bereitstellen können, um dessen Mehrbenutzerzugriff, Aufgabenplanung und andere Funktionen zu nutzen. Ebenso lassen sich SPSS Modeler-Streams zusammen mit folgenden Anwendungen nutzen:

- IBM® SPSS® Modeler Advantage
- Predictive Applications 5.0-Anwendungen
- Anwendungen, die Dateien im PMML-Format importieren und exportieren können

Weitere Informationen über das Verwenden von Streams in IBM SPSS Modeler Advantage finden Sie unter Öffnen eines Streams in IBM SPSS Modeler Advantage auf S. 219.

Zum Exportieren eines Streams für die Verwendung mit Predictive Applications 5.0 folgen Sie den Anleitungen für die Bereitstellung als Szenario. [Für weitere Informationen siehe Thema Bereitstellen von Streams in Kapitel 9 auf S. 207.](#)

Informationen zum Exportieren und Importieren von Modellen als PMML-Dateien, um Modelle mit beliebigen anderen Anwendungen, die dieses Format unterstützen, gemeinsam zu nutzen, finden Sie unter Importieren und Exportieren von Modellen als PMML auf S. 220.

Hinweis: Das Produkt Predictive Applications wurde durch IBM® Analytical Decision Management ersetzt. Die Unterstützung für Predictive Applications wird in einer zukünftigen Version von SPSS Modeler aufgehoben.

Öffnen eines Streams in IBM SPSS Modeler Advantage

IBM® SPSS® Modeler-Streams können zusammen mit der Thin-Client-Anwendung IBM® SPSS® Modeler Advantage verwendet werden. Es ist zwar möglich, benutzerdefinierte Anwendungen komplett in IBM SPSS Modeler Advantage zu erstellen, jedoch können Sie auch einen bereits in SPSS Modeler erstellten Stream als Basis eines Anwendungsablaufs verwenden.

So öffnen Sie einen Stream in IBM SPSS Modeler Advantage:

- ▶ Stellen Sie den Stream im IBM® SPSS® Collaboration and Deployment Services-Repository bereit, wobei die Option Als Stream bereitstellen aktiviert sein muss. [Für weitere Informationen siehe Thema Bereitstellen von Streams in Kapitel 9 auf S. 207.](#)

- ▶ Klicken Sie in der Symbolleiste auf die Schaltfläche In IBM SPSS Modeler Advantage öffnen oder klicken Sie im Hauptmenü auf:
Datei > In IBM SPSS Modeler Advantage öffnen

- ▶ Geben Sie gegebenenfalls Verbindungseinstellungen zum Repository ein. [Für weitere Informationen siehe Thema Verbinden mit dem Repository in Kapitel 9 auf S. 183.](#) Für spezielle Port-, Passwort- und andere Verbindungsdaten wenden Sie sich an den lokalen Systemadministrator.

Hinweis: Auf dem Repository-Server muss ebenfalls die IBM SPSS Modeler Advantage-Software installiert sein.

- ▶ Wählen Sie im Dialogfeld “Repository: Speichern” den Ordner aus, in dem Sie das Objekt speichern möchten, geben Sie alle weiteren Informationen an, die Sie aufzeichnen möchten, und klicken Sie auf die Schaltfläche Speichern. [Für weitere Informationen siehe Thema Festlegen der Objekteigenschaften in Kapitel 9 auf S. 186.](#)

Dadurch wird IBM SPSS Modeler Advantage gestartet, wobei der Stream bereits geöffnet ist. Der Stream wird in SPSS Modeler geschlossen.

Importieren und Exportieren von Modellen als PMML

PMML (Predictive Model Markup Language) ist ein XML-Format zur Beschreibung von Data Mining-Modellen und statistischen Modellen, einschließlich der Eingaben zu den Modellen, der zur Vorbereitung der Daten für das Data Mining verwendeten Transformationen sowie der Parameter, die die Modelle selbst definieren. IBM® SPSS® Modeler kann PMML importieren und exportieren, wodurch es möglich ist, Modelle mit anderen Anwendungen auszutauschen, die dieses Format unterstützen, zum Beispiel IBM® SPSS® Statistics.

Hinweis: Für den Export von PMML ist eine IBM® SPSS® Modeler Server-Lizenz erforderlich.

Weitere Informationen zu PMML finden Sie auf der Website der Data Mining Group (<http://www.dmg.org>).

So exportieren Sie ein Modell:

PMML-Export wird für die meisten der in SPSS Modeler erstellten Modelltypen unterstützt. [Für weitere Informationen siehe Thema Modelltypen, die PMML unterstützen auf S. 223.](#)

- ▶ Klicken Sie mit der rechten Maustaste auf ein Modell-Nugget in der Modellpalette. (Alternativ können Sie auf ein Modell-Nugget im Zeichenbereich klicken und das Menü “Datei” auswählen.)
- ▶ Klicken Sie im Menü auf PMML exportieren.

Abbildung 10-1
Exportieren von Modellen im PMML-Format



- ▶ Geben Sie im Dialogfeld “Exportieren” (oder “Speichern”) ein Zielverzeichnis und einen eindeutigen Namen für das Modell an.

Hinweis: Im Dialogfeld “Benutzeroptionen” können Sie die Optionen für den PMML-Export ändern. Klicken Sie im Hauptmenü auf Folgendes:

Werkzeuge > Optionen > Benutzeroptionen

und wählen Sie die Registerkarte “PMML”.

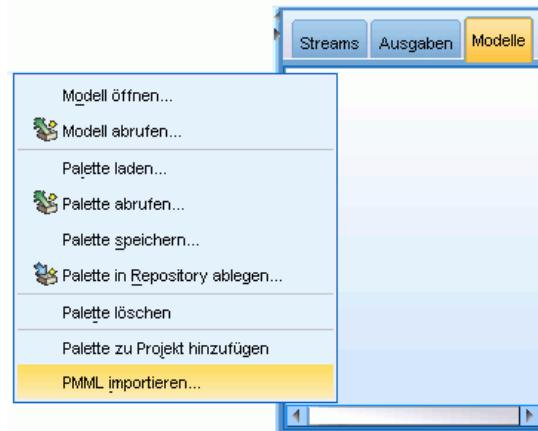
Für weitere Informationen siehe Thema Festlegen von PMML-Exportoptionen in Kapitel 12 auf S. 248.

So importieren Sie ein als PMML gespeichertes Modell:

Modelle, die aus SPSS Modeler oder einer anderen Anwendung als PMML exportiert wurden, können in die Modellpalette importiert werden. Für weitere Informationen siehe Thema [Modelltypen, die PMML unterstützen auf S. 223](#).

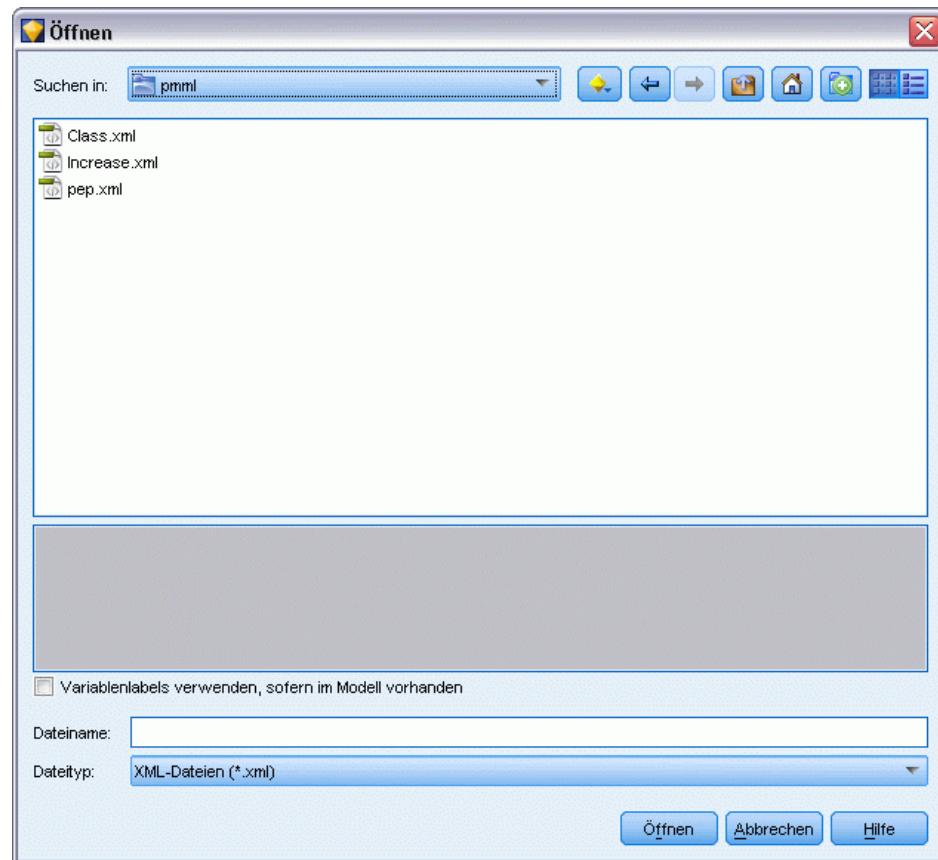
- ▶ Klicken Sie in der Modellpalette mit der rechten Maustaste auf die Palette und wählen Sie aus dem Menü die Option PMML importieren.

Abbildung 10-2
Importieren von Modellen im PMML-Format



- ▶ Wählen Sie die zu importierende Datei aus und geben Sie nach Bedarf Optionen für Variablenlabels an.
- ▶ Klicken Sie auf Öffnen.

Abbildung 10-3
Auswählen der XML-Datei für ein mit PMML gespeichertes Modell



Verwenden Sie Variablenlabels, falls im Modell vorhanden. Im PMML-Code können sowohl die Variablennamen als auch die Variablenlabels (beispielsweise “Referrer ID” für *RefID*) für Variablen im Datenlexikon gefunden. Wählen Sie diese Option aus, um Variablenlabels zu verwenden, wenn diese im ursprünglich exportierten PMML-Code vorhanden sind.

Wenn Sie die Optionen zur Variablenbeschriftung ausgewählt haben, im PMML-Code jedoch keine Variablenlabels vorhanden sind, werden die Variablennamen wie üblich verwendet.

Modelltypen, die PMML unterstützen

PMMML-Export

SPSS Modeler Modelle. Die folgenden in IBM® SPSS® Modeler erstellten Modelle können als PMML 4.0 exportiert werden:

- C&R-Baum
- QUEST
- CHAID
- Lineare Regression
- Netzwerk
- C5.0
- Logistische Regression
- Genlin
- SVM
- Bayes-Netz
- A Priori
- Carma
- K-Means
- Kohonen
- Two Step
- KNN
- Statistics Modell

Das folgende in SPSS Modeler erstellte Modell kann als PMML 3.2 exportiert werden:

- Entscheidungsliste

Datenbankeigene Modelle. Bei Modellen, die mithilfe von datenbankeigenen Algorithmen generiert wurden, steht der PMML-Export nur bei IBM InfoSphere Warehouse-Modellen zur Verfügung. Modelle, die mithilfe von Analysis Services von Microsoft oder mit Oracle Data Miner erstellt wurden, können nicht exportiert werden. Beachten Sie außerdem, dass als PMML exportierte IBM-Modelle danach nicht wieder in SPSS Modeler importiert werden können. Für weitere Informationen siehe Thema Übersicht über die Datenbank-Modellierung in Kapitel 2 in *IBM SPSS Modeler 15 In-Database Mining-Handbuch*.

PMML-Import

SPSS Modeler kann PMML-Modelle importieren und sceren, die von aktuellen Versionen aller IBM® SPSS® Statistics-Produkte erstellt wurden, darunter Modelle, die aus SPSS Modeler exportiert wurden, sowie Modell- bzw. Transformations-PMML, die von SPSS Statistics 17.0 oder höher generiert wurde. Dies gilt also im Grunde für jegliche PMML, die die Scoring-Engine sceren kann – mit folgenden Ausnahmen:

- A Priori-, CARMA-, Anomaly Detection-(Anomalieerkennungs-)Modelle und Sequenzmodelle können nicht importiert werden.
- PMML-Modelle können nach dem Import in SPSS Modeler nicht durchsucht werden, obwohl sie für das Scoring verwendet werden können. (Dies gilt auch für Modelle, die ursprünglich aus SPSS Modeler exportiert wurden. Um diese Einschränkung zu vermeiden, sollten Sie das betreffende Modell als generierte Modelldatei (*.gm) und nicht als PMML exportieren.)
- Als PMML exportierte IBM InfoSphere Warehouse-Modelle können nicht importiert werden.
- Eine eingeschränkte Validierung findet beim Import statt, aber die vollständige Validierung erfolgt beim Versuch, das Modell zu sceren. Daher kann der Import erfolgreich durchgeführt werden, das Scoring aber fehlschlagen oder falsche Ergebnisse erzeugen.

Projekte und Berichte

Einführung in Projekte

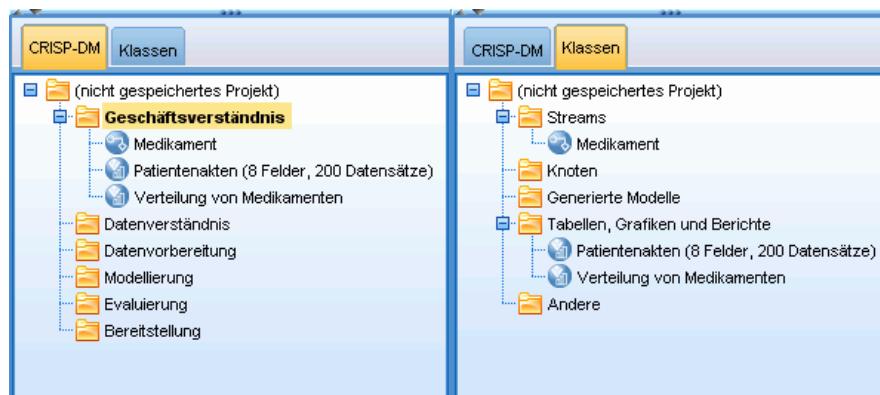
Ein **Projekt** ist eine Gruppe von Dateien, die in Bezug zu einer Data Mining-Aufgabe stehen. Zu Projekten gehören Daten-Streams, Diagramme, generierte Modelle, Berichte und alles andere, was Sie in IBM® SPSS® Modeler erstellt haben. Auf den ersten Blick kann der Eindruck entstehen, SPSS Modeler-Projekte seien einfach eine Methode zur Organisation von Ausgaben, doch sie haben noch viele weitere Funktionen. Mit Projekten haben Sie folgende Möglichkeiten:

- Erstellen von Anmerkungen für jedes Objekt in der Projektdatei.
- Verwenden der CRISP-DM-Methodologie als Orientierung beim Data Mining. Projekte können außerdem ein CRISP-DM-Hilfesystem enthalten, das Einzelheiten und reale Beispiele für Data Mining mit CRISP-DM bietet.
- Hinzufügen von nicht aus SPSS Modeler stammenden Objekten zum Projekt, beispielsweise einer PowerPoint-Präsentation zur Vorstellung Ihrer Data Mining-Ziele oder White Papers für die Algorithmen, die Sie verwenden möchten.
- Erstellen von umfassenden und gleichzeitig einfachen Aktualisierungsberichten auf der Grundlage Ihrer Anmerkungen. Diese Berichte können in HTML generiert werden, um eine einfache Veröffentlichung im Intranet Ihrer Organisation zu ermöglichen.

Hinweis: Wenn der Projektbereich nicht im SPSS Modeler-Fenster sichtbar ist, klicken Sie auf die Option Projekt im Menü “Ansicht”.

Es gibt zwei Möglichkeiten zur Anzeige der zu einem Projekt hinzugefügten Objekte: Die **Klassenansicht** und die **CRISP-DM-Ansicht**. Alles, was Sie zu einem Projekt hinzufügen, wird zu beiden Ansichten hinzugefügt. Sie können zwischen den Ansichten umschalten, um die Organisation zu erstellen, die am besten geeignet ist.

Abbildung 11-1
CRISP-DM-Ansicht und Klassenansicht einer Projektdatei



CRISP-DM-Ansicht

Durch Unterstützung von CRISP-DM (Cross-Industry Standard Process for Data Mining) bieten IBM® SPSS® Modeler-Projekte eine in der Branche bewährte, nicht eigentumsrechtlich geschützte Methode zur Organisation der einzelnen Bruchstücke Ihrer Data Mining-Arbeit. CRISP-DM verwendet sechs Phasen zur Beschreibung des Prozesses vom Start (Zusammenstellen der Geschäftsanforderungen) bis zum Ende (Bereitstellen der Ergebnisse). Auch wenn einige Phasen bei SPSS Modeler normalerweise keine Arbeit beinhalten, umfasst der Projektbereich alle sechs Phasen, sodass Ihnen ein Zentrum zum Speichern und Nachverfolgen aller Materialien zur Verfügung steht, die mit dem Projekt zu tun haben. Zum Geschäftsverständnis gehören üblicherweise das Sammeln von Anforderungen sowie Besprechungen mit Kollegen zur Bestimmung der Ziele und weniger die Arbeit mit den Daten in SPSS Modeler. Mit dem Projektbereich können Sie Ihre Aufzeichnungen aus solchen Besprechungen im Ordner *Geschäftsverständnis* speichern, um sie später nachlesen oder in Berichte aufnehmen zu können.

Abbildung 11-2
CRISP-DM, Ansicht



Die CRISP-DM-Ansicht im Projektbereich ist mit einem eigenen Hilfesystem ausgestattet, das Sie durch den Data Mining-Zyklus führt. In SPSS Modeler ist der Zugriff auf diese Hilfe durch Klicken auf CRISP-DM-Hilfe im Hilfemenü möglich.

Hinweis: Wenn der Projektbereich nicht im Fenster sichtbar ist, klicken Sie auf die Option Projekt im Menü „Ansicht“.

Einrichten der Standard-Projektphase

Die zu einem Projekt hinzugefügten Objekte werden zu einer Standardphase von CRISP-DM hinzugefügt. Daher müssen Sie die Objekte manuell entsprechend der Data Mining-Phase organisieren, in der Sie sie verwendet haben. Es ist sinnvoll, den Standardordner auf die Phase einzustellen, in der Sie gerade arbeiten.

So legen Sie fest, welche Phase als Standard verwendet wird:

- ▶ Klicken Sie in der CRISP-DM-Ansicht mit der rechten Maustaste auf den Ordner für die Phase, die als Standard festgelegt werden soll.
- ▶ Klicken Sie im Menü auf Als Standard festlegen.

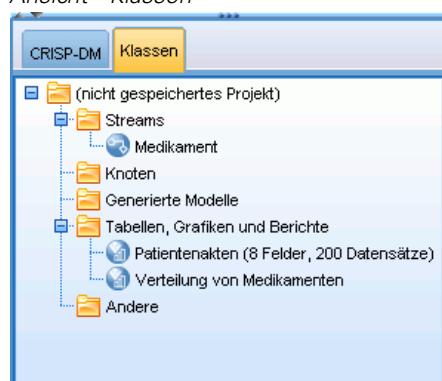
Der Standardordner wird fett gedruckt angezeigt.

Klassenansicht

Mit der Ansicht "Klassen" im Projektbereich können Sie Ihre Arbeit in IBM® SPSS® Modeler nach den erstellten Objekttypen in Kategorien organisieren. Die gespeicherten Objekte können zu einer der folgenden Kategorien hinzugefügt werden:

- Streams
- Knoten
- Models
- Tabellen, Diagramme, Berichte
- Andere (Nicht-SPSS Modeler-Dateien, wie beispielsweise Präsentationen oder White Papers, die für Ihre Data Mining-Arbeit relevant sind)

Abbildung 11-3
Ansicht "Klassen"



Wenn Objekte zur Klassenansicht hinzugefügt werden, werden sie auch zum Standardphasenordner in der CRISP-DM-Ansicht hinzugefügt.

Hinweis: Wenn der Projektbereich nicht im Fenster sichtbar ist, klicken Sie auf die Option Projekt im Menü "Ansicht".

Erstellen von Projekten

Bei einem Projekt handelt es sich im Wesentlichen um eine Datei mit Verweisen auf alle Dateien, die Sie dem Projekt zuordnen. Projektelemente werden also sowohl einzeln als auch als Verweis in der Projektdatei (.cpj) gespeichert. Aufgrund dieser referentiellen Struktur sollten Sie auf folgende Punkte achten:

- Projektelemente müssen zunächst einzeln gespeichert werden, bevor sie zu einem Projekt hinzugefügt werden. Wenn ein Element nicht gespeichert ist, werden Sie aufgefordert, es zu speichern, bevor Sie es zum aktuellen Projekt hinzufügen.

- Einzeln aktualisierte Objekte, beispielsweise Streams, werden auch in der Projektdatei aktualisiert.
- Beim manuellen Verschieben oder Löschen von Objekten (beispielsweise Streams, Knoten und Ausgabeobjekte) aus dem Dateisystem werden die Verknüpfungen in der Projektdatei ungültig.

Erstellen eines neuen Projekts

Neue Projekte können problemlos im IBM® SPSS® Modeler-Fenster erstellt werden. Sie können entweder mit dem Erstellen eines Projekts beginnen, wenn noch keines geöffnet ist, oder Sie können ein bestehendes Objekt schließen und von vorne anfangen.

- ▶ Klicken Sie im Hauptmenü auf Folgendes:
Datei > Projekt > Neues Projekt ...

Hinzufügen zu einem Projekt

Nach dem Erstellen oder Öffnen von Projekten stehen Ihnen mehrere Methoden zur Verfügung, um Objekte, wie beispielsweise Daten-Streams, Knoten und Berichte, hinzuzufügen.

Hinzufügen von Objekten aus den Managern

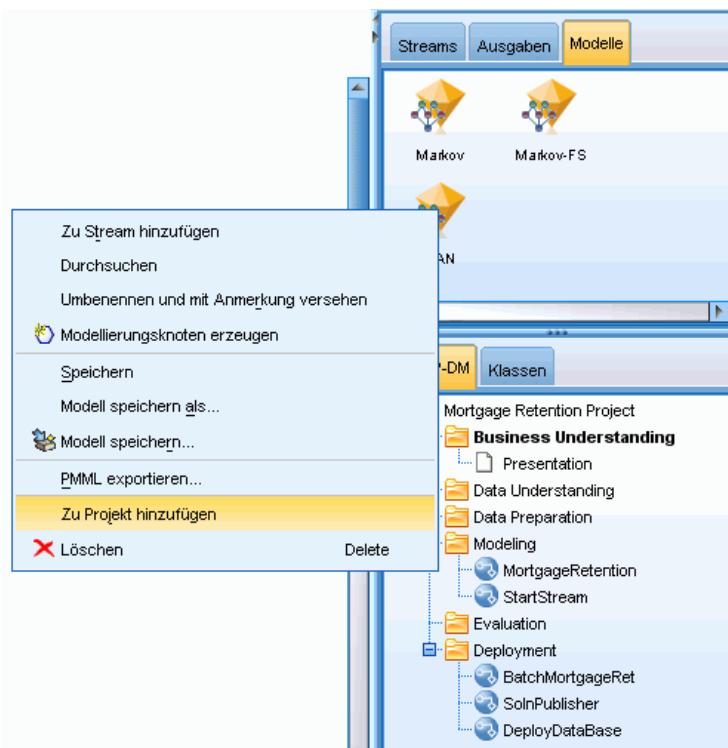
Mithilfe der Manager in der rechten oberen Ecke des IBM® SPSS® Modeler-Fensters können Sie Streams oder Ausgaben hinzufügen.

- ▶ Wählen Sie ein Objekt, beispielsweise eine Tabelle oder einen Stream aus einer der Manager-Registerkarten aus.
- ▶ Klicken Sie mit der rechten Maustaste und klicken Sie auf die Option Zu Projekt hinzufügen.

Wenn das Objekt zuvor gespeichert wurde, wird es automatisch zum entsprechenden Objektordner (in der Klassenansicht) oder zum Ordner der Standardphase (in der CRISP-DM-Ansicht) hinzugefügt.
- ▶ Alternativ können Sie durch Ziehen und Ablegen Objekte aus den Managern in den Projektbereich verschieben.

Hinweis: Sie werden möglicherweise dazu aufgefordert, das Objekt zuerst zu speichern. Beim Speichern muss Datei zu Projekt hinzufügen im Dialogfeld "Speichern" ausgewählt sein. Dadurch wird das Objekt nach dem Speichern automatisch zum Projekt hinzugefügt.

Abbildung 11-4
Hinzufügen von Elementen zu einem Projekt



Hinzufügen von Knoten aus dem Zeichenbereich

Mit dem Dialogfeld “Speichern” können Sie einzelne Knoten aus dem Stream-Zeichenbereich hinzufügen.

- ▶ Wählen Sie einen Knoten im Zeichenbereich aus.
- ▶ Klicken Sie mit der rechten Maustaste und klicken Sie auf die Option **Knoten speichern**. Klicken Sie alternativ im Hauptmenü auf Folgendes:
Bearbeiten > Knoten > Speicherknoten ...
- ▶ Wählen Sie im Dialogfeld “Speichern” die Option **Datei zum Projekt hinzufügen**.
- ▶ Erstellen Sie einen Namen für den Knoten und klicken Sie auf **Speichern**.

Dadurch wird die Datei gespeichert und zum Projekt hinzugefügt. Knoten werden in der Klassenansicht zum Ordner **Knoten** und in der CRISP-DM-Ansicht zum Ordner der Standardphase hinzugefügt.

Hinzufügen externer Dateien

Sie können eine große Vielzahl von Nicht-SPSS Modeler-Objekten zu einem Projekt hinzufügen. Dies ist sinnvoll, wenn der gesamte Data-Mining-Vorgang innerhalb von SPSS Modeler verwaltet werden soll. Sie können beispielsweise Verknüpfungen zu Daten, Notizen, Präsentationen und

Grafiken in einem Objekt speichern. In der CRISP-DM-Ansicht können externe Dateien zum Ordner Ihrer Wahl hinzugefügt werden. In der Klassenansicht können externe Dateien nur im Ordner *Andere* gespeichert werden.

So fügen Sie externe Dateien zu einem Projekt hinzu:

- ▶ Ziehen Sie die Dateien vom Desktop auf das Projekt.
oder
- ▶ Klicken Sie mit der rechten Maustaste auf den Zielordner in der CRISP-DM- oder Klassen-Ansicht.
- ▶ Klicken Sie im Menü auf die Option Zu Ordner hinzufügen.
- ▶ Wählen Sie eine Datei im Dialogfeld aus und klicken Sie auf Öffnen.

Dadurch wird ein Verweis zum ausgewählten Objekt innerhalb der SPSS Modeler-Projekte hergestellt.

Übertragen von Projekten an das IBM SPSS Collaboration and Deployment Services Repository

Sie können ein gesamtes Projekt, einschließlich aller Komponentendateien, in einem Schritt an das IBM® SPSS® Collaboration and Deployment Services Repository übertragen. Alle Objekte, die sich bereits am Zielspeicherort befinden, werden nicht verschoben. Diese Funktion kann auch in umgekehrter Richtung angewendet werden: Sie können gesamte Projekte vom IBM SPSS Collaboration and Deployment Services Repository in Ihr lokales Dateisystem übertragen.

Hinweis: Für den Zugriff auf ein IBM® SPSS® Collaboration and Deployment Services-Repository ist eine separate Lizenz erforderlich. Weitere Informationen finden Sie im Dokument <http://www.ibm.com/software/analytics/spss/products/deployment/cds/>

Übertragen von Projekten

Vergewissern Sie sich, dass das zu übertragende Projekt im Projektbereich geöffnet ist.

So übertragen Sie ein Projekt:

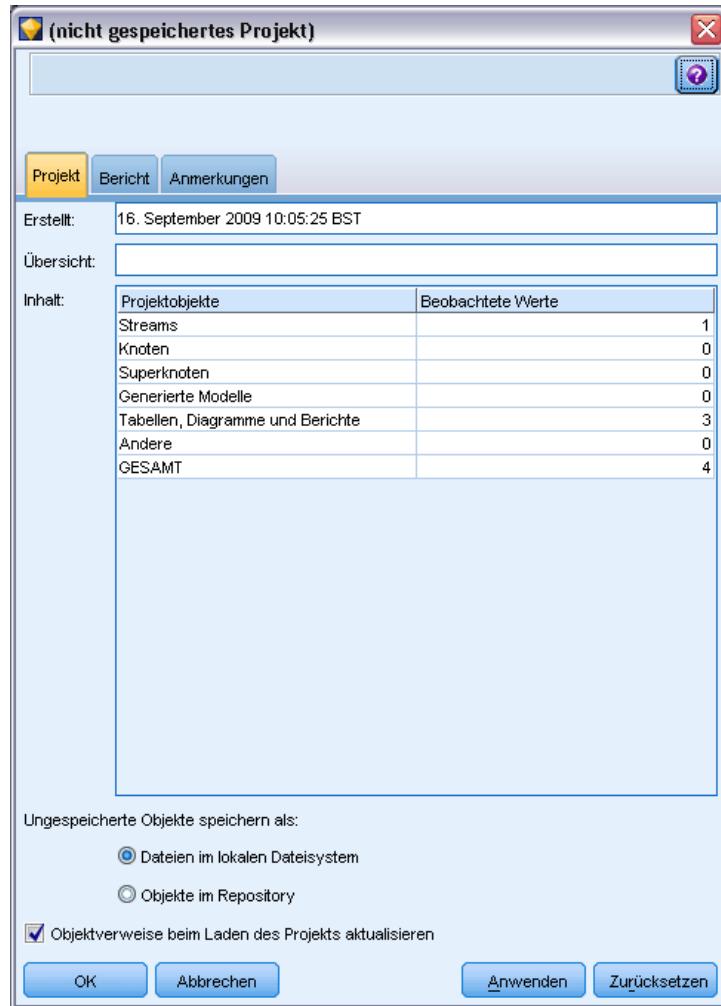
- ▶ Klicken Sie mit der rechten Maustaste auf den Projektordner und klicken Sie auf die Option Projekt übertragen.
- ▶ Wenn Sie dazu aufgefordert werden, melden Sie sich bei IBM SPSS Collaboration and Deployment Services Repository an.
- ▶ Geben Sie den neuen Speicherort für das Projekt an und klicken Sie auf OK.

Festlegen der Projekteigenschaften

Sie können den Inhalt und die Dokumentation eines Projekts mithilfe des Dialogfelds “Projekteigenschaften” anpassen. So können Sie auf die Projekteigenschaften zugreifen:

- ▶ Klicken Sie mit der rechten Maustaste auf ein Objekt oder einen Ordner im Projektbereich und klicken Sie auf die Option Projekteigenschaften.
- ▶ Klicken Sie auf die Registerkarte Projekt, um grundlegende Projektinformationen anzugeben.

Abbildung 11-5
Festlegen der Projekteigenschaften



Erstellt. Zeigt das Erstellungsdatum des Projekts an (nicht bearbeitbar).

Zusammenfassung. Sie können eine Übersicht für Ihr Data Mining-Projekt eingeben, die im Projektbericht angezeigt werden soll.

Inhalt. Listet Typ und Anzahl der von der Projektdatei referenzierten Komponenten auf (nicht bearbeitbar).

Ungespeicherte Objekte speichern als. Gibt an, ob ungespeicherte Objekte im lokalen Dateisystem oder im Repository gespeichert werden sollen. [Für weitere Informationen siehe Thema Informationen zu IBM SPSS Collaboration and Deployment Services Repository in Kapitel 9 auf S. 180.](#)

Objektverweise beim Laden des Projekts aktualisieren. Wählen Sie diese Option, um die Verweise des Projekts auf seine Komponenten zu aktualisieren. *Hinweis:* Die zu einem Projekt hinzugefügten Dateien werden nicht in der Projektdatei selbst gespeichert. Stattdessen wird ein Verweis auf die Dateien im Projekt gespeichert. Das bedeutet, dass durch Verschieben oder Löschen einer Datei das betreffende Objekt aus dem Projekt entfernt wird.

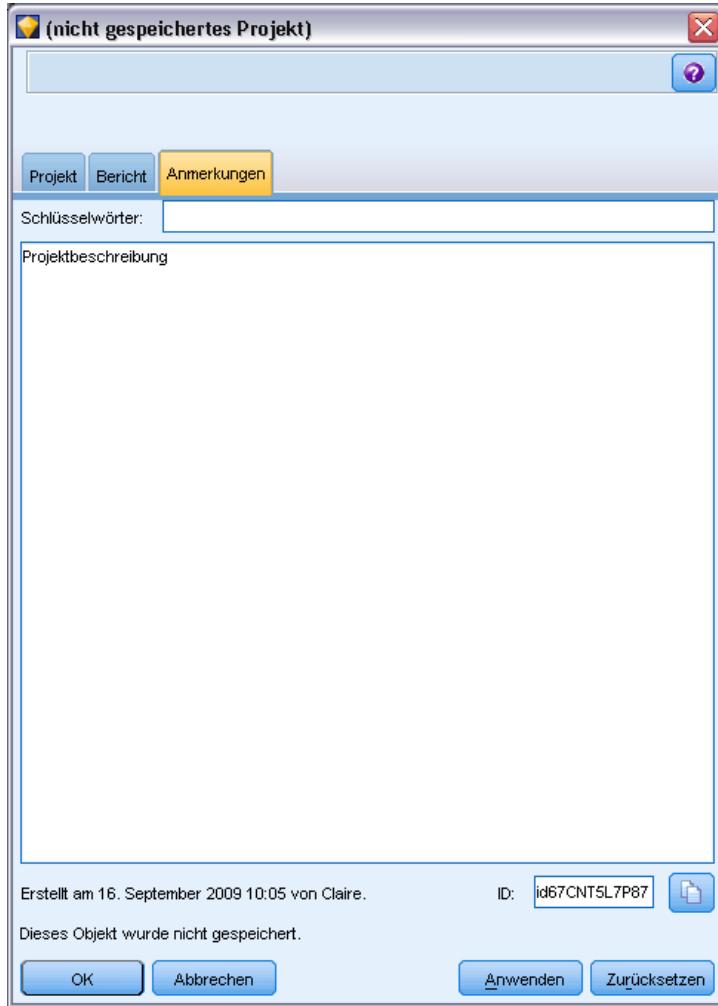
Erstellen von Anmerkungen für ein Projekt

Der Projektbereich bietet eine Reihe von Möglichkeiten, um Anmerkungen für Ihre Data Mining-Arbeit zu erstellen. Anmerkungen auf der Projektebene werden häufig verwendet, um weiter reichende Ziele und Entscheidungen aufzuzeichnen, während die Anmerkungen auf der Ordner- oder Knotenebene zusätzliche Details bieten. Die Registerkarte “Anmerkungen” bietet genügend Raum zum Dokumentieren von Informationen für die Projektebene, wie beispielsweise den Ausschluss von Datensätzen mit nicht wiederherstellbaren fehlenden Daten oder vielversprechenden Hypothesen, die während der Datenuntersuchung aufgestellt wurden.

So erstellen Sie Anmerkungen zu einem Projekt:

- ▶ Wählen Sie den Projektordner entweder in der CRISP-DM- oder in der Klassenansicht aus.
- ▶ Klicken Sie mit der rechten Maustaste auf den Ordner und klicken Sie auf die Option Projekteigenschaften.
- ▶ Klicken Sie auf die Registerkarte Anmerkungen.

Abbildung 11-6
Registerkarte "Anmerkungen" im Dialogfeld "Projekteigenschaften"



- Geben Sie Schlüsselwörter und Text zur Beschreibung des Projekts ein.

Ordnereigenschaften und Anmerkungen

Einzelne Projektordner (in der CRISP-DM-Ansicht und der Klassenansicht) können mit Anmerkungen versehen werden. In der CRISP-DM-Ansicht kann dies eine sehr effektive Methode sein, um die Ziele Ihrer Organisation für die einzelnen Phasen des Data Mining zu dokumentieren. Beispielsweise können Sie mit dem Anmerkungstool im Ordner *Geschäftsverständnis* Hinweise aufnehmen wie "Das Geschäftsziel für diese Studie besteht darin, die Abwanderung hochwertiger Kunden zu reduzieren". Dieser Text könnte dann durch Auswahl der Option In Bericht einschließen automatisch in den Projektbericht aufgenommen werden.

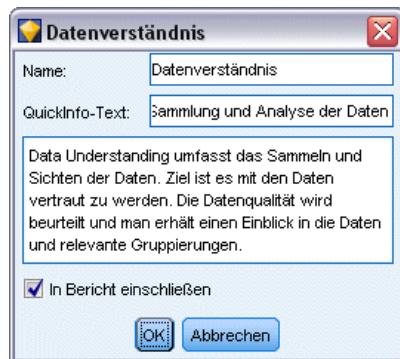
So erstellen Sie Anmerkungen zu einem Ordner:

- Wählen Sie einen Ordner im Projektbereich aus.

- Klicken Sie mit der rechten Maustaste auf den Ordner und klicken Sie auf die Option **Ordnereigenschaften**.

In der CRISP-DM-Ansicht bestehen die Anmerkungen für die Ordner aus einer Übersicht über den Zweck der einzelnen Phasen sowie in Hinweisen zur Durchführung der relevanten Data Mining-Aufgaben. Sie können diese Anmerkungen nach Belieben entfernen oder bearbeiten.

Abbildung 11-7
Projektordner mit CRISP-DM-Anmerkung



Name. In diesem Bereich wird der Name des ausgewählten Felds angezeigt.

QuickInfo-Text. Dient zum Erstellen benutzerdefinierter QuickInfos, die angezeigt werden, wenn der Mauszeiger über einen Projektordner geführt wird. Dies ist vor allem in der CRISP-DM-Ansicht sinnvoll, beispielsweise um einen schnellen Überblick über die Ziele der einzelnen Phasen zu bieten oder den Status einer Phase (z. B. "In Arbeit" oder "Abgeschlossen") anzugeben.

Anmerkungsfeld. Dieses Feld ist für längere Anmerkungen gedacht, die im Projektbericht zusammengestellt werden können. Die CRISP-DM-Ansicht enthält eine Beschreibung der einzelnen Data Mining-Phasen in der Anmerkung, Sie können diese jedoch für Ihr Projekt nach Belieben anpassen.

In Bericht einschließen. Um die Anmerkung in die Berichte einzuschließen, wählen Sie **In Bericht einschließen** aus.

Objekteigenschaften

Sie können Objekteigenschaften anzeigen und auswählen, ob einzelne Objekte in den Projektbericht aufgenommen werden sollen. So können Sie auf die Objekteigenschaften zugreifen:

- Klicken Sie mit der rechten Maustaste auf ein Objekt im Projektbereich.
- Klicken Sie im Menü auf die Option **Objekteigenschaften**.

Abbildung 11-8
Dialogfeld "Objekteigenschaften"



Name. In diesem Bereich wird der Name des gespeicherten Objekts aufgeführt.

Pfad. In diesem Bereich wird der Speicherort des gespeicherten Objekts aufgeführt.

In Bericht einschließen. Wählen Sie diese Option aus, um die Objektdetails in einem generierten Bericht einzuschließen.

Schließen von Projekten

Wenn Sie IBM® SPSS® Modeler beenden oder ein neues Projekt öffnen, wird die vorhandene Projektdatei (.cpj) geschlossen.

Einige Dateien, die mit dem Projekt assoziiert sind (z. B. Streams, Knoten oder Grafiken), sind möglicherweise noch geöffnet. Wenn diese Dateien geöffnet bleiben sollen, antworten Sie mit **Nein** auf die Meldung ... Diese Dateien speichern und schließen?

Wenn Sie zugehörige Dateien nach dem Schließen eines Projekts bearbeiten und speichern, werden diese aktualisierten Versionen beim nächsten Öffnen des Projekts in das Projekt aufgenommen. Um dies zu verhindern, müssen Sie die betreffende Datei aus dem Projekt entfernen oder unter einem anderen Dateinamen speichern.

Generieren von Berichten

Eine der nützlichsten Funktionen von Projekten ist die Möglichkeit, Berichte auf der Grundlage der Projektelemente und der Anmerkungen zu erstellen. Dies ist eine äußerst wichtige Komponente des effektiven Data Mining, wie in der gesamten CRISP-DM-Methodologie. Sie können einen Bericht direkt in einem von mehreren Dateitypen erstellen oder in einem Ausgabefenster auf dem Bildschirm, um ihn sofort betrachten zu können. Von dort können Sie den Bericht drucken, speichern oder in einem Webbrowser anzeigen. Sie können gespeicherte Berichte an andere Personen in Ihrer Organisation verteilen.

Berichte werden häufig mehrmals während des Data Mining-Prozesses aus den Berichtdateien generiert, um sie an die am Projekt beteiligten Personen weiterzugeben. Der Bericht entnimmt Informationen zu den referenzierten Objekten aus der Projektdatei sowie aus den erstellten Anmerkungen (sofern vorhanden). Sie können die Berichte entweder auf der Grundlage der Klassenansicht oder der CRISP-DM-Ansicht erstellen.

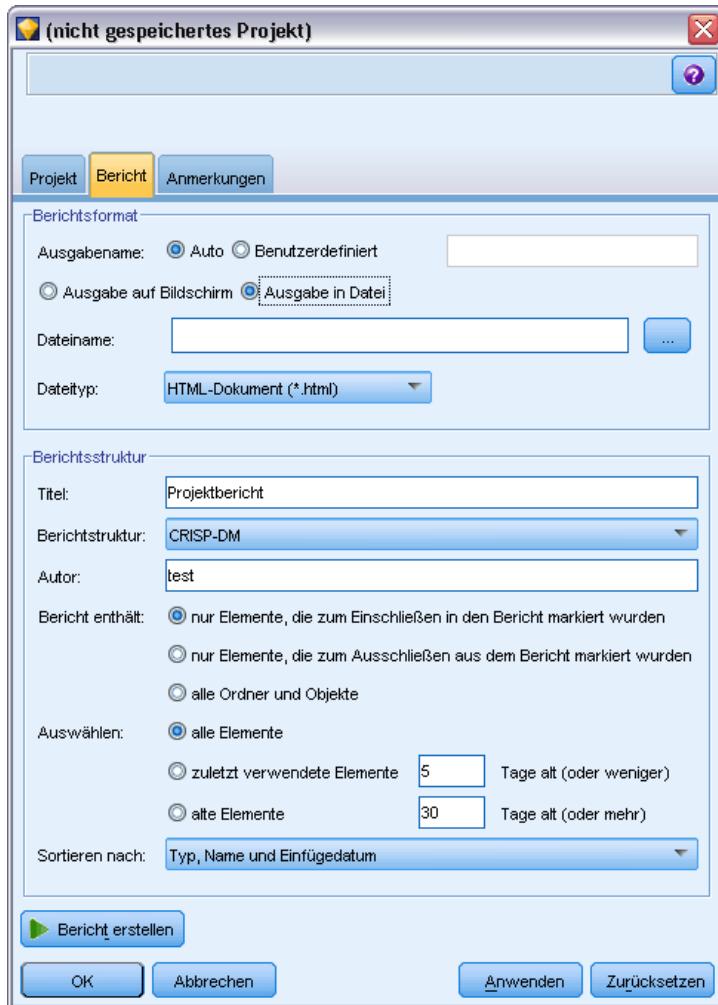
Abbildung 11-9
Fenster für generierte Berichte



So generieren Sie einen Bericht:

- ▶ Wählen Sie den Projektordner entweder in der CRISP-DM- oder in der Klassenansicht aus.
- ▶ Klicken Sie mit der rechten Maustaste auf den Ordner und klicken Sie auf die Option Projektbericht.
- ▶ Geben Sie die Berichtsoptionen an und klicken Sie auf Bericht erstellen.

Abbildung 11-10
Auswählen der Optionen für einen Bericht



Die Optionen im Berichtdialogfeld bieten mehrere Möglichkeiten zur Generierung des erforderlichen Berichtstyps:

Ausgabename. Geben Sie den Namen des Ausgabefensters an, wenn die Ausgabe des Berichts an den Bildschirm gesendet werden soll. Sie können einen benutzerdefinierten Namen angeben oder IBM® SPSS® Modeler das Fenster automatisch benennen lassen.

Ausgabe auf Bildschirm. Wählen Sie diese Option, um den Bericht zu generieren und in einem Ausgabefenster anzuzeigen. Beachten Sie, dass Sie die Möglichkeit haben, den Bericht aus dem Ausgabefenster in verschiedene Dateitypen zu exportieren.

Ausgabe in Datei. Wählen Sie diese Option, um den Bericht zu erstellen und als Datei mit dem in der Liste "Dateityp" angegebenen Typ zu speichern.

Dateiname. Geben Sie einen Dateinamen für den generierten Bericht an. Die Dateien werden standardmäßig im \bin-Verzeichnis von SPSS Modeler gespeichert. Mit der Auslassungsschaltfläche (...) können Sie einen anderen Speicherort angeben.

Dateityp. Die folgenden Dateitypen stehen zur Auswahl:

- **HTML-Dokument.** Der Bericht wird als einzelne HTML-Datei gespeichert. Wenn Ihr Bericht Diagramme enthält, werden diese als PNG-Dateien gespeichert und von der HTML-Datei referenziert. Wenn Sie Ihren Bericht im Internet veröffentlichen, müssen Sie die HTML-Datei und alle von ihr referenzierten Bilder hochladen.
- **Textdokument.** Der Bericht wird als einzelne Textdatei gespeichert. Wenn der Bericht Diagramme enthält werden nur die Referenzen auf Dateinamen und Pfad in den Bericht aufgenommen.
- **Microsoft Word-Dokument.** Der Bericht wird als Einzeldokument gespeichert. Alle Diagramme sind direkt in das Dokument eingebettet.
- **Microsoft Excel-Dokument.** Der Bericht wird als einzelnes Arbeitsblatt gespeichert. Alle Diagramme sind direkt in das Arbeitsblatt eingebettet.
- **Microsoft PowerPoint-Dokument.** Jede Phase wird auf einer Folie angezeigt. Alle Diagramme sind direkt in die PowerPoint-Folien eingebettet.
- **Ausgabeobjekt.** Wenn diese Datei (.cou) in SPSS Modeler geöffnet wird, stimmt sie mit der oben genannten Option Ausgabe auf Bildschirm in der Gruppe Berichtformat überein.

Hinweis: Um eine Microsoft Office-Datei zu exportieren, muss die entsprechende Anwendung installiert sein.

Titel. Geben Sie einen Titel für den Bericht an.

Berichtstruktur. Wählen Sie entweder CRISP-DM oder Klassen. Die CRISP-DM-Ansicht bietet einen Statusbericht mit überblicksartigen Synapsen sowie Einzelheiten zu den einzelnen Data Mining-Phasen. Bei der Ansicht "Klassen" handelt es sich um eine objektbasierte Ansicht, die besser für die interne Verfolgung von Daten und Streams geeignet ist.

Autor. Der Standardbenutzername wird angezeigt. Sie können ihn jedoch ändern.

Bericht enthält. Wählen Sie eine Methode für die Aufnahme von Objekten in den Bericht aus. Wählen Sie alle Ordner und Objekte, um alle zur Projektdatei hinzugefügten Elemente einzuschließen. Sie können auch nur die Elemente einschließen lassen, bei denen in den Objekteigenschaften In Bericht einschließen ausgewählt ist. Alternativ können Sie, um ungemeldete Elemente zu überprüfen, festlegen, dass nur Elemente aufgenommen werden sollen, die für den Ausschluss markiert sind (bei denen In Bericht einschließen deaktiviert ist).

Auswählen. Mit dieser Option können Sie Projektaktualisierungen bereitstellen, indem Sie nur Elemente der Kategorie zuletzt verwendete Elemente in den Bericht aufnehmen. Alternativ können Sie ältere und vielleicht ungelöste Probleme nachverfolgen, indem Sie die Parameter für alte Elemente festlegen. Wählen Sie alle Elemente aus, um die Zeit als Berichtsparameter zu ignorieren.

Sortieren nach. Sie können eine Kombination der folgenden Objektmerkmale auswählen, um sie innerhalb eines Ordners zu sortieren:

- **Typ.** Objekte werden nach Typ gruppiert.
- **Name.** Objekte werden alphabetisch angeordnet.
- **Einfügedatum.** Objekte werden nach dem Datum ihrer Hinzufügung zum Projekt sortiert.

Speichern und Exportieren von generierten Berichten

Ein auf dem Bildschirm generierter Bericht wird in einem neuen Ausgabefenster angezeigt. Alle im Bericht enthaltenen Diagramme werden als Inline-Bilder angezeigt.

Im Bericht wird die Gesamtzahl der Knoten in den einzelnen Streams aufgeführt. Die Zahlen werden unter folgenden Überschriften angegeben, wobei IBM® SPSS® Modeler-Terminologie, nicht CRISP-DM-Terminologie, verwendet wird:

- **Datenleser.** Quellenknoten. Für weitere Informationen siehe Thema Übersicht in Kapitel 2 in *IBM SPSS Modeler 15 Quellen-, Prozess- und Ausgabeknoten*.
- **Datenschreiber.** Exportknoten. Für weitere Informationen siehe Thema Überblick über Exportknoten in Kapitel 7 in *IBM SPSS Modeler 15 Quellen-, Prozess- und Ausgabeknoten*.
- **Modellersteller.** Erstellungs- bzw. Modellierungsknoten. Für weitere Informationen siehe Thema Überblick über Modellierungsknoten in Kapitel 3 in *IBM SPSS Modeler 15 Modellierungsknoten*.
- **Modellanwender.** Generierte Modelle, auch als “Nuggets” bezeichnet. Für weitere Informationen siehe Thema Modell-Nuggets in Kapitel 3 in *IBM SPSS Modeler 15 Modellierungsknoten*.
- **Ausgabeersteller.** Diagramm- oder Ausgabeknoten. Für weitere Informationen siehe Thema Häufig verwendete Funktionen von Diagrammknoten in Kapitel 5 in *IBM SPSS Modeler 15 Quellen-, Prozess- und Ausgabeknoten*. Für weitere Informationen siehe Thema Überblick über Ausgabeknoten in Kapitel 6 in *IBM SPSS Modeler 15 Quellen-, Prozess- und Ausgabeknoten*.
- **Sonstige.** Alle anderen Knoten, die zum Projekt gehören. Dazu gehören beispielsweise die Knoten, die auf der Registerkarte “Feldoperationen” oder “Datensatzoperationen” auf der Knotenpalette verfügbar sind.

So speichern Sie einen Bericht:

- ▶ Klicken Sie im Menü “Datei” auf die Option Speichern.
- ▶ Geben Sie einen Dateinamen an.

Der Bericht wird als Ausgabeobjekt gespeichert.

So exportieren Sie einen Bericht:

- ▶ Klicken Sie im Menü “Datei” auf die Option Exportieren und wählen Sie den zu exportierenden Dateityp aus.
- ▶ Geben Sie einen Dateinamen an.

Der Bericht wird im gewählten Format gespeichert.

Ein Export ist in folgende Dateitypen möglich:

- HTML
- Text
- Microsoft Word

- Microsoft Excel
- Microsoft PowerPoint

Hinweis: Um eine Microsoft Office-Datei zu exportieren, muss die entsprechende Anwendung installiert sein.

Mit den Schaltflächen oben im Fenster haben Sie folgende Möglichkeiten:

- Bericht drucken.
- Bericht in einem externen Webbrowser als HTML anzeigen.

Abbildung 11-11

In einem Webbrowser angezeigter Bericht

The screenshot shows a web browser window with a report. At the top, the title 'Business Understanding' is displayed in a large, bold, dark font. Below the title, there is a short text block: 'This initial phase focuses on understanding the project objectives and requirements from a business perspective.' Underneath this text, there is a section titled 'Objektliste' (Object List) in bold. A table follows, showing one item: 'Presentation' with the file path 'C:/SPSS Screenshooting/Modeler_Shots_Kit/Data_Files/Mortgage Retention/Presentation.ppt'. The table has columns for 'Name', 'Dateiname', and 'Hinzugefügt'. To the right of the table, there is a vertical scroll bar. Below the object list, there are three more sections: 'Presentation', 'Data Understanding', and 'Data Preparation', each with a brief description below it.

Name	Dateiname	Hinzugefügt
Presentation	C:/SPSS Screenshooting/Modeler_Shots_Kit/Data_Files/Mortgage Retention/Presentation.ppt	10.08.2004 18:31:32

Presentation

Data Understanding

The data understanding involves collecting data in order to become familiar with the data, identify data quality problems and identify initial insights or interesting subsets.

Data Preparation

This phase covers all activities involved in constructing the final datasets to be used in the modeling phase.

Anpassen von IBM SPSS Modeler

Anpassen von IBM SPSS Modeler-Optionen

Es gibt eine Reihe von Operationen, mit denen Sie IBM® SPSS® Modeler an Ihre Bedürfnisse anpassen können. Diese Anpassung besteht hauptsächlich aus der Festlegung bestimmter Benutzeroptionen, wie Arbeitsspeicherzuordnung, Standardverzeichnisse und die Verwendung von Klängen und Farben. Außerdem können Sie die Knotenpalette anpassen, die sich unten im SPSS Modeler-Fenster befindet.

Festlegen von IBM SPSS Modeler-Optionen

Es gibt verschiedene Möglichkeiten, Optionen für IBM® SPSS® Modeler anzupassen und festzulegen:

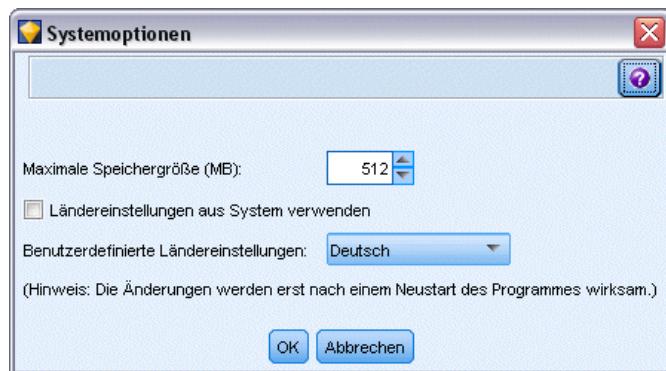
- Festlegen von Systemoptionen, wie Speicherbelegung und Ländereinstellung, durch Klicken auf Systemoptionen im Menü Extras > Optionen.
- Festlegen von Benutzeroptionen, wie Schriftarten und Farben für die Anzeige, durch Klicken auf Benutzeroptionen im Menü Extras > Optionen.
- Festlegen des Verzeichnisses für die Anwendungen, die mit SPSS Modeler verwendet werden können, durch Klicken auf Hilfsprogramme im Menü Extras > Optionen.
- Festlegen der in SPSS Modeler verwendeten Standardverzeichnisse durch Klicken auf Verzeichnis festlegen oder Serververzeichnis festlegen im Menü “Datei”.

Sie können auch Optionen festlegen, die für einige Ihrer Streams oder für alle Streams gelten. Für weitere Informationen siehe Thema [Festlegen von Optionen für Streams in Kapitel 5 auf S. 62](#).

Systemoptionen

Sie können die bevorzugte Sprache oder Ländereinstellung für IBM® SPSS® Modeler festlegen, indem Sie im Menü Extras > Optionen auf Systemoptionen klicken. Hier können Sie auch die maximale Speicherbelegung für SPSS Modeler angeben. Beachten Sie, dass die in diesem Dialogfeld vorgenommenen Änderungen so lange unwirksam sind, bis Sie SPSS Modeler erneut starten.

Abbildung 12-1
Dialogfeld "Systemoptionen"



Maximale Speichergröße. Wählen Sie diese Option, um einen Grenzwert in Megabyte für die Speicherbelegung von SPSS Modeler festzulegen. Auf einigen Plattformen beschränkt SPSS Modeler die Prozessgröße, um die Belastung der Computer mit begrenzten Ressourcen oder hoher Auslastung zu reduzieren. Wenn Sie große Datenmengen bearbeiten, führt dies eventuell zu einer Fehlermeldung, dass nicht genügend Speicher vorhanden ist. Sie können die Speicherauslastung verbessern, indem Sie einen neuen Schwellenwert angeben.

Ländereinstellungen aus System verwenden. Diese Option wird standardmäßig ausgewählt und auf "Englisch" (USA) gesetzt. Deaktivieren Sie diese Option, um eine andere Sprache aus der Liste mit verfügbaren Sprachen und Ländereinstellungen auszuwählen.

Speicherverwaltung

Zusätzlich zur Einstellung Maximale Speichergröße, die Sie im Dialogfeld "Systemoptionen" angegeben haben, gibt es mehrere Möglichkeiten zur Optimierung der Speicherbelegung:

- Festlegen eines Caches an jedem beliebigen Knoten, der kein Endknoten ist, sodass die Daten aus dem Cache gelesen und nicht aus der Datenquelle abgerufen werden, wenn Sie den Daten-Stream ausführen. So wird die Speicherauslastung für große Datenmengen reduziert. [Für weitere Informationen siehe Thema Cachen von Optionen für Knoten in Kapitel 5 auf S. 57.](#)
- Ändern Sie die Option Maximale Elemente für nominale Felder im Dialogfeld für die Stream-Eigenschaften. Diese Option legt eine maximale Anzahl Mitglieder für nominale Felder fest, nach denen das Messniveau des Felds zu *ohne Typ* wird. [Für weitere Informationen siehe Thema Festlegen von allgemeinen Optionen für Streams in Kapitel 5 auf S. 63.](#)
- Zwingen Sie IBM® SPSS® Modeler, Speicher freizugeben, indem Sie in die rechte untere Ecke des Fensters klicken, in der der von SPSS Modeler belegte Speicher und die zugewiesene Speichermenge angezeigt werden (xxMB/xxMB). Wenn Sie in diesen Bereich klicken, wird dieser dunkler und die Zahlen für die Speicherbelegung werden niedriger. Sobald der Bereich erneut die reguläre Farbe annimmt, hat SPSS Modeler den gesamten möglichen Speicher freigegeben.

Festlegen von Standardverzeichnissen

Sie können das für Dateibrowser und die Ausgabe verwendete Standardverzeichnis angeben, indem Sie im Menü “Datei” Verzeichnis festlegen oder Server-Verzeichnis festlegen auswählen.

- **Verzeichnis festlegen.** Sie können diese Option verwenden, um das Arbeitsverzeichnis festzulegen. Das Standardarbeitsverzeichnis basiert auf dem Installationspfad Ihrer IBM® SPSS® Modeler-Version oder auf dem Befehlszeilenpfad, der zum Aufrufen von SPSS Modeler verwendet wird. Im lokalen Modus ist das Arbeitsverzeichnis der Pfad, der für alle clientseitigen Operationen und Ausgabedateien verwendet wird (falls diese mit relativen Pfaden angegeben werden).
- **Server-Verzeichnis festlegen.** Die Option “Server-Verzeichnis festlegen” im Menü “Datei” wird aktiviert, wenn eine Remote-Server-Verbindung hergestellt wird. Verwenden Sie diese Option, um das Standardverzeichnis für alle Server- und Datendateien festzulegen, die für die Ein- und Ausgabe angegeben sind. Das Server-Standardverzeichnis lautet `$CLEO/data`, wobei `$CLEO` das Verzeichnis ist, in dem die Server-Version von SPSS Modeler installiert ist. Mithilfe der Befehlszeile können Sie diesen Standardwert auch überschreiben, indem Sie das Flag `-server_directory` mit dem Befehlszeilenargument `modelerclient` verwenden.

Festlegen von Benutzeroptionen

Sie können allgemeine Optionen für IBM® SPSS® Modeler festlegen, indem Sie im Menü Extras > Optionen den Eintrag Benutzeroptionen auswählen. Diese Optionen gelten für alle in SPSS Modeler verwendeten Streams.

Die folgenden Optionen können festgelegt werden, indem Sie auf die entsprechende Registerkarte klicken:

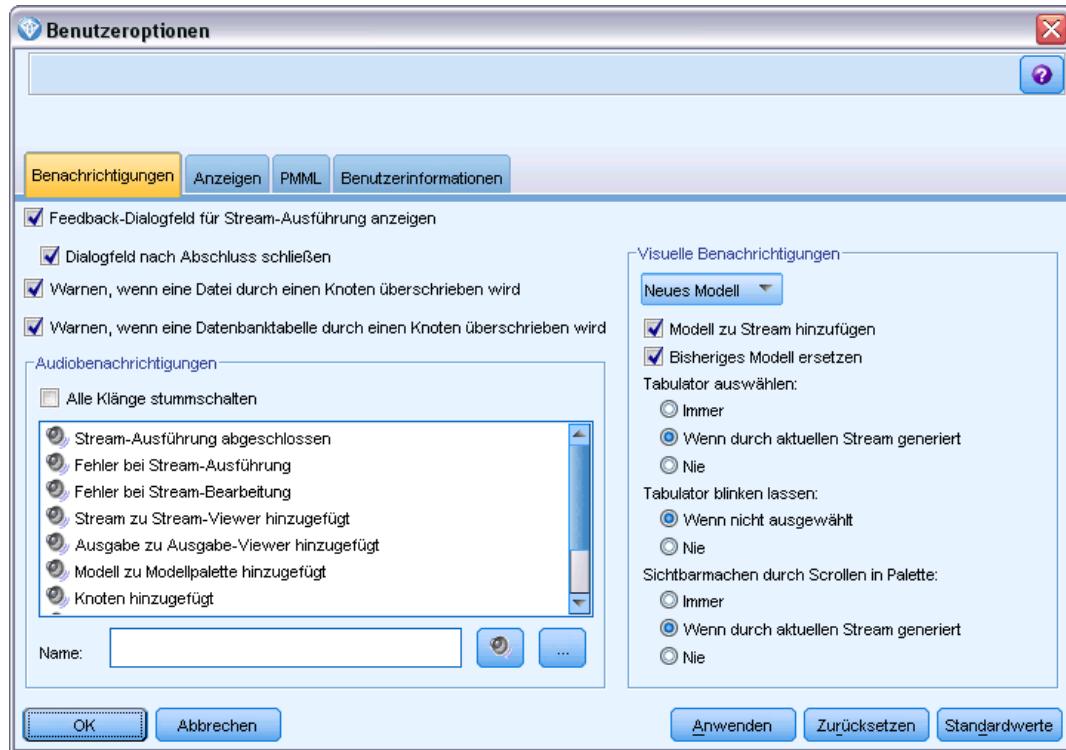
- Benachrichtigungsoptionen, wie Modellüberschreibung und Fehlermeldungen.
- Anzeigeoptionen, wie Diagramm- und Hintergrundfarben.
- PMML-Exportoptionen, die beim Export von Modellen in Predictive Model Markup Language (PMML) verwendet werden.
- Informationen zum Benutzer bzw. Autor, wie Name, Initialen und E-Mail-Adresse. Diese Informationen können auf der Registerkarte “Anmerkungen” für Knoten und andere von Ihnen erstellte Objekte angezeigt werden.

Um Stream-spezifische Optionen, wie Dezimaltrennzeichen, Uhrzeit- und Datumsformate, Optimierung, Stream-Layout sowie Stream-Skripts, festzulegen, verwenden Sie das Dialogfeld “Stream-Eigenschaften”, das über die Menüs “Datei” und “Extras” aufgerufen werden kann.

Festlegen von Benachrichtigungsoptionen

Mithilfe der Registerkarte “Benachrichtigungen” im Dialogfeld “Benutzeroptionen” können Sie verschiedene Optionen bzgl. des Auftretens und des Typs von Warnungen und Bestätigungsfenstern in IBM® SPSS® Modeler festlegen. Sie können auch das Verhalten der Registerkarten “Ausgaben” und “Modelle” im Manager-Bereich festlegen, wenn eine neue Ausgabe und neue Modelle generiert werden.

Abbildung 12-2
Dialogfeld "Benutzeroptionen," Registerkarte "Benachrichtigungen"



Feedback-Dialogfeld für Stream-Ausführung anzeigen. Wählen Sie diese Option aus, um ein Dialogfeld mit Fortschrittsanzeige einzublenden, wenn ein Stream seit drei Sekunden ausgeführt wird. Das Dialogfeld enthält auch Einzelheiten zu den Ausgabeobjekten, die vom Stream erstellt werden.

- **Dialogfeld nach Beenden schließen.** Das Dialogfeld wird standardmäßig am Ende der Stream-Ausführung geschlossen. Deaktivieren Sie dieses Kontrollkästchen, wenn das Dialogfeld sichtbar bleiben soll, nachdem der Stream beendet ist.

Warnen, wenn eine Datei durch einen Knoten überschrieben wird. Wählen Sie diese Option, wenn eine Fehlermeldung erscheinen soll, wenn eine bereits vorhandene Datei durch Knotenoperationen überschrieben wird.

Warnen, wenn eine Datenbanktabelle durch einen Knoten überschrieben wird. Wählen Sie diese Option, wenn eine Fehlermeldung erscheinen soll, wenn eine bereits vorhandene Datenbanktabelle durch Knotenoperationen überschrieben wird.

Akustische Benachrichtigungen

Verwenden Sie die Liste, um festzulegen, ob Sie bei Ereignissen und Fehlern durch ein Tonsignal benachrichtigt werden möchten. Es stehen mehrere Klänge zur Verfügung. Mit der Wiedergabe-Schaltfläche (Lautsprecher) geben Sie einen ausgewählten Klang wieder. Mit der Auslassungsschaltfläche (...) wählen Sie einen Klang aus.

Hinweis: Die .wav-Dateien, mit denen in SPSS Modeler Klänge erstellt werden, sind im Verzeichnis */media/sounds* Ihrer Installation gespeichert.

- **Alle Klänge stummschalten.** Wählen Sie diese Option, um die akustische Benachrichtigung für alle Ereignisse zu deaktivieren.

Visuelle Benachrichtigungen

Die Optionen in dieser Gruppe werden verwendet, um das Verhalten der Manager-Registerkarten "Ausgaben" und "Modelle" oben rechts beim Generieren neuer Elemente festzulegen. Wählen Sie Neues Modell oder Neue Ausgabe in der Liste aus, um das Verhalten der entsprechenden Registerkarte festzulegen.

Die folgenden Optionen sind für Neues Modell verfügbar:

Modell zu Stream hinzufügen. Wenn ausgewählt (Standard), wird dem Stream sowie der Registerkarte "Modelle" ein neues Modell hinzugefügt, sobald das Modell erstellt wird. Im Stream wird das Modell mit einer Verknüpfung zum Modellierungsknoten gezeigt, von dem aus das Modell erstellt wurde. Wenn Sie dieses Kontrollkästchen deaktivieren, wird das Modell nur der Registerkarte "Modelle" hinzugefügt.

Bisheriges Modell ersetzen. Wenn ausgewählt (Standard), wird ein bestehendes Modell aus diesem Stream in der Registerkarte "Modelle" und im Stream-Zeichenbereich überschrieben. Wenn dieses Kontrollkästchen deaktiviert ist, wird das Modell den vorhandenen Modellen in der Registerkarte und im Zeichenbereich hinzugefügt. Beachten Sie, dass diese Einstellung durch die Modellersetzungseinstellung bei einer Modellverknüpfung überschrieben wird. Für weitere Informationen siehe Thema [Ersetzen eines Modells in Kapitel 3 in IBM SPSS Modeler 15](#) *Modellierungsknoten*.

Die folgenden Optionen sind für Neue Ausgabe verfügbar:

Warnung, wenn Ausgabe größer als [n]. Wählen Sie aus, ob eine Warnung angezeigt werden soll, wenn die Anzahl der Elemente auf der Registerkarte "Ausgaben" einen zuvor angegebenen Wert übersteigt. Der Standardwert ist 20; Sie können diesen Wert jedoch bei Bedarf ändern.

Die folgenden Optionen sind in allen Fällen verfügbar:

Tabulator auswählen. Wählen Sie, ob Sie zur Registerkarte "Ausgaben" oder "Modelle" wechseln möchten, wenn das entsprechende Objekt während der Stream-Ausführung generiert wird.

- Wählen Sie Immer, um zur entsprechenden Registerkarte im Manager-Bereich zu wechseln.
- Wählen Sie Wenn durch aktuellen Stream generiert, um nur bei Objekten, die von dem Stream generiert wurden, der derzeit im Zeichenbereich sichtbar ist, zur entsprechenden Registerkarte zu wechseln.
- Wählen Sie Nie, um die Software so zu beschränken, dass nicht zu der entsprechenden Registerkarte gewechselt wird, um Sie über generierte Ausgaben oder Modelle zu benachrichtigen.

Tabulator blinken lassen. Legen Sie hier fest, ob die Registerkarte "Ausgaben" oder "Modelle" im Manager-Bereich blinken soll, wenn neue Ausgaben bzw. Modelle generiert wurden.

- Wählen Sie Wenn nicht ausgewählt, damit die entsprechende Registerkarte blinkt (falls nicht bereits ausgewählt), sobald neue Objekte im Manager-Bereich generiert wurden.
- Wählen Sie Nie, um die Software so zu beschränken, dass die entsprechende Registerkarte nicht blinkt, um Sie über generierte Objekte zu benachrichtigen.

Sichtbarmachen durch Scrollen in Palette (Nur neue Modelle). Legen Sie fest, ob Sie durch die Registerkarte “Modelle” im Manager-Bereich automatisch blättern möchten, damit das aktuellste Modell sichtbar wird.

- Wählen Sie Immer, um das Scrollen zu aktivieren.
- Wählen Sie Wenn durch aktuellen Stream generiert, um nur nach Objekten zu suchen, die von dem Stream generiert wurden, der derzeit im Zeichenbereich sichtbar ist.
- Wählen Sie Nie, um die Software so zu beschränken, dass durch die Registerkarte “Modelle” nicht automatisch geblättert wird.

Fenster öffnen (Nur Neue Ausgabe). Legen Sie fest, ob nach der Generierung automatisch ein Ausgabefenster geöffnet werden soll.

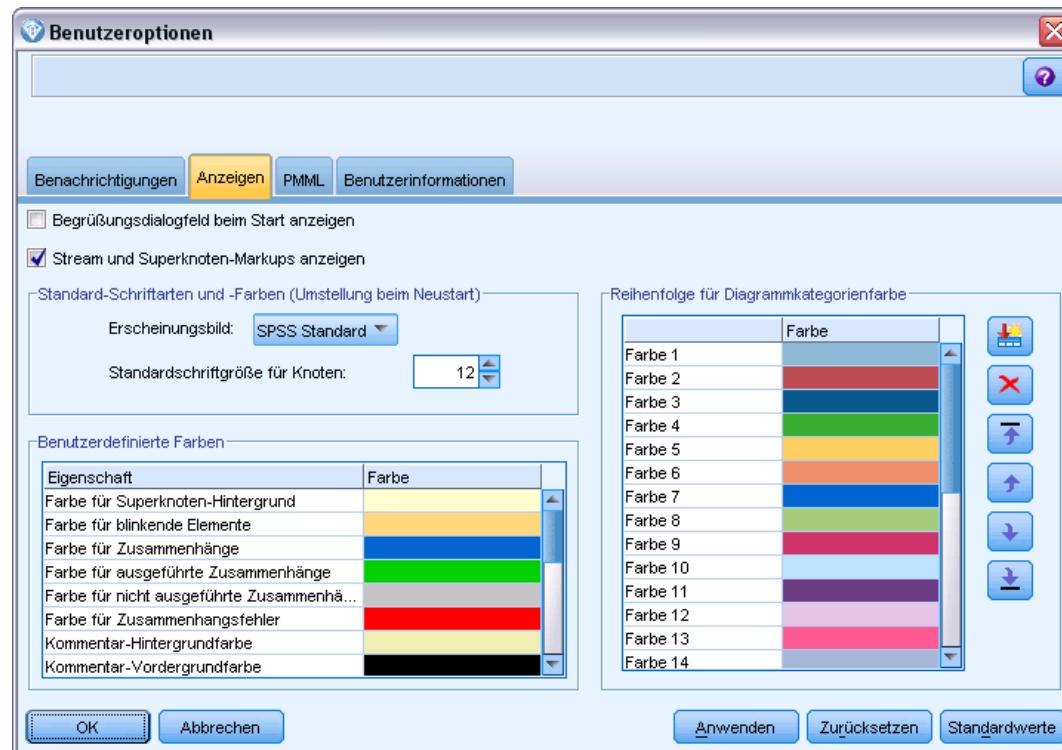
- Wählen Sie Immer, um immer ein neues Ausgabefenster zu öffnen.
- Wählen Sie Wenn durch aktuellen Stream generiert, um ein neues Fenster für die Ausgabe zu öffnen, die von dem Stream generiert wurde, der derzeit im Zeichenbereich sichtbar ist.
- Wählen Sie Nie, um die Software so zu beschränken, dass keine neuen Fenster für die generierte Ausgabe automatisch geöffnet werden.

Klicken Sie auf Standardwerte, um zu den Systemstandardeinstellungen für diese Registerkarte zurückzukehren.

Festlegen von Anzeigeoptionen

Auf der Registerkarte “Anzeige” im Dialogfeld “Benutzeroptionen” können Sie die Optionen für die Anzeige von Schriftarten und Farben in IBM® SPSS® Modeler festlegen.

Abbildung 12-3
Dialogfeld "Benutzeroptionen," Registerkarte "Anzeige"



Begrüßungsdialogfeld beim Start anzeigen. Wählen Sie diese Option aus, um das Begrüßungsdialogfeld beim Start anzuzeigen. Das Begrüßungsdialogfeld enthält Optionen, mit denen das Lernprogramm mit Anwendungsbeispielen gestartet, ein Demo-Stream bzw. ein bestehender Stream oder ein bestehendes Projekt geöffnet oder ein neuer Stream erstellt werden kann.

Stream- und Superknoten-Markups anzeigen. Wählen Sie diese Option, damit vorhandene Markups zu Streams und Superknoten standardmäßig angezeigt werden. Markup umfasst Stream-Kommentare, Modellverknüpfungen und die Hervorhebung von Scoring-Verzweigungen.

Standard-Schriftarten & Farben (wirksam nach Neustart). Die Optionen in diesem Kontrollfeld werden verwendet, um das Bildschirmdesign und das Farbschema von SPSS Modeler sowie die Größe der angezeigten Schriftarten festzulegen. Die hier ausgewählten Optionen werden erst wirksam, wenn Sie SPSS Modeler schließen und neu starten.

- **Erscheinungsbild.** Ermöglicht Ihnen, ein Standard-Farbschema und Bildschirmdesign zu wählen. Sie haben folgende Möglichkeiten:

SPSS Standard (Standard), ein gemeinsames Design für alle Produkte von IBM SPSS.

SPSS Classic, ein Design, mit dem Benutzer früherer SPSS Modeler-Versionen vertraut sind.

Windows, ein Windows-Design, das u. U. hilfreich für einen höheren Kontrast im Stream-Zeichenbereich und in den Paletten ist.

- **Standardschriftgröße für Knoten.** Geben Sie eine Schriftartengröße an, die in den Knotenpaletten verwendet werden kann, sowie für Knoten, die im Stream-Zeichenbereich angezeigt werden.

Hinweis: Sie können die Größe der Knotensymbole für einen Stream im Fenster “Layout” der Registerkarte “Optionen” des Dialogfelds mit den Stream-Eigenschaften festlegen. Wählen Sie im Hauptmenü Folgendes:

Werkzeuge > Stream-Eigenschaften > Optionen > Layout.

Benutzerdefinierte Farben. In dieser Tabelle werden die aktuell ausgewählten Farben aufgelistet, die für verschiedene Anzeigeelemente verwendet werden. Sie können für jedes in der Tabelle aufgeführte Element die aktuelle Farbe ändern, indem Sie in die entsprechende Zeile in der Spalte Farbe doppelklicken und eine Farbe aus der Liste auswählen. Um eine benutzerdefinierte Farbe festzulegen, blättern Sie in der Liste nach unten und klicken Sie auf den Eintrag Farbe....

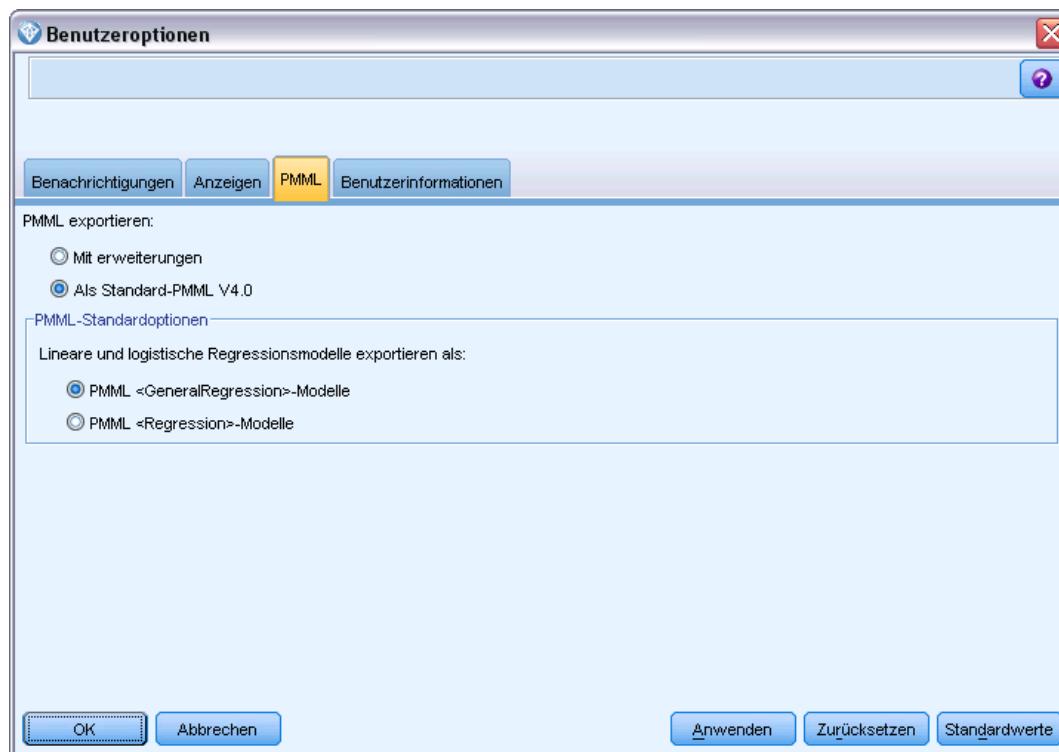
Reihenfolge für Diagrammkategorienfarbe. In dieser Tabelle werden die aktuell ausgewählten Farben aufgelistet, die in den neu erstellten Diagrammen verwendet werden. Die Reihenfolge der Farben spiegelt die Reihenfolge wider, in der Sie im Diagramm verwendet werden. Wenn z. B. ein als eine Farüberlagerung verwendetes nominales Feld vier eindeutige Werte enthält, werden lediglich die ersten vier der hier aufgeführten Farben verwendet. Sie können für jedes in der Tabelle aufgeführte Element die aktuelle Farbe ändern, indem Sie in die entsprechende Zeile in der Spalte Farbe doppelklicken und eine Farbe aus der Liste auswählen. Um eine benutzerdefinierte Farbe festzulegen, blättern Sie in der Liste nach unten und klicken Sie auf den Eintrag Farbe.... Hier vorgenommene Änderungen wirken sich nicht auf zuvor erstellte Diagramme aus.

Klicken Sie auf Standardwerte, um zu den Systemstandardeinstellungen für diese Registerkarte zurückzukehren.

Festlegen von PMML-Exportoptionen

Auf der Registerkarte “PMML” können Sie steuern, wie IBM® SPSS® Modeler Modelle in die Predictive Model Markup Language (PMML) exportiert. Für weitere Informationen siehe Thema Importieren und Exportieren von Modellen als PMML in Kapitel 10 auf S. 220.

Abbildung 12-4
Dialogfeld "Benutzeroptionen," Registerkarte "PMML"



PMML exportieren. Hier können Sie Variationen von PMML konfigurieren, die am besten mit Ihrer Zielanwendung funktionieren.

- Wählen Sie mit Erweiterungen, um PMML-Erweiterungen für Sonderfälle zuzulassen, für die es keine Standard-PMML-Entsprechung gibt. Beachten Sie, dass dies in den meisten Fällen zum selben Ergebnis führt wie Standard-PMML.
- Wählen Sie Als Standard-PMML..., um PMML zu exportieren, die so nah wie möglich am PMML-Standard bleibt.

PMML-Standardoptionen. Wenn oben Als Standard-PMML... ausgewählt ist, können Sie eine von zwei gültigen Möglichkeiten auswählen, um lineare und logistische Regressionsmodelle zu exportieren:

- Als PMML <GeneralRegression<-Modelle
- Als PMML <Regression>-Modelle

Weitere Informationen zu PMML finden Sie auf der Website der Data Mining Group unter www.dmg.org.

Festlegen von Benutzerinformationen

Benutzer-/Autoreninformationen. Hier eingegebene Informationen können auf der Registerkarte "Anmerkungen" für Knoten und andere von Ihnen erstellte Objekte angezeigt werden.

Anpassen der Knotenpalette

Streams werden mithilfe von Knoten erstellt. Die Knotenpalette unten im IBM® SPSS® Modeler-Fenster enthält alle Knoten, die bei der Stream-Erstellung verwendet werden können. Für weitere Informationen siehe Thema Knotenpalette in Kapitel 3 auf S. 22.

Sie haben zwei Möglichkeiten für die Neuanordnung der Knotenpalette:

- Anpassen des Paletten-Managers. Für weitere Informationen siehe Thema Anpassen des Paletten-Managers auf S. 250.
- Ändern der Art und Weise, wie Paletten-Registerkarten, die Unterpaletten enthalten, auf der Knotenpalette angezeigt werden. Für weitere Informationen siehe Thema Erstellen einer Unterpalette auf S. 255.

Abbildung 12-5
Die Registerkarte "Datensatzoperationen" in der Knotenpalette



Anpassen des Paletten-Managers

Der Paletten-Manager kann entsprechend Ihrer Nutzung von IBM® SPSS® Modeler angepasst werden. Wenn Sie z. B. häufig Zeitreihendaten aus einer Datenbank analysieren, möchten Sie u. U. sicherstellen, dass der Datenbankquellenknoten, der Zeitintervallknoten, der Zeitreihenknoten und der Zeitdiagrammknoten gemeinsam auf einer speziellen Registerkarte der Palette zur Verfügung stehen. Im Paletten-Manager können Sie diese Anpassungen problemlos vornehmen, indem Sie auf der Knotenpalette benutzerdefinierte Paletten-Registerkarten erstellen.

Mit dem Paletten-Manager können Sie verschiedene Aufgaben durchführen:

- Festlegen, welche Palettenregisterkarten auf der Knotenpalette unterhalb des Stream-Zeichenbereichs angezeigt werden sollen.
- Ändern der Reihenfolge, in der die Palettenregisterkarten auf der Knotenpalette angezeigt werden.
- Erstellen und Bearbeiten eigener Palettenregisterkarten und ggf. zugehöriger Unterpaletten.
- Bearbeiten der Standard-Knotenauswahl auf der Registerkarte "Favoriten".

So rufen Sie den Paletten-Manager auf:

- ▶ Klicken Sie im Menü "Extras" auf die Option Paletten verwalten.

Abbildung 12-6
Paletten-Manager mit den auf der Knotenpalette angezeigten Registerkarten



Palettenname. Jede verfügbare Palettenregisterkarte wird aufgeführt, unabhängig davon, ob sie auf der Knotenpalette angezeigt wird oder nicht. Dies umfasst ggf. auch alle von Ihnen selbst erstellten Palettenregisterkarten. [Für weitere Informationen siehe Thema Erstellen einer Palettenregisterkarte auf S. 252.](#)

Anzahl der Knoten. Die Anzahl der auf den einzelnen Palettenregisterkarten angezeigten Knoten. Ein hoher Wert bedeutet, dass es praktischer sein kann, Unterpaletten zu erstellen, um die Knoten auf der Registerkarte zu unterteilen. [Für weitere Informationen siehe Thema Erstellen einer Unterpalette auf S. 255.](#)

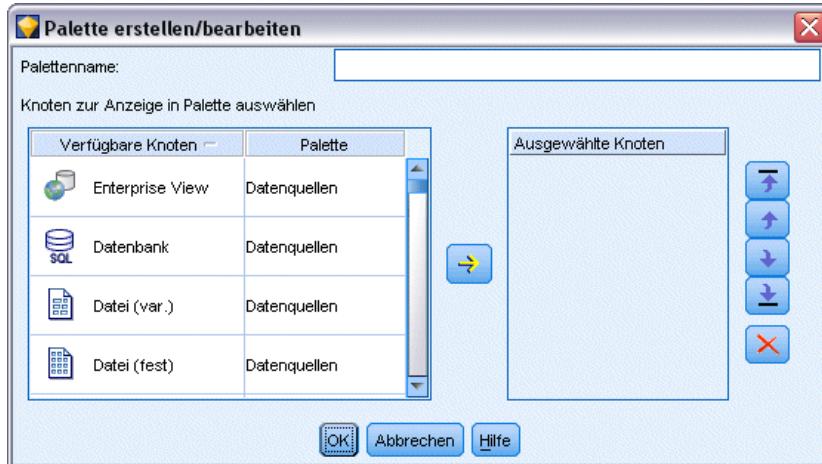
Angezeigt? Wählen Sie dieses Feld aus, um die Palettenregisterkarte auf der Knotenpalette anzuzeigen. [Für weitere Informationen siehe Thema Anzeigen von Palettenregisterkarten auf der Knotenpalette auf S. 252.](#)

Unterpalten. Um Unterpaletten zur Anzeige auf einer Palettenregisterkarte auszuwählen, markieren Sie den entsprechenden Palettennamen und klicken Sie auf diese Schaltfläche, um die Registerkarte "Unterpalten" anzuzeigen. [Für weitere Informationen siehe Thema Erstellen einer Unterpalette auf S. 255.](#)

Standardeinstellungen wiederherstellen. Durch Klicken auf diese Schaltfläche können Sie alle Änderungen und Zusätze, die Sie an den Paletten und Unterpaletten vorgenommen haben, vollständig entfernen und zu den Standardeinstellungen für die Palette zurückkehren.

Erstellen einer Palettenregisterkarte

Abbildung 12-7
Erstellung von Palettenregisterkarten mit dem Dialogfeld "Palette erstellen/bearbeiten"



So erstellen Sie eine benutzerdefinierte Palettenregisterkarte:

- ▶ Öffnen Sie im Menü "Extras" die Option "Paletten-Manager".
- ▶ Klicken Sie rechts neben der Spalte *Angezeigt?* auf die Schaltfläche zum Hinzufügen von Paletten; das Dialogfeld "Palette erstellen/bearbeiten" wird angezeigt.
- ▶ Geben Sie einen eindeutigen Palettennamen ein.
- ▶ Wählen Sie im Bereich *Verfügbare Knoten* den Knoten aus, der zur Palettenregisterkarte hinzugefügt werden soll.
- ▶ Klicken Sie unter "Knoten hinzufügen" auf den nach rechts weisenden Pfeil, um den markierten Knoten in den Bereich *Ausgewählte Knoten* zu verschieben. Wiederholen Sie den Vorgang, bis Sie alle gewünschten Knoten hinzugefügt haben.

Nachdem alle erforderlichen Knoten hinzugefügt wurden, können Sie die Reihenfolge ändern, in der sie auf der Palettenregisterkarte angezeigt werden:

- ▶ Mithilfe der einfachen Pfeilschaltflächen können Sie einen Knoten eine Zeile nach oben bzw. unten verschieben.
- ▶ Mithilfe der Schaltflächen mit Pfeil und Balken können Sie einen Knoten ganz nach unten oder oben in der Liste verschieben.
- ▶ Um einen Knoten aus einer Palette zu entfernen, markieren Sie den Knoten und klicken Sie auf die Schaltfläche "Löschen" rechts neben dem Bereich *Ausgewählte Knoten*.

Anzeigen von Palettenregisterkarten auf der Knotenpalette

Es kann in IBM® SPSS® Modeler Optionen geben, die Sie nie benutzen. In diesem Fall können Sie die Registerkarten mit den entsprechenden Knoten mithilfe des Paletten-Managers ausblenden.

Abbildung 12-8
Paletten-Manager mit den auf der Knotenpalette angezeigten Registerkarten



So können Sie auswählen, welche Registerkarten auf der Knotenpalette angezeigt werden sollen:

- ▶ Öffnen Sie im Menü “Extras” die Option “Paletten-Manager”.
- ▶ Mithilfe der Kontrollkästchen in der Spalte *Angezeigt?* können Sie auswählen, ob die einzelnen Palettenregisterkarten ein- oder ausgeblendet werden sollen.

Um eine Palettenregisterkarte endgültig aus der Knotenpalette zu entfernen, markieren Sie den Knoten und klicken Sie auf die Schaltfläche “Löschen” rechts neben der Spalte *Angezeigt?*. Gelöschte Palettenregisterkarten können nicht wiederhergestellt werden.

Hinweis: Mit Ausnahme der Registerkarte “Favoriten” können die im Lieferumfang von SPSS Modeler enthaltenen Standard-Palettenregisterkarten nicht gelöscht werden.

Ändern der Anzeigereihenfolge auf der Knotenpalette

Nachdem Sie ausgewählt haben, welche Palettenregisterkarten angezeigt werden sollen, können Sie die Reihenfolge ändern, in der sie auf der Knotenpalette angezeigt werden:

- ▶ Mithilfe der einfachen Pfeilschaltflächen können Sie eine Palettenregisterkarte eine Zeile nach oben bzw. unten verschieben. Beim Verschieben nach oben werden sie links neben die Knotenpalette verschoben und umgekehrt.
- ▶ Mithilfe der Schaltflächen mit Pfeil und Balken können Sie eine Palettenregisterkarte ganz nach unten oder oben in der Liste verschieben. Die Elemente oben in der Liste werden links neben der Knotenpalette angezeigt.

Anzeigen von Unterpaletten auf einer Palettenregisterkarte

So wie Sie festlegen können, welche Palettenregisterkarten auf der Knotenpalette angezeigt werden sollen, so können Sie auch festlegen, welche Unterpaletten auf der zugehörigen übergeordneten Palettenregisterkarte verfügbar sein sollen.

Abbildung 12-9

Für die Palettenregisterkarte "Modellierung" zur Verfügung stehende Unterpaletten



So wählen Sie die Unterpaletten aus, die auf einer Palettenregisterkarte angezeigt werden sollen:

- ▶ Öffnen Sie im Menü "Extras" die Option "Paletten-Manager".
- ▶ Wählen Sie die erforderliche Palette aus.
- ▶ Klicken Sie auf die Schaltfläche "Unterpaletten"; das Dialogfeld "Unterpaletten" wird angezeigt.
- ▶ Mithilfe der Kontrollkästchen in der Spalte *Angezeigt?* können Sie auswählen, ob die einzelnen Unterpaletten auf der Palettenregisterkarte erscheinen sollen. Die Unterpalette Alle wird immer angezeigt und kann nicht gelöscht werden.
- ▶ Um eine Unterpalette endgültig aus der Palettenregisterkarte zu entfernen, markieren Sie die Unterpalette und klicken Sie auf die Schaltfläche "Löschen" rechts neben der Spalte *Angezeigt?*.

Hinweis: Die bereits im Lieferumfang enthaltenen Standard-Unterpaletten der Palettenregisterkarte "Modellierung" können nicht gelöscht werden.

Ändern der Anzeigereihenfolge auf der Palettenregisterkarte

Nachdem Sie ausgewählt haben, welche Unterpaletten angezeigt werden sollen, können Sie die Reihenfolge ändern, in der sie auf der übergeordneten Palettenregisterkarte angezeigt werden:

- ▶ Mithilfe der einfachen Pfeilschaltflächen können Sie eine Unterpalette eine Zeile nach oben bzw. unten verschieben.
- ▶ Mithilfe der Schaltflächen mit Pfeil und Balken können Sie eine Unterpalette ganz nach unten oder oben in der Liste verschieben.

Die von Ihnen erstellten Unterpaletten werden auf der Knotenpalette angezeigt, wenn Sie die zugehörige übergeordnete Palettenregisterkarte auswählen. Für weitere Informationen siehe Thema Ändern der Ansicht einer Palettenregisterkarte auf S. 256.

Erstellen einer Unterpalette

Da Sie in die von Ihnen erstellten benutzerdefinierten Palettenregisterkarten jeden beliebigen bestehenden Knoten aufnehmen können, ist es möglich, dass Sie mehr Knoten auswählen, als ohne Bildlauf auf dem Bildschirm angezeigt werden können. Um die Notwendigkeit eines Bildlaufs zu vermeiden, können Sie Unterpaletten erstellen, in die Sie die Knoten aufnehmen können, die Sie für die Palettenregisterkarte ausgewählt haben. Wenn Sie beispielsweise eine Palettenregisterkarte erstellt haben, die die Knoten enthält, die Sie am häufigsten zum Erstellen von Streams verwenden, können Sie vier Unterpaletten erstellen, die die Auswahl nach Quellenknoten, Feldoperationen, Modellierung und Ausgabe aufgliedern.

Hinweis: Als Knoten für die Unterpaletten stehen nur diejenigen Knoten zur Auswahl, die zur übergeordneten Palettenregisterkarte hinzugefügt wurden.

Abbildung 12-10
Erstellung von Unterpaletten mit dem Dialogfeld „Unterpalette anzeigen/bearbeiten“



So erstellen Sie eine Unterpalette:

- ▶ Öffnen Sie im Menü „Extras“ die Option „Paletten-Manager“.
- ▶ Wählen Sie die Palette aus, zu der Sie Unterpaletten hinzufügen möchten.
- ▶ Klicken Sie auf die Schaltfläche „Unterpaletten“; das Dialogfeld „Unterpaletten“ wird angezeigt.
- ▶ Klicken Sie rechts neben der Spalte *Angezeigt?* auf die Schaltfläche zum Hinzufügen von Unterpaletten; das Dialogfeld „Unterpalette anzeigen/bearbeiten“ wird angezeigt.
- ▶ Geben Sie einen eindeutigen Unterpalettennamen ein.
- ▶ Wählen Sie im Bereich *Verfügbare Knoten* den Knoten aus, der zur Unterpalette hinzugefügt werden soll.

- ▶ Klicken Sie unter “Knoten hinzufügen” auf den nach rechts weisenden Pfeil, um den ausgewählten Knoten zum Bereich Ausgewählte Knoten hinzuzufügen.
- ▶ Klicken Sie nach dem Hinzufügen der erforderlichen Knoten auf OK, um zum Dialogfeld “Unterpaletten” zurückzukehren.

Die von Ihnen erstellten Unterpaletten werden auf der Knotenpalette angezeigt, wenn Sie die zugehörige übergeordnete Palettenregisterkarte auswählen. Für weitere Informationen siehe [Thema Ändern der Ansicht einer Palettenregisterkarte auf S. 256](#).

Ändern der Ansicht einer Palettenregisterkarte

Aufgrund der großen Anzahl an Knoten, die in IBM® SPSS® Modeler verfügbar sind, sind möglicherweise auf kleinen Bildschirmen nicht alle Knoten sichtbar, ohne dass auf der Knotenpalette ein Bildlauf nach links oder rechts durchgeführt wird. Dieses Problem tritt verstärkt auf der Palettenregisterkarte “Modellierung” auf. Um die Notwendigkeit eines Bildlaufs zu verringern, können Sie auswählen, dass nur die auf einer Unterpalette (sofern verfügbar) enthaltenen Knoten angezeigt werden sollen. Für weitere Informationen siehe [Thema Erstellen einer Unterpalette auf S. 255](#).

Um die auf einer Palettenregisterkarte angezeigten Knoten zu ändern, wählen Sie die Palettenregisterkarte aus und legen dann im Menü auf der linken Seite fest, ob alle Knoten angezeigt werden sollen oder nur diejenigen, die sich auf einer bestimmten Unterpalette befinden.

Abbildung 12-11
Palettenregisterkarte “Modellierung” mit Unterpalette “Klassifizierung”



CEMI-Knotenverwaltung

CEMI wird nun nicht mehr unterstützt und wurde durch CLEF ersetzt, das einen flexibleren und bedienerfreundlichen Funktionssatz bietet. Weitere Informationen finden Sie im *IBM® SPSS® Modeler 15 CLEF-Entwicklerhandbuch*, das in dieser Version enthalten ist.

Erläuterungen zur Leistung für Streams und Knoten

Sie können Ihre Streams so entwerfen, dass die Leistungsfähigkeit maximiert wird, indem Sie die Knoten in der effizientesten Konfiguration anordnen, an den entsprechenden Stellen Knoten-Caches aktivieren und andere Punkte beachten, die in diesem Abschnitt ausgeführt sind.

Neben den hier erörterten Punkten können weitere und deutlichere Leistungsverbesserungen normalerweise durch die effektive Nutzung der Datenbank, insbesondere durch SQL-Optimierung, erzielt werden. [Für weitere Informationen siehe Thema Leistungsüberblick in Kapitel 5 in IBM SPSS Modeler Server 15-Verwaltungs- und Leistungshandbuch.](#)

Knotenreihenfolge

Selbst wenn Sie keine SQL-Optimierung verwenden, kann sich die Reihenfolge der Knoten in einem Stream auf die Leistung auswirken. Das allgemeine Ziel besteht in der Minimierung des Verarbeitungsaufwands weiter unten im Stream. Daher sollten Knoten, die die Datenmenge verringern, sofern vorhanden, relativ an den Anfang des Streams gestellt werden. IBM® SPSS® Modeler Server kann während der Kompilierung automatisch einige Neuordnungsregeln anwenden, um bestimmte Knoten an eine Stelle weiter oben im Stream zu rücken, wenn nachgewiesen werden kann, dass dieser Vorgang keine Probleme mit sich bringt. (Diese Funktion ist standardmäßig aktiviert. Fragen Sie bei Ihrem Systemadministrator nach, ob diese Funktion in Ihrer Installation tatsächlich aktiviert ist.)

Bei der Verwendung der SQL-Optimierung sollen Verfügbarkeit und Effizienz maximiert werden. Da die Optimierung anhält, wenn ein Stream eine Operation enthält, die nicht in der Datenbank durchgeführt werden kann, sollten SQL-optimierte Operationen zusammen am Anfang des Streams gruppiert werden. Diese Strategie sorgt dafür, dass ein größerer Teil der Verarbeitung in der Datenbank erfolgt und somit weniger Daten in IBM® SPSS® Modeler eingebracht werden.

Die folgenden Operationen können in den meisten Datenbanken ausgeführt werden. Versuchen Sie, sie an den *Anfang* des Streams zu stellen:

- Merge nach Schlüssel (join)
- Auswählen
- Aggregieren
- Sortieren
- Muster
- Anhängen
- Verschiedene Vorgänge im Modus *Einschließen*, in dem alle Felder ausgewählt werden
- Füller-Operationen

- Grundlegende Ableitungsoperationen unter Verwendung von Standardarithmetik oder Zeichenkettenbearbeitung (je nachdem, welche Operationen von der Datenbank unterstützt werden)
- Dichotomie

Die folgenden Operationen können in den meisten Datenbanken nicht ausgeführt werden. Sie sollten im Stream *hinter* den Operationen in der vorstehenden Liste platziert werden:

- Operationen für andere Daten als Datenbankdaten, beispielsweise für Textdateien
- Merge nach Reihenfolge
- Balance
- Bestimmte Operationen im Modus *Verwerfen* oder wenn nur eine Untermenge von Feldern als unterschiedlich ausgewählt wurde
- Alle Operationen, bei denen auf Daten zugegriffen werden muss, die sich in einem anderen Datensatz als dem gerade verarbeiteten befinden.
- Ableitungen von Status- und Anzahlfeldern
- Verlaufsknoten-Operationen
- Operationen, die Zeitreihenfunktionen (mit “@”) beinhalten
- Die Typprüfungsmodi *Warnen* und *Abbrechen*
- Modellerstellung, -anwendung und -analyse.

Hinweis: Entscheidungsbäume, Regelmengen, lineare Regression und faktorgenerierte Modelle können SQL erzeugen und daher per Pushback an die Datenbank zurückübertragen werden.

- Datenausgabe an einen anderen Ort als die Datenbank, die die Daten verarbeitet

Knoten-Caches

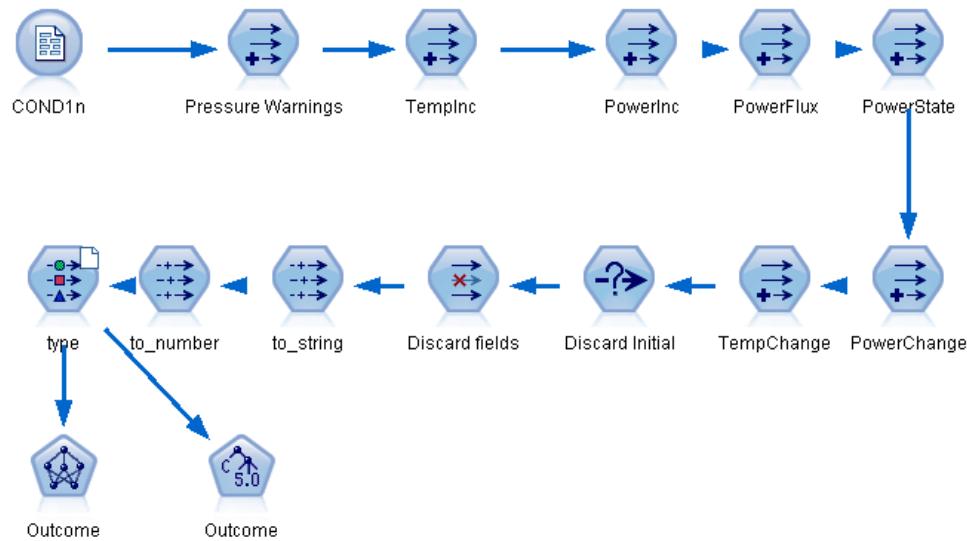
Um die Stream-Ausführung zu optimieren, können Sie einen **Cache** für Knoten einrichten, die keine Endknoten sind. Wenn Sie einen Cache für einen Knoten einrichten, wird der Cache mit den Daten gefüllt, die bei der nächsten Ausführung des Daten-Streams durch den Knoten laufen. Ab diesem Zeitpunkt werden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

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

Hinweis: Ein Caching bei Quellenknoten, bei denen einfach eine Kopie der ursprünglichen Dateien, die in IBM® SPSS® Modeler eingelesen wurden, gespeichert wird, führt unter den meisten Umständen nicht zu einer besseren Leistung.

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

Abbildung 13-1
Cachen am Typknoten zum Speichern neu abgeleiteter Felder



So aktivieren Sie einen Cache:

- ▶ Klicken Sie im Stream-Zeichenbereich mit der rechten Maustaste auf den Knoten und klicken Sie im Menü auf die Option Cache.
- ▶ Klicken Sie im Untermenü auf die Option Aktivieren.
- ▶ Sie können den Cache deaktivieren, indem Sie mit der rechten Maustaste auf den Knoten klicken und im Untermenü auf Deaktivieren klicken.

Cachen von Knoten in einer Datenbank

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

Bei Verwendung von Datenbank-Caching mit Zeichenketten, die mehr als 255 Zeichen umfassen, müssen Sie entweder sicherstellen, dass sich ein Typknoten im Stream oberhalb des Caching-Knotens befindet und dass die Feldwerte gelesen werden, oder Sie müssen die Zeichenkettenlänge mithilfe des Parameters `default_sql_string_length` in der Datei `options.cfg` festlegen. Dies stellt sicher, dass die entsprechende Spalte in der temporären Tabelle auf die korrekte Breite eingestellt wird, um die Strings anzuzeigen. [Für weitere Informationen](#)

siehe Thema Optionen nur in options.cfg sichtbar in Kapitel 4 in *IBM SPSS Modeler Server 15-Verwaltungs- und Leistungshandbuch*.

Um das Caching von Datenbanken nutzen zu können, müssen sowohl SQL-Optimierung als auch Datenbank-Caching aktiviert sein. Die Optimierungseinstellungen auf dem Server haben Vorrang vor jenen auf dem Client. Für weitere Informationen siehe Thema Festlegen von Optimierungsoptionen für Streams in Kapitel 5 auf S. 68.

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

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

Leistung: Prozessknoten

Sortieren. Der Sortierknoten muss das gesamte Eingabe-Daten-Set lesen, bevor es sortiert werden kann. Die Daten werden bis zu einer gewissen Obergrenze im Arbeitsspeicher gespeichert, der Rest wird auf den Datenträger geschrieben. Beim Sortieralgorithmus handelt es sich um einen Kombinationsalgorithmus: Daten werden bis zur angegebenen Grenze in den Arbeitsspeicher eingelesen und mithilfe eines schnellen hybriden Quick-Sort-Algorithmus. Wenn alle Daten in den Arbeitsspeicher passen, ist der Sortiervorgang abgeschlossen. Andernfalls wird der Merge-Sort-Algorithmus angewendet. Die sortierten Daten werden in eine Datei geschrieben und die nächste Datenmenge wird in den Arbeitsspeicher eingelesen, sortiert und auf den Datenträger geschrieben. Dieser Vorgang wird wiederholt, bis alle Daten gelesen wurden; anschließend werden die sortierten Datenmengen zusammengeführt (Merge). Für die Zusammenführung können mehrere Durchläufe über die auf dem Datenträger gespeicherten Daten erforderlich sein. Bei Spitzennutzung hat der Sortierknoten zwei vollständige Kopien des Daten-Sets auf dem Datenträger gespeichert: eine sortierte und eine unsortierte Version.

Die Gesamlaufzeit des Algorithmus liegt in der Größenordnung von $N * \log(N)$. Dabei ist N die Anzahl der Datensätze. Die Sortierung im Arbeitsspeicher ist schneller als die Zusammenführung über den Datenträger, sodass sich die tatsächliche Laufzeit durch die Zuweisung von mehr Arbeitsspeicher zum Sortiervorgang reduziert lässt. Der Algorithmus weist sich selbst einen Teil des physikalischen RAM zu, der von der IBM® SPSS® Modeler Server-Konfigurationsoption *Multiplikator für Arbeitsspeicherverwendung* gesteuert wird. Um den für die Sortierung verwendeten Arbeitsspeicher zu erhöhen, müssen Sie mehr physikalischen RAM bereitstellen oder diesen Wert erhöhen. Beachten Sie: Wenn der Anteil des verwendeten Arbeitsspeichers die Arbeitsmenge des Prozesses überschreitet, sodass ein Teil des Arbeitsspeichers auf den Datenträger ausgelagert wird, lässt die Leistung nach, da das Muster für den Arbeitsspeicherzugriff des Arbeitsspeicherinternen Sortieralgorithmus zufällig ist und zu übermäßigem Paging führen kann. Der Sortieralgorithmus wird außer vom Sortierknoten noch von mehreren anderen Knoten verwendet; es gelten jedoch dieselben Leistungsregeln.

Klassieren. Der Klasserknoten liest das gesamte Eingabe-Daten-Set ein, um die Klassengrenzen zu berechnen, bevor er Datensätze zu Klassen zuweist. Das Daten-Set wird während der Berechnung der Grenzen im Cache gespeichert; anschließend wird es für die Zuweisung erneut überprüft. Bei der Klassermethode *Feste Breite* oder *Mittelwert/Standardabweichung* wird das Daten-Set direkt auf dem Datenträger im Cache gespeichert. Diese Methoden weisen eine lineare Laufzeit auf und erfordern so viel Speicherplatz auf dem Datenträger, dass das gesamte Daten-Set gespeichert werden kann. Bei der Klassermethode *Ränge* oder *n-tile* wird das Daten-Set mithilfe des zuvor beschriebenen Sortieralgorithmus sortiert und das sortierte Daten-Set wird als Cache verwendet. Die Sortierung führt bei diesen Methoden zu einer Laufzeit von $M*N*\log(N)$. Dabei ist M die Anzahl der klassierten Felder und N ist die Anzahl der Datensätze. Auf dem Datenträger muss das Doppelte der Daten-Set-Größe an Speicherplatz verfügbar sein.

Das Generieren eines Ableitungsknotens auf der Grundlage von generierten Klassen verbessert die Leistung in den nachfolgenden Durchgängen. Ableitungsoperationen sind wesentlich schneller als das Klassieren. Für weitere Informationen siehe Thema [Vorschau der generierten Klassen in Kapitel 4 in IBM SPSS Modeler 15 Quellen-, Prozess- und Ausgabeknoten](#).

Merge nach Schlüssel (join) Bei der Merge-Methode *Schlüssel* (entspricht einem Datenbank-Join) sortiert der Merge-Knoten alle Eingabe-Daten-Sets anhand der Schlüsselfelder. Dieser Teil des Verfahrens weist eine Laufzeit von $M*N*\log(N)$ auf. Dabei ist M die Anzahl der Eingaben und N die Anzahl der Datensätze in der größten Eingabe: Es erfordert genügend Datenträgerspeicher, um alle Eingabe-Daten-Sets sowie eine zweite Kopie des größten Daten-Sets zu speichern. Die Laufzeit des Merging selbst ist proportional zur Größe des Ausgabe-Daten-Sets, die von der Häufigkeit übereinstimmender Schlüssel abhängt. Im schlechtesten Fall, bei dem die Ausgabe das kartesische Produkt der Eingaben ist, kann sich die Laufzeit dem Wert NM nähern. Dieser Fall kommt nur selten vor. Die meisten Joins weisen deutlich weniger übereinstimmende Schlüssel auf. Wenn ein Daten-Set größer ist als die anderen oder wenn die eingehenden Daten bereits durch ein Schlüsselfeld sortiert wurden, können Sie mithilfe der Registerkarte “Optimierung” die Leistung dieses Knotens verbessern. Für weitere Informationen siehe Thema [Optimierungseinstellungen für das Zusammenführen in Kapitel 3 in IBM SPSS Modeler 15 Quellen-, Prozess- und Ausgabeknoten](#).

Aggregieren. Wenn die Option *Schlüssel sind zusammenhängend* nicht ausgewählt ist, liest dieser Knoten das gesamte Eingabe-Daten-Set ein (speichert es jedoch nicht), bevor eine aggregierte Ausgabe erstellt wird. In extremeren Situationen, wenn die Größe der aggregierten Daten einen Grenzwert erreicht (von der Konfigurationsoption *Multiplikator für Arbeitsspeicherverwendung* von SPSS Modeler Server festgelegt), wird der Rest des Daten-Sets so sortiert und verarbeitet, als wäre die Option *Schlüssel sind zusammenhängend* ausgewählt worden. Wenn diese Option nicht ausgewählt wurde, werden keine Daten gespeichert, weil die aggregierten Ausgabedatensätze beim Lesen der Eingabedaten produziert werden.

Duplikat. Der Duplikatknoten speichert alle eindeutigen Schlüsselfelder im Eingabe-Daten-Set. In Fällen, in denen alle Felder Schlüsselfelder und alle Datensätze eindeutig sind, wird das gesamte Daten-Set gespeichert. Standardmäßig sortiert der Duplikatknoten die Daten in den Schlüsselfeldern und wählt (oder verwirft) dann den ersten distinkten Datensatz aus jeder Gruppe. Für kleinere Daten-Sets mit einer geringen Anzahl distinkter Schlüssel oder für diejenigen, die vorsortiert wurden, können Sie Optionen wählen, um Geschwindigkeit und Effizienz der Verarbeitung zu steigern.

Typ. In einigen Fällen speichert der Typknoten beim Lesen der Werte die Eingabedaten im Cache; der Cache wird für die nachfolgende Verarbeitung verwendet. Für den Cache ist soviel Datenträgerspeicher erforderlich, um das gesamte Daten-Set zu speichern. Dafür wird jedoch die Verarbeitung beschleunigt.

Evaluierung. Der Evaluierungsknoten muss die Eingabedaten sortieren, um n-tile zu berechnen. Der Sortiervorgang wird für jedes evaluierte Modell wiederholt, da die Scores und folglich die Datensatzreihenfolge in jedem Fall unterschiedlich sind. Die Laufzeit beträgt $M*N*\log(N)$. Dabei ist M die Anzahl der Modelle und N die Anzahl der Datensätze.

Leistung: Modellierungsknoten

Netzwerk und Kohonen. Trainingsalgorithmen für neuronale Netzwerke (einschließlich des Kohonen-Algorithmus) führen zahlreiche Durchläufe über die Trainingsdaten durch. Die Daten werden bis zu einer gewissen Obergrenze im Arbeitsspeicher gespeichert, der Rest wird auf den Datenträger geschrieben. Der Zugriff auf die Trainingsdaten über den Datenträger ist aufwendig, da es sich um eine Zufallszugriffsmethode handelt, die zu übermäßiger Datenträgeraktivität führen kann. Sie können die Verwendung des Datenträgerspeichers für diese Algorithmen deaktivieren, sodass alle Daten im Arbeitsspeicher gespeichert werden müssen. Wählen Sie dazu im Dialogfeld des Knotens auf der Registerkarte "Modell" die Option Auf Geschwindigkeit optimieren aus. Beachten Sie, wenn der zum Speichern der Daten erforderliche Arbeitsspeicher größer ist als die Arbeitsmenge des Serverprozesses, wird ein Teil davon auf dem Datenträger ausgelastet und die Leistung wird entsprechend beeinträchtigt.

Bei Auswahl von Speicher optimiert wird dem Algorithmus ein Prozentsatz des physikalischen RAM zugeordnet. Dies geschieht gemäß dem Wert in der Konfigurationsoption *Obergrenze für Modellierungsarbeitsspeicher (in Prozent)* von IBM® SPSS® Modeler Server. Um mehr Arbeitsspeicher für das Trainieren neuronaler Netzwerke zu verwenden, stellen Sie entweder mehr RAM zur Verfügung oder erhöhen Sie den Wert für diese Option. Beachten Sie jedoch, dass ein zu hoher Wert zu Paging führt.

Die Laufzeit des Netzwerkalgorithmus hängt von der gewünschten Genauigkeitsstufe ab. Sie können die Laufzeit durch Festlegung einer Stoppbedingung im Dialogfeld des Knotens steuern.

K-Means. Der K-Means-Cluster-Algorithmus weist dieselben Optionen für die Steuerung der Arbeitsspeicherverwendung auf wie die Netzwerkalgorithmen. Die Leistung für auf Datenträger gespeicherte Daten ist jedoch besser, weil der Zugriff auf die Daten sequenziell erfolgt.

Leistung: CLEM-Ausdrücke

CLEM-Sequenzfunktionen ("@-Funktionen"), die einen Rückblick auf den Daten-Stream ermöglichen, müssen so viele Daten speichern, wie für den längsten Rückblick erforderlich sind. Bei Operationen mit unbegrenztem Rückblicksgrad müssen alle Werte im Feld gespeichert werden. Eine unbegrenzte Operation ist eine Operation, bei der es sich beim Offset-Wert nicht um ein ganzzahliges Literal handelt. z. B. @OFFSET(Sales, Month). Der Offset-Wert ist der Feldname *Monat*, dessen Wert bis zur Ausführung unbekannt ist. Der Server muss alle Werte des Felds *Verkäufe* speichern, um genaue Ergebnisse zu gewährleisten. Wenn eine Obergrenze bekannt ist, sollten Sie sie als zusätzliches Argument angeben; Beispiel: @OFFSET(Sales, Month, 12). Diese

Operation weist den Server an, nicht mehr als die 12 aktuellsten Werte von *Verkäufe* zu speichern. Sequenzfunktionen, ob mit Grenzen oder ohne, behindern fast immer die SQL-Generierung.

Eingabehilfen in IBM SPSS Modeler

Übersicht über die Eingabehilfen in IBM SPSS Modeler

Diese Version bietet erweiterte Eingabehilfen für alle Benutzer sowie besondere Unterstützung für Benutzer mit eingeschränkter Sehfähigkeit und anderen Körperbehinderungen. In diesem Abschnitt werden die Funktionen und Methoden für die Arbeit mit erweiterten Eingabehilfen, wie Bildschirm-Lesesystem und Direktzugriffstasten, beschrieben.

Arten der Unterstützung von Eingabehilfen

Ob Sie nun unter einer eingeschränkten Sehfähigkeit leiden oder aus anderen Gründen auf die Tastatur angewiesen sind, es gibt eine Vielzahl alternativer Methoden zur Verwendung dieses Data-Mining-Toolkits. Sie können z. B. Streams erstellen, Optionen angeben und die Ausgabe lesen, ohne die Maus zu verwenden. Die verfügbaren Direktzugriffstasten werden in den folgenden Themen beschrieben. Außerdem bietet IBM® SPSS® Modeler umfangreiche Unterstützung für Bildschirm-Lesesysteme, wie JAWS für Windows. Sie können auch das Farbschema ändern, um für zusätzlichen Kontrast zu sorgen. Die verschiedenen Möglichkeiten zur Unterstützung werden in den folgenden Themen beschrieben.

Zugriffsmöglichkeiten für Sehbehinderte

In IBM® SPSS® Modeler gibt es viele Eigenschaften, die Sie individuell festlegen können, um besser mit der Software arbeiten zu können.

Anzeigeoptionen

Sie können zum Anzeigen von Diagrammen verschiedene Farben auswählen. Sie können für die Software auch Ihre individuellen Windows-Einstellungen übernehmen. So wird der visuelle Kontrast unter Umständen erhöht.

- ▶ Um die Anzeigeoptionen einzustellen, klicken Sie im Menü “Extras” auf die Option Benutzeroptionen.
- ▶ Klicken Sie auf die Registerkarte Anzeige. Die Optionen auf dieser Registerkarte beziehen sich auf das Farbschema der Software, die Farben für Diagramme sowie die Schriftgrößen für Knoten.

Verwenden von Tönen zur Benachrichtigung

Durch das Aktivieren bzw. Deaktivieren von Tönen können Sie steuern, wie Ihnen bestimmte Operationen in der Software gemeldet werden. Zum Beispiel können Sie Töne für Ereignisse

aktivieren, wie die Erstellung von Knoten und die Löschung bzw. Generierung neuer Ausgabe bzw. Modelle.

- ▶ Um die Benachrichtigungsoptionen einzustellen, klicken Sie im Menü “Extras” auf die Option Benutzeroptionen.
- ▶ Klicken Sie auf die Registerkarte Benachrichtigungen.

Steuern des automatischen Aufrufs neuer Fenster

Die Registerkarte “Benachrichtigungen” im Dialogfeld “Benutzeroptionen” wird auch dazu verwendet, zu steuern, ob die neu generierte Ausgabe, wie Tabellen und Diagramme, in einem separaten Fenster aufgerufen wird. Unter Umständen ist es für Sie einfacher, diese Option zu deaktivieren und nur im Bedarfsfall ein Ausgabefenster zu öffnen.

- ▶ Um diese Optionen einzustellen, klicken Sie im Menü “Extras” auf die Option Benutzeroptionen.
- ▶ Klicken Sie auf die Registerkarte Benachrichtigungen.
- ▶ Wählen Sie im Dialogfeld in der Liste in der Gruppe Visuelle Benachrichtigungen die Option Neue Ausgabe aus.
- ▶ Wählen Sie unter Fenster öffnen die Option Nie.

Knotengröße

Knoten können entweder in der Standard- oder in einer kleineren Größe angezeigt werden. Sie können die verschiedenen Größeneinstellungen an Ihre Anforderungen anpassen.

- ▶ Um die Größenoptionen für die Knoten festzulegen, klicken Sie im Menü “Datei” auf die Option Stream-Eigenschaften.
- ▶ Klicken Sie auf die Registerkarte Layout.
- ▶ Wählen Sie in der Symbolgröße die Option Standard aus.

Zugriffsmöglichkeiten für Blinde

Die Unterstützung für blinde Benutzer ist nur möglich, wenn ein Bildschirm-Lesesystem verwendet wird wie JAWS für Windows. Um die Verwendung eines Bildschirm-Lesesystems in IBM® SPSS® Modeler zu optimieren, können Sie verschiedene Einstellungen festlegen.

Anzeigeoptionen

Bildschirm-Lesesysteme funktionieren in der Regel besser, wenn der visuelle Bildschirmkontrast höher ist. Wenn bereits eine Windows-Einstellung mit hohem Kontrast gewählt ist, können Sie diese Windows-Einstellungen für die Software verwenden.

- ▶ Um die Anzeigeoptionen einzustellen, klicken Sie im Menü “Extras” auf die Option Benutzeroptionen.
- ▶ Klicken Sie auf die Registerkarte Anzeige.

Verwenden von Tönen zur Benachrichtigung

Durch das Aktivieren bzw. Deaktivieren von Tönen können Sie steuern, wie Ihnen bestimmte Operationen in der Software gemeldet werden. Zum Beispiel können Sie Töne für Ereignisse aktivieren, wie die Erstellung von Knoten und die Löschung bzw. Generierung neuer Ausgabe bzw. Modelle.

- ▶ Um die Benachrichtigungsoptionen einzustellen, klicken Sie im Menü “Extras” auf die Option Benutzeroptionen.
- ▶ Klicken Sie auf die Registerkarte Benachrichtigungen.

Steuern des automatischen Aufrufs neuer Fenster

Die Registerkarte “Benachrichtigungen” im Dialogfeld “Benutzeroptionen” wird auch dazu verwendet, zu steuern, ob die neu generierte Ausgabe in einem separaten Fenster aufgerufen wird. Unter Umständen ist es für Sie einfacher, diese Option zu deaktivieren und nur im Bedarfsfall ein Ausgabefenster zu öffnen.

- ▶ Um diese Optionen einzustellen, klicken Sie im Menü “Extras” auf die Option Benutzeroptionen.
- ▶ Klicken Sie auf die Registerkarte Benachrichtigungen.
- ▶ Wählen Sie im Dialogfeld in der Liste in der Gruppe Visuelle Benachrichtigungen die Option Neue Ausgabe aus.
- ▶ Wählen Sie unter Fenster öffnen die Option Nie.

Tastatureingabehilfen

Die Produktfunktionen können über die Tastatur aufgerufen werden. So können Sie ALT und die entsprechende Taste drücken, um Fenstermenüs zu aktivieren (z. B. Alt+D, um das Menü “Datei” aufzurufen), oder die Tabulatortaste drücken, um durch die Dialogfeldsteuerelemente zu blättern. Darüber hinaus gibt es spezielle Zugriffsmöglichkeiten, die nur für die einzelnen Hauptfenster des Produkts gelten, und hilfreiche Hinweise zum Navigieren in Dialogfeldern.

In diesem Abschnitt werden die wichtigsten Tastatureingabehilfen beschrieben, vom Öffnen eines Streams über die Verwendung von Knotendialogfeldern bis hin zur Arbeit mit Ausgaben. Für eine noch wirksamere Navigation stehen Listen mit Direktzugriffstasten zur Verfügung.

Direktzugriffstasten für die Navigation im Hauptfenster

Die meiste Data-Mining-Arbeit findet im Hauptfenster von IBM® SPSS® Modeler statt. Der Hauptbereich ist der **Stream-Zeichenbereich**, der zum Erstellen und Ausführen von Daten-Streams verwendet wird. Der untere Teil des Fensters enthält die **Knotenpaletten** mit allen verfügbaren Knoten. Die Paletten werden je nach Typ der Data-Mining-Operation für jede Knotengruppe auf Registerkarten angeordnet. Knoten, die zum Importieren von Daten in SPSS Modeler verwendet werden, sind auf der Registerkarte “Datenquellen” gruppiert und Knoten, die zum Ableiten, Filtern und Eingeben von Feldern verwendet werden, sind auf der Registerkarte “Feldoperationen” gruppiert.

Die rechte Seite des Fensters enthält mehrere Tools zur Verwaltung von Streams, Ausgaben und Projekten. Die obere Hälfte auf der rechten Seite enthält die **Manager** mit drei Registerkarten, die zum Verwalten von Streams, Ausgaben und generierten Modellen verwendet werden. Sie können diese Objekte auswählen, indem Sie die Registerkarte und ein Objekt aus der Liste auswählen. Die untere Hälfte auf der rechten Seite enthält den **Projektbereich**, mit dem Sie Ihre Arbeit in Projekte aufteilen können. In diesem Bereich gibt es zwei Registerkarten, mit denen zwei unterschiedliche Ansichten eines Projekts möglich sind. Die Ansicht **Klassen** sortiert Projektobjekte nach Typ, während die Ansicht **CRISP-DM** Objekte nach der entsprechenden Data-Mining-Phase sortiert, z. B. Datenvorbereitung oder Modellierung. Diese verschiedenen Aspekte des SPSS Modeler-Fensters werden im gesamten Hilfesystem sowie im Benutzerhandbuch erläutert.

Im Folgenden erhalten Sie eine Tabelle mit Direktzugriffstasten, mit denen Sie im SPSS Modeler-Hauptfenster navigieren und Streams erstellen können. Die Direktzugriffstasten für Dialogfelder und die Ausgabe werden in den folgenden Themen aufgelistet. Beachten Sie, dass diese Tastenkürzel nur im Hauptfenster zur Verfügung stehen.

Direktzugriffstasten im Hauptfenster

Tastenkürzel	Funktion
Ctrl+F5	Verschiebt den Fokus zu den Knotenpaletten.
Strg+F6	Verschiebt den Fokus zum Stream-Zeichenbereich.
Strg+F7	Verschiebt den Fokus zum Manager-Bereich.
Strg+F8	Verschiebt den Fokus zum Projektbereich.

Knoten- und Stream-Direktzugriffstasten

Tastenkürzel	Funktion
Strg+N	Erstellt einen neuen, leeren Stream-Zeichenbereich.
Strg+O	Zeigt das Dialogfeld "Öffnen" an, über das Sie bereits vorhandene Streams auswählen und öffnen können.
Strg+Zifferntasten	Verschiebt den Fokus in einem Fenster zur entsprechenden Registerkarte. In einem Fensterbereich oder in einem Fenster mit Registerkarten springt Strg+1 zur ersten Registerkarte von links, Strg+2 zur zweiten usw.
Strg+Pfeil nach unten	Wird in der Knotenpalette zum Verschieben des Fokus von einer Palettenregisterkarte zum ersten Knoten auf dieser Registerkarte verwendet.
Strg+Pfeil nach oben	Wird in der Knotenpalette zum Verschieben des Fokus von einem Knoten zu dessen Palettenregisterkarte verwendet.
Enter	Wenn in der Knotenpalette ein Knoten ausgewählt ist (einschließlich verfeinerter Modelle in der Palette der generierten Modelle), wird der Knoten durch Drücken dieser Taste in den Stream-Zeichenbereich eingefügt. Durch Drücken der Eingabetaste wird bei bereits ausgewähltem Knoten im Zeichenbereich das Dialogfeld für diesen Knoten geöffnet.
Strg+Eingabetaste	Wenn in der Palette ein Knoten ausgewählt ist, wird dieser Knoten durch diese Tastenkombination dem Stream-Zeichenbereich hinzugefügt, ohne dass dieser ausgewählt wird. Gleichzeitig wird der Fokus auf den ersten Knoten in der Palette verschoben.

Tastenkürzel	Funktion
ALT+Eingabetaste	Wenn in der Palette ein Knoten ausgewählt ist, wird dieser Knoten durch diese Tastenkombination dem Stream-Zeichenbereich hinzugefügt und ausgewählt. Gleichzeitig wird der Fokus auf den ersten Knoten in der Palette verschoben.
Umschalt+Leertaste	Wenn ein Knoten in der Palette einen Fokus aufweist, wird hiermit der Knoten oder Kommentar ausgewählt bzw. seine Auswahl aufgehoben. Wenn auch andere Knoten oder Kommentare ausgewählt sind, wird deren Auswahl ebenfalls aufgehoben.
Strg+Umschalt+Leertaste	Wenn ein Knoten oder Kommentar im Stream oder ein Knoten oder Kommentar in der Palette einen Fokus aufweist, wird hiermit der Knoten oder Kommentar ausgewählt bzw. seine Auswahl aufgehoben. Dies beeinflusst keine anderen ausgewählten Knoten oder Kommentare.
Pfeil nach links/rechts	Verschiebt den gesamten Stream horizontal, wenn der Fokus auf dem Stream-Zeichenbereich ist. Wenn der Fokus auf einer Palettenregisterkarte liegt, wird zwischen den Registerkarten gewechselt. Wenn der Fokus auf einem Palettenknoten liegt, wird zwischen den Knoten in der Palette gewechselt.
Pfeil nach oben/unten	Verschiebt den gesamten Stream vertikal, wenn der Fokus auf dem Stream-Zeichenbereich ist. Wenn der Fokus auf einem Palettenknoten liegt, wird zwischen den Knoten in der Palette gewechselt. Wenn der Fokus auf einer Unterpalette liegt, wird zwischen anderen Unterpaletten dieser Palettenregisterkarte gewechselt.
Alt+Pfeil nach links/rechts	Verschiebt ausgewählte Knoten und Kommentare im Stream-Zeichenbereich horizontal in Richtung der Pfeiltaste.
Alt+Pfeil nach oben/unten	Verschiebt ausgewählte Knoten und Kommentare im Stream-Zeichenbereich vertikal in Richtung der Pfeiltaste.
Strg+A	Wählt alle Knoten in einem Stream aus.
Strg+Q	Mit dieser Tastenkombination werden der Knoten, auf dem der Fokus ist, sowie alle abwärts liegenden Knoten ausgewählt. Die Auswahl aller aufwärts liegenden Knoten wird aufgehoben.
Strg+W	Mit dieser Tastenkombination wird bei einem ausgewählten Knoten, auf dem der Fokus ist, sowie allen abwärts liegenden Knoten die Auswahl aufgehoben.
Strg+Alt+D	Mit dieser Tastenkombination wird ein ausgewählter Knoten dupliziert.
Strg+Alt+L	Wenn im Stream ein Modell-Nugget ausgewählt ist, wird ein Dialogfeld "Einfügen" geöffnet, über das Sie ein gespeichertes Modell aus einer .nod-Datei in den Stream laden können.
Strg+Alt+R	Zeigt für einen ausgewählten Knoten die Registerkarte "Anmerkungen" an, über die Sie den Knoten umbenennen können.
Strg+Alt+U	Erstellt einen Benutzereingabe-Quellenknoten.
Strg+Alt+C	Mit dieser Tastenkombination wird der Cache für einen Knoten aktiviert/deaktiviert.
Strg+Alt+F	Leert den Cache für einen Knoten.
Tabulator	Auf dem Stream-Zeichenbereich wird zwischen allen Quellenknoten und Kommentaren im aktuellen Stream gewechselt. In einer Knotenpalette wird zwischen den Knoten in der Palette gewechselt. In einer ausgewählten Unterpalette wird zum ersten Knoten in der Unterpalette gewechselt.

Tastenkürzel	Funktion
Umschalt+Tabulatortaste	Führt den gleichen Befehl wie die Tabulatortaste aus, jedoch in umgekehrter Reihenfolge.
Strg+Tabulatortaste	Wenn sich der Fokus im Manager-Bereich oder im Projektbereich befindet, wird der Fokus in den Stream-Zeichenbereich verschoben. Wenn sich der Fokus in einer Knotenpalette befindet, wird der Fokus zwischen einem Knoten und seiner Palettenregisterkarte gewechselt.
Beliebige Buchstabentaste	Wenn sich der Fokus auf einem Knoten im aktuellen Stream befindet, wird der Fokus auf den nächsten Knoten gesetzt, dessen Name mit der gedrückten Taste beginnt, und springt dorthin.
F1	Startet das Hilfesystem bei einem für den Fokus relevanten Thema.
F2	Startet den Verbindungsprozess für einen im Zeichenbereich ausgewählten Knoten. Verwenden Sie die Tabulatortaste, um zum gewünschten Knoten im Zeichenbereich zu gelangen, und drücken Sie Umschalt+Leertaste, um die Verbindung zu beenden.
F3	Löscht alle Verbindungen für den ausgewählten Knoten im Zeichenbereich.
F6	Verschiebt den Fokus zwischen dem Manager-Bereich, dem Projektbereich und der Knotenpalette.
F10	Öffnet das Menü "Datei".
Umschalt+F10	Öffnet das Popup-Menü für den Knoten oder Stream.
Delete	Löscht einen ausgewählten Knoten aus dem Zeichenbereich.
ESC	Schließt ein Popup-Menü oder Dialogfeld.
Strg+Alt+X	Mit dieser Tastenkombination wird der Superknoten erweitert.
Strg+Alt+Z	Vergrößert einen Superknoten.
Strg+Alt+Umschalt+Z	Verkleinert einen Superknoten.
Strg+E	Wenn sich der Fokus im Stream-Zeichenbereich befindet, wird mit dieser Tastenkombination der aktuelle Stream ausgeführt.

In SPSS Modeler werden auch viele der Standard-Direktzugriffstasten verwendet, wie Strg+C zum Kopieren. Für weitere Informationen siehe Thema [Verwenden von Direktzugriffstasten in Kapitel 3 auf S. 31](#).

Direktzugriffstasten für Dialogfelder und Tabellen

Einige der Direktzugriffstasten für die Tastatur und das Bildschirm-Lesesystem sind hilfreich, wenn Sie mit Dialogfeldern, Tabellen und Tabellen in Dialogfeldern arbeiten. Nachfolgend erhalten Sie eine vollständige Liste mit speziellen Direktzugriffstasten für die Tastatur und das Bildschirm-Lesesystem.

Dialogfeld- und Expression Builder-Direktzugriffstasten

Tastenkürzel	Funktion
Alt+4	Wird zum Verwerfen aller geöffneten Dialogfelder oder Ausgabefenster verwendet. Die Ausgabe kann auf der Registerkarte "Ausgaben" im Manager-Bereich abgerufen werden.
Strg+Ende	Wenn sich der Fokus auf einem beliebigen Steuerelement im Expression Builder befindet, wird die Einfügemarke mit dieser Tastenkombination an das Ende des Ausdrucks verschoben.

Tastenkürzel	Funktion
Strg+1	Im Expression Builder wird der Fokus mit dieser Tastenkombination zum Steuerelement für die Bearbeitung des Ausdrucks verschoben.
Strg+2	Im Expression Builder wird der Fokus mit dieser Tastenkombination zur Funktionsliste verschoben.
Strg+3	Mit dieser Tastenkombination wird der Fokus im Expression Builder zur Feldliste verschoben.

Direktzugriffstasten für Tabellen

Tabellen-Direktzugriffstasten werden für Ausgabetabellen sowie für Tabellensteuerelemente in Dialogfeldern für Knoten, wie Typ, Filter und Verbinden, verwendet. In der Regel verwenden Sie die Tabulatortaste, um zwischen Tabellenzellen zu navigieren, und Strg+Tab, um die Tabellensteuerung zu verlassen. *Hinweis:* Gelegentlich kann es vorkommen, dass ein Bildschirm-Lesesystem nicht sofort mit dem Lesen des Inhalts einer Zelle beginnt. Wenn Sie die Pfeiltasten ein- oder zweimal drücken, wird die Software zurückgesetzt und die Sprachausgabe beginnt.

Tastenkürzel	Funktion
Strg+W	Mit dieser Tastenkombination wird die Kurzbeschreibung der ausgewählten Zeile in Tabellen gelesen. Zum Beispiel: "Die ausgewählten Werte aus Zeile 2 sind Geschlecht, Flag, m/w usw."
Strg+Alt+W	Mit dieser Tastenkombination wird die lange Beschreibung der ausgewählten Zeile in Tabellen gelesen. Zum Beispiel: "Die ausgewählten Werte aus Zeile 2 sind Feld = Geschlecht, Typ = Flag, Geschlecht = m/w usw."
Strg+D	Mit dieser Tastenkombination wird die Kurzbeschreibung des ausgewählten Bereichs in Tabellen gelesen. Zum Beispiel: "Auswahl ist eine Zeile mal sechs Spalten."
Strg+Alt+D	Mit dieser Tastenkombination wird die lange Beschreibung des ausgewählten Bereichs in Tabellen gelesen. Zum Beispiel: "Auswahl ist eine Zeile mal sechs Spalten. Die ausgewählten Spalten sind Feld, Typ, Fehlend. Die ausgewählte Zeile ist 1."
Strg+T	Diese Tastenkombination liefert in Tabellen eine Kurzbeschreibung der ausgewählten Spalten. Zum Beispiel: "Felder, Typ, Fehlend."
Strg+Alt+T	Diese Tastenkombination liefert in Tabellen eine lange Beschreibung der ausgewählten Spalten. Zum Beispiel: "Ausgewählte Spalten sind Felder, Typ, Fehlend."
Strg+R	Mit dieser Tastenkombination wird in Tabellen die Anzahl der in der Tabelle vorhandenen Datensätze angezeigt.
Strg+Alt+R	Mit dieser Tastenkombination wird in Tabellen die Anzahl der in der Tabelle vorhandenen Datensätze sowie Spaltennamen angezeigt.
Strg+I	Mit dieser Tastenkombination werden in Tabellen die ZellenInformation bzw. der Inhalt für die Zelle, auf der der Fokus ist, gelesen.
Strg+Alt+I	Mit dieser Tastenkombination wird in Tabellen die lange Beschreibung der ZellenInformation (Spaltenname und Inhalt der Zelle) für die Zelle, auf der der Fokus ist, gelesen.

Tastenkürzel	Funktion
Strg+G	Diese Tastenkombination liefert für Tabellen kurze Allgemeine Auswahlinformationen.
Strg+Alt+G	Diese Tastenkombination liefert für Tabellen lange Allgemeine Auswahlinformationen.
Strg+Q	Ermöglicht in Tabellen ein schnelles Wechseln zwischen den Tabellenzellen. Mit Strg+Q werden lange Beschreibungen, wie "Geschlecht=Weiblich", gelesen, wenn Sie mit den Pfeiltasten durch die Tabelle blättern. Wenn Sie Strg+Q erneut auswählen, werden die kurzen Beschreibungen (Zelleninhalte) angezeigt.

Direktzugriffstasten für Kommentare

Beim Arbeiten mit Kommentaren am Bildschirm können Sie die folgenden Direktzugriffstasten verwenden.

Tastenkürzel	Funktion
Alt+C	Aktiviert/deaktiviert die Funktion "Kommentare ein-/ausblenden".
Alt+M	Fügt einen neuen Kommentar ein, falls Kommentare angezeigt werden, und zeigt Kommentare an, wenn Sie gerade ausgeblendet sind.
Tabulator	Auf dem Stream-Zeichenbereich wird zwischen allen Quellenknoten und Kommentaren im aktuellen Stream gewechselt.
Enter	Wenn ein Kommentar den Fokus aufweist, gibt dies den Beginn der Bearbeitung an.
Alt+Eingabetaste oder Strg+Tabulatortaste	Beendet die Bearbeitung und speichert vorgenommene Änderungen
ESC	Bricht die Bearbeitung ab. Vorgenommene Änderungen gehen verloren.
Alt+Umschalt+Pfeil nach oben	Verringert die Höhe des Textbereichs um eine Rasterzelle (oder ein Pixel), wenn die Funktion zum Ausrichten am Gitter aktiviert (bzw. deaktiviert) ist.
Alt+Umschalt+Pfeil nach unten	Vergrößert die Höhe des Textbereichs um eine Rasterzelle (oder ein Pixel), wenn die Funktion zum Ausrichten am Gitter aktiviert (bzw. deaktiviert) ist.
Alt+Umschalt+Pfeil nach links	Verringert die Breite des Textbereichs um eine Rasterzelle (oder ein Pixel), wenn die Funktion zum Ausrichten am Gitter aktiviert (bzw. deaktiviert) ist.
Alt+Umschalt+Pfeil nach rechts	Vergrößert die Breite des Textbereichs um eine Rasterzelle (oder ein Pixel), wenn die Funktion zum Ausrichten am Gitter aktiviert (bzw. deaktiviert) ist.

Direktzugriffstasten für Cluster-Viewer und Modell-Viewer

Für die Navigation in den Fenstern des Cluster-Viewers und des Modell-Viewers stehen Direktzugriffstasten zur Verfügung.

Allgemein - Cluster-Viewer und Modell-Viewer

Tastenkürzel	Funktion
Tabulator	Verschiebt den Fokus zum nächsten Bildschirm-Steuerelement.
Umschalt+Tabulatortaste	Verschiebt den Fokus zum vorherigen Bildschirm-Steuerelement.
Pfeil nach unten	Wenn eine Dropdown-Liste den Fokus enthält, wird die Liste geöffnet bzw. erhält das nächste Element in der Liste den Fokus. Wenn ein Menü den Fokus enthält, wird der Fokus zum nächsten Element im Menü verschoben. Wenn sich der Fokus auf einem Miniaturdiagramm befindet, wird der Fokus zum nächsten Element im Set (oder zum ersten, wenn der Fokus auf dem letzten Miniaturdiagramm liegt) verschoben.
Pfeil nach oben	Wenn eine Dropdown-Liste den Fokus enthält, wird der Fokus zum vorherigen Element in der Liste verschoben. Wenn ein Menü den Fokus enthält, wird der Fokus zum vorherigen Element im Menü verschoben. Wenn sich der Fokus auf einem Miniaturdiagramm befindet, wird der Fokus zum vorherigen Element im Set (oder zum letzten, wenn der Fokus auf dem ersten Miniaturdiagramm liegt) verschoben.
Enter	Schließt eine geöffnete Dropdown-Liste oder nimmt eine Auswahl in einem geöffneten Menü vor.
F6	Schaltet den Fokus zwischen dem linken und rechten Bereich des Fensters um.
Pfeil nach links/rechts	Wenn eine Registerkarte den Fokus enthält, wird der Fokus zur vorherigen bzw. nächsten Registerkarte verschoben. Wenn ein Menü den Fokus enthält, wird der Fokus zum vorherigen bzw. nächsten Menü verschoben.
Alt+Buchstabe	Wählt die Schaltfläche bzw. das Menü aus, in dessen Namen dieser Buchstabe unterstrichen ist.
ESC	Schließt ein geöffnetes Menü oder Dropdown-Listenfeld.

Nur Cluster-Viewer

Im Cluster-Viewer kann ein Cluster-nach-Funktionen-Raster angezeigt werden. Für weitere Informationen siehe Thema Clusteransicht in Kapitel 11 in *IBM SPSS Modeler 15 Modellierungsknoten*.

So wählen Sie die Cluster-Ansicht anstelle der Ansicht "Modellübersicht":

- ▶ Drücken Sie wiederholt die Tabulatortaste, bis die Schaltfläche Ansicht ausgewählt ist.
- ▶ Drücken Sie zweimal die Pfeil-nach-unten-Taste, um Cluster auszuwählen.

Von hier aus können Sie eine einzelne Zelle im Raster auswählen:

- ▶ Drücken Sie wiederholt die Tabulatortaste, bis Sie das letzte Symbol in der Visualisierungssymbolleiste erreichen.

Abbildung A-1
Symbol "Visualisierungsbaum anzeigen"



- ▶ Drücken Sie noch einmal die Tabulatortaste, anschließend die Leertaste und dann eine Pfeiltaste.

Nun stehen die folgenden Tastenkürzel zur Verfügung:

Tastenkürzel	Funktion
Pfeiltaste	Verschiebt den Fokus zwischen einzelnen Zellen im Raster. Die Anzeige Zellenverteilung im rechten Bereich ändert sich mit dem Verschieben des Fokus.
Strg+, (Komma)	Wählt die ganze Spalte im Raster aus, in der eine Zelle den Fokus enthält, bzw. hebt die Auswahl der ganzen Spalte auf. Sie fügen der Auswahl eine Spalte hinzu, indem Sie mithilfe der Pfeiltasten zu einer Zelle in der entsprechenden Spalte wechseln und erneut Strg+, drücken.
Tabulator	Verschiebt den Fokus aus dem Raster und zum nächsten Bildschirm-Steuerelement.
Umschalt+Tabulatortaste	Verschiebt den Fokus aus dem Raster und zum vorherigen Bildschirm-Steuerelement.
F2	Wechselt in den Bearbeitungsmodus (nur Beschriftungen und Beschreibungszellen).
Enter	Speichert Bearbeitungsänderungen und beendet den Bearbeitungsmodus (nur Beschriftungen und Beschreibungszellen).
ESC	Beendet den Bearbeitungsmodus, ohne Änderungen zu speichern (nur Beschriftungen und Beschreibungszellen).

Beispiel für Direktzugriffstasten: Erstellen von Streams

Um den Stream-Erstellungsprozess für Benutzer zu verdeutlichen, die auf die Tastatur oder ein Bildschirm-Lesesystem angewiesen sind, folgt nun ein Beispiel für eine Stream-Erstellung ohne Verwendung der Maus. In diesem Beispiel erstellen Sie einen Stream, der einen Knoten vom Typ “Datei (var.)”, einen Knoten vom Typ “Ableitung” und einen Knoten vom Typ “Histogramm” enthält. Gehen Sie dazu wie folgt vor:

- ▶ **Start SPSS Modeler.** Beim Start von IBM® SPSS® Modeler befindet sich der Fokus in der Registerkarte “Favoriten” der Knotenpalette.
- ▶ **Strg+Pfeil nach unten.** Verschiebt den Fokus von der Registerkarte selbst zum Inhalt der Registerkarte.
- ▶ **Pfeil nach rechts.** Verschiebt den Fokus zum Knoten “Variable Datei”.
- ▶ **Leertaste.** Wählt den Knoten “Variable Datei” aus.
- ▶ **Strg+Eingabetaste.** Nimmt den neuen Variablenknoten in den Stream-Zeichenbereich auf. Mit dieser Tastenkombination bleibt die Auswahl am Knoten “Variable Datei”, sodass der nächste hinzugefügte Knoten damit verbunden wird.
- ▶ **Tabulatortaste.** Verschiebt den Fokus zurück in die Knotenpalette.
- ▶ **Drücken Sie viermal auf die Pfeil-nach-rechts-Taste.** So gelangen Sie zum Ableitungsknoten.
- ▶ **Leertaste.** Mit dieser Taste wird der Ableitungsknoten ausgewählt.

- ▶ **ALT+Eingabetaste.** Fügt den Ableitungsknoten im Zeichenbereich ein und verschiebt die Auswahl zum Ableitungsknoten. Dieser Knoten kann nun mit dem nächsten hinzugefügten Knoten verbunden werden.
- ▶ **Tabulatortaste.** Verschiebt den Fokus zurück in die Knotenpalette.
- ▶ **Drücken Sie fünfmal auf die Pfeil-nach-rechts-Taste.** Verschiebt den Fokus zum Knoten “Histogramm” in der Palette.
- ▶ **Leertaste.** Wählt den Knoten “Histogramm” aus.
- ▶ **Eingabe.** Fügt den Knoten dem Stream hinzu und verschiebt den Fokus in den Stream-Zeichenbereich.

Fahren Sie mit dem nächsten Beispiel fort oder speichern Sie den Stream, falls Sie das nächste Beispiel zu einem späteren Zeitpunkt ausprobieren möchten.

Beispiel für Direktzugriffstasten: Bearbeiten von Knoten

In diesem Beispiel verwenden Sie den im vorherigen Beispiel erstellten Stream. Der Stream besteht aus einem Knoten vom Typ “Datei (var.)”, einem Ableitungsknoten und einem Knoten vom Typ “Histogramm”. Die Anweisungen beginnen mit dem Fokus auf dem dritten Knoten im Stream, dem Knoten vom Typ “Histogramm”.

- ▶ **Drücken Sie zweimal auf die Pfeil-nach-links-Taste.** Verschiebt den Fokus zurück zum Knoten “Variable Datei”.
- ▶ **Eingabe.** Ruft das Dialogfeld “Variable Datei” auf. Wechseln Sie mithilfe der Tabulatortaste in das Feld “Datei” und geben Sie einen Dateipfad und einen Dateinamen ein, um die entsprechende Datei auszuwählen. Wechseln Sie dann mithilfe der Tastenkombination Strg+Tab in den unteren Bereich des Dialogfelds, drücken Sie die Tabulatortaste, bis die Schaltfläche “OK” ausgewählt ist, und drücken Sie die Eingabetaste, um das Dialogfeld zu schließen.
- ▶ **Strg+Pfeil nach rechts:** Verschiebt den Fokus zum zweiten Knoten, einem Ableitungsknoten.
- ▶ **Eingabe.** Ruft das Dialogfeld “Ableitungsknoten” auf. Blättern Sie mit der Tabulatortaste durch dieses Dialogfeld, um die gewünschten Felder auszuwählen und die Ableitungsbedingungen festzulegen. Wechseln Sie dann mithilfe der Tastenkombination Strg+Tab zur Schaltfläche “OK” und drücken Sie die Eingabetaste, um das Dialogfeld zu schließen.
- ▶ **Strg+Pfeil nach rechts:** Verschiebt den Fokus zum dritten Knoten, einem Knoten vom Typ “Histogramm”.
- ▶ **Eingabe.** Ruft das Dialogfeld “Histogrammknoten” auf. Blättern Sie mit der Tabulatortaste durch dieses Dialogfeld, um die gewünschten Felder auszuwählen und Diagrammoptionen festzulegen. Für Dropdown-Listen drücken Sie die Pfeil-nach-unten-Taste, um die Liste zu öffnen und ein Element zu markieren, und drücken Sie dann die Eingabetaste, um das Listenelement auszuwählen. Wechseln Sie dann mithilfe der Tabulatortaste zur Schaltfläche “OK” und drücken Sie die Eingabetaste, um das Dialogfeld zu schließen.

Zu diesem Zeitpunkt können Sie weitere Knoten hinzufügen oder den aktuellen Stream ausführen. Beachten Sie beim Erstellen von Streams folgende Hinweise:

- Verwenden Sie beim manuellen Verbinden von Knoten die Taste F2, um den Startpunkt einer Verbindung zu erstellen. Navigieren Sie dann mithilfe der Tabulatortaste zum Endpunkt und drücken Sie Umschalt+Leertaste, um die Verbindung fertigzustellen.
- Verwenden Sie die Taste F3, um alle Verbindungen für einen ausgewählten Knoten im Zeichenbereich zu zerstören.
- Nachdem Sie einen Stream erstellt haben, drücken Sie Strg+E, um den aktuellen Stream auszuführen.

Eine vollständige Liste mit Direktzugriffstasten ist vorhanden. Für weitere Informationen siehe Thema [Direktzugriffstasten für die Navigation im Hauptfenster auf S. 266](#).

Verwenden eines Bildschirm-Lesesystems

Auf dem Markt sind verschiedene Bildschirm-Lesesysteme erhältlich. IBM® SPSS® Modeler ist für die Unterstützung von JAWS für Windows mit Java Access Bridge konfiguriert, das zusammen mit SPSS Modeler installiert wird. Wenn JAWS installiert ist, starten Sie JAWS einfach vor dem Start von SPSS Modeler, um das Produkt zu verwenden.

Aufgrund der Natur der einmaligen grafischen Darstellung des Data-Mining-Prozesses von SPSS Modeler werden die Diagramme und Grafiken optimal visuell verwendet. Es ist jedoch möglich, dass Sie Entscheidungen basierend auf der Ausgabe und der Modelle treffen, die unter Verwendung eines Bildschirm-Lesesystems textuell angezeigt werden.

Hinweis: Bei 64-Bit-Client-Computern funktionieren manche Hilfstechnologiefunktionen nicht. Das liegt daran, dass die Java Access Bridge nicht für den 64-Bit-Betrieb konfiguriert ist.

Verwenden der Wörterbuchdatei von IBM SPSS Modeler

Eine Wörterbuchdatei von SPSS Modeler (*Awt.JDF*) steht für die Verwendung mit JAWS zur Verfügung. So verwenden Sie diese Datei:

- ▶ Navigieren Sie zum Unterverzeichnis */accessibility* Ihrer SPSS Modeler-Installation und kopieren Sie die Wörterbuchdatei (*Awt.JDF*).
- ▶ Kopieren Sie die Datei in das Verzeichnis, in dem sich Ihre JAWS-Skripts befinden.

Unter Umständen befindet sich bereits eine Datei namens You *Awt.JDF* auf Ihrem Rechner, falls andere JAVA-Anwendungen ausgeführt werden. In diesem Fall können Sie diese Wörterbuchdatei eventuell nicht verwenden, ohne die Wörterbuchdatei manuell zu bearbeiten.

Verwenden eines Bildschirm-Lesesystems mit HTML-Ausgabe

Wenn Sie unter Verwendung eines Bildschirm-Lesesystems eine Ausgabe anzeigen, die in IBM® SPSS® Modeler als HTML angezeigt wird, können eventuell einige Schwierigkeiten auftreten. Es sind mehrere Ausgabetypen betroffen, darunter:

- Ausgabe, die auf der Registerkarte “Erweitert” für Knoten vom Typ “Regression”, “Logistische Regression” und “Faktor/PCA” angezeigt wird
- Ausgabe des Berichtsknotens

In jedem dieser Fenster oder Dialogfelder gibt es ein Tool in der Symbolleiste, das zum Aufrufen der Ausgabe in Ihrem Standardbrowser verwendet wird, der Unterstützung für ein Standard-Bildschirm-Lesesystem bietet. Sie können das Bildschirm-Lesesystem dann verwenden, um die Ausgabeinformationen zu übermitteln.

Eingabehilfen im Fenster “Interaktiver Baum”

Die Standardanzeige eines Entscheidungsbaummodells im Fenster “Interaktiver Baum” kann zu Problemen mit Bildschirm-Lesesystemen führen. Um eine Version mit Eingabehilfen aufzurufen, wählen Sie in den Menüs “Interaktiver Baum”:

Ansicht > Eingabehilfen

Die Anzeige ist ähnlich wie die Standardbaumübersicht, kann jedoch von JAWS ordnungsgemäß gelesen werden. Mit den Standardpfeiltasten können Sie sich nach oben, nach unten, nach rechts oder nach links bewegen. Bei der Navigation im Fenster mit Eingabehilfen ändert sich der Fokus im Fenster “Interaktiver Baum” entsprechend. Verwenden Sie die Leertaste, um die Auswahl zu ändern, oder Strg+Leertaste, um die aktuelle Auswahl zu erweitern.

Tipps zur Verwendung

Es gibt mehrere Tipps, wie die IBM® SPSS® Modeler-Umgebung für Sie benutzerfreundlicher gestaltet werden kann. Im Folgenden erhalten Sie allgemeine Tipps für die Arbeit in SPSS Modeler.

- **Verlassen von erweiterten Textfeldern.** Verwenden Sie Strg+Tab, um die erweiterten Textfelder zu verlassen. *Hinweis:* Strg+Tab wird auch zum Beenden von Tabellensteuerelementen verwendet.
- **Verwenden der Tabulatortaste anstelle der Pfeiltasten.** Verwenden Sie bei der Auswahl von Optionen für ein Dialogfeld die Tabulatortasten, um zwischen den Optionsschaltflächen hin- und herzuwechseln. Die Pfeiltasten funktionieren in diesem Zusammenhang nicht.
- **Dropdown-Listen.** In einer Dropdown-Liste für Dialogfelder können Sie entweder die Escape-Taste oder die Leertaste verwenden, um ein Element auszuwählen, und dann die Liste schließen. Sie können die Escape-Taste auch verwenden, um die Dropdown-Listen zu schließen, die nicht geschlossen werden können, wenn Sie mit der Tabulatortaste zu einem anderen Steuerelement wechseln.

- **Ausführungsstatus.** Wenn Sie einen Stream für eine große Datenbank ausführen, kann JAWS mit dem Lesen des Stream-Status hinterherhinken. Drücken Sie die Strg-Taste in regelmäßigen Abständen, um den Statusbericht zu aktualisieren.
- **Verwenden der Knotenpaletten.** Wenn Sie zum ersten Mal eine Registerkarte der Knotenpaletten aufrufen, liest JAWS manchmal "Gruppenfeld" anstatt den Namen des Knotens. In diesem Fall können Sie Strg+Pfeil nach rechts und dann Strg+Pfeil nach links verwenden, um das Bildschirm-Lesesystem zurückzusetzen und den Knotennamen zu hören.
- **Lesemenüs.** Beim erstmaligen Öffnen eines Menüs kann es gelegentlich vorkommen, dass JAWS den ersten Menüeintrag nicht lesen kann. Wenn Sie vermuten, dass dies der Fall gewesen sein könnte, verwenden Sie die Pfeil-nach-unten-Taste und dann die Pfeil-nach-oben-Taste, um das erste Element im Menü zu hören.
- **Überlappende Menüs.** JAWS liest die erste Ebene eines überlappenden Menüs nicht. Wenn Sie beim Sprechen eine Unterbrechung hören, während Sie durch das Menü blättern, drücken Sie die Pfeil-nach-rechts-Taste, um die untergeordneten Menüelemente zu hören.

Wenn auf Ihrem Rechner Text Analytics for IBM® SPSS® Modeler installiert ist, können Ihnen außerdem die folgenden Tipps die Arbeit mit der interaktiven Workbench-Oberfläche erleichtern.

- **Aufrufen von Dialogfeldern.** Möglicherweise müssen Sie die Tabulatortaste drücken, um beim Aufrufen eines Dialogfelds den Fokus auf das erste Steuerelement zu legen.
- **Verlassen von erweiterten Textfeldern.** Verwenden Sie Strg+Tab, um die erweiterten Textfelder zu verlassen und zum nächsten Steuerelement überzugehen. *Hinweis:* Strg+Tab wird auch zum Beenden von Tabellensteuerelementen verwendet.
- **Eingabe des ersten Buchstabens zum Auffinden von Elementen in der Baumliste.** Beim Suchen nach einem Element im Fensterrbereich für Kategorien, im Fensterrbereich für die extrahierten Ergebnisse oder im Bibliotheksbaum können Sie den ersten Buchstaben des Elements eingeben, wenn sich der Fokus im entsprechenden Fensterrbereich befindet. Dadurch wird das nächste Vorkommnis eines Elements ausgewählt, das mit dem eingegebenen Buchstaben beginnt.
- **Dropdown-Listen.** In einer Dropdown-Liste für Dialogfelder können Sie die Leertaste verwenden, um ein Element auszuwählen, und dann die Liste schließen.

In den folgenden Themen werden weitere zusätzliche Tips erörtert.

Kompatibilität mit anderer Software

Beim Testen von IBM® SPSS® Modeler mit Bildschirm-Lesesystemen, wie JAWS, hat unser Entwicklungsteam festgestellt, dass die Verwendung eines Systems Management Servers (SMS) in Ihrem Unternehmen eventuell mit der Funktion von JAWS kollidiert, Java-basierte Anwendungen, wie SPSS Modeler, zu lesen. Die Deaktivierung von SMS behebt diese Situation. Weitere Informationen zu SMS finden Sie auf der Microsoft-Website.

JAWS und Java

Die verschiedenen Versionen von JAWS bieten unterschiedliche Unterstützung für Java-basierte Software-Anwendungen. IBM® SPSS® Modeler funktioniert zwar mit allen neueren Versionen von JAWS, bei einigen Versionen können jedoch kleinere Probleme auftreten, wenn Sie zusammen mit Java-basierten Systemen eingesetzt werden. Besuchen Sie die Website von JAWS für Windows unter <http://www.FreedomScientific.com>.

Verwenden von Diagrammen in IBM SPSS Modeler

Die visuelle Anzeige von Informationen, wie Histogramme, Evaluierungsdiagramme, Multidiagramme und Streudiagramme, sind mit einem Bildschirm-Lesesystem schwer zu interpretieren. Beachten Sie jedoch, dass die Webdiagramme und Verteilungen mithilfe der Textübersicht im Ausgabefenster angezeigt werden können.

Unicode-Unterstützung

Unicode-Unterstützung in IBM SPSS Modeler

IBM® SPSS® Modeler 10.0 ist für IBM® SPSS® Modeler und IBM® SPSS® Modeler Server vollständig Unicode-kompatibel. Dadurch können Daten mit anderen Anwendungen ausgetauscht werden, die Unicode unterstützen, einschließlich mehrsprachiger Datenbanken, und das ohne jeden Informationsverlust, der durch die Konvertierung in ein oder aus einem Kodierungsschema verursacht werden könnte, das für eine bestimmte Ländereinstellung spezifisch ist.

- SPSS Modeler speichert Unicode-Daten intern und kann mehrsprachige Daten, die als Unicode in Datenbanken gespeichert sind, verlustfrei lesen und schreiben.
- SPSS Modeler kann UTF-8-kodierte Textdateien lesen und schreiben. Für den Import und Export von Textdateien wird standardmäßig die für die Ländereinstellung spezifische Kodierung verwendet, doch UTF-8 wird als Alternative unterstützt. Diese Einstellung kann in den Import- bzw. Exportknoten der Datei festgelegt werden. Alternativ kann die Standardkodierung im Dialogfeld “Stream-Eigenschaften” geändert werden. **Für weitere Informationen siehe Thema Festlegen von allgemeinen Optionen für Streams in Kapitel 5 auf S. 63.**
- Statistics-, SAS- und Texdatendateien, die in der für die Ländereinstellung spezifischen Kodierung gespeichert sind, werden beim Import in UTF-8 konvertiert und beim Export wieder in das ursprüngliche Format zurückkonvertiert. Wenn beim Schreiben in eine Datei Unicode-Zeichen vorhanden sind, die im Zeichensatz der Ländereinstellung nicht vorhanden sind, werden diese Zeichen ersetzt und eine Warnung wird angezeigt. Dies sollte nur passieren, wenn die Daten aus einer Datenquelle importiert wurden, die Unicode unterstützt (Datenbank oder UTF-8-Textdatei) und Zeichen aus einer anderen Ländereinstellung oder aus mehreren Ländereinstellungen oder Zeichensätzen enthält.
- Bilder aus IBM® SPSS® Modeler Solution Publisher sind UTF-8-kodiert und vollständig zwischen Plattformen und Ländereinstellungen übertragbar.

Informationen zu Unicode

Das Ziel des Unicode-Standards besteht darin, ein konsistentes Verfahren zur Kodierung von mehrsprachigem Text bereitzustellen, sodass dieser problemlos über Grenzen, Ländereinstellungen und Anwendungen hinweg verwendet werden kann. Der Unicode-Standard, der mittlerweile in Version 4.0.1 vorliegt, definiert einen Zeichensatz, der eine Obermege aller derzeit häufig verwendeten Zeichensätze weltweit darstellt, und weist jedem Zeichen einen eindeutigen Namen und Codepunkt zu. Die Zeichen und ihre Codepunkte stimmen mit denen des in ISO-10646 definierten Universal Character Set (“universeller Zeichensatz”, Abk. UCS) überein. Weitere Informationen finden Sie auf der [Unicode-Homepage \(*http://www.unicode.org*\)](http://www.unicode.org).

Hinweise

Diese Informationen wurden für weltweit angebotene Produkte und Dienstleistungen erarbeitet.

IBM bietet die in diesem Dokument behandelten Produkte, Dienstleistungen oder Merkmale möglicherweise nicht in anderen Ländern an. Informationen zu den derzeit in Ihrem Land erhältlichen Produkten und Dienstleistungen erhalten Sie bei Ihrem zuständigen IBM-Mitarbeiter vor Ort. Mit etwaigen Verweisen auf Produkte, Programme oder Dienste von IBM soll nicht behauptet oder impliziert werden, dass nur das betreffende Produkt oder Programm bzw. der betreffende Dienst von IBM verwendet werden kann. Stattdessen können alle funktional gleichwertigen Produkte, Programme oder Dienste verwendet werden, die keine geistigen Eigentumsrechte von IBM verletzen. Es obliegt jedoch der Verantwortung des Benutzers, die Funktionsweise von Produkten, Programmen oder Diensten von Drittanbietern zu bewerten und zu überprüfen.

IBM verfügt möglicherweise über Patente oder hat Patentanträge gestellt, die sich auf in diesem Dokument beschriebene Inhalte beziehen. Durch die Bereitstellung dieses Dokuments werden Ihnen keinerlei Lizizenzen an diesen Patenten gewährt. Lizenzanfragen können schriftlich an folgende Adresse gesendet werden:

*IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785,
U.S.A.*

Bei Lizenzanfragen in Bezug auf DBCS-Daten (Double-Byte Character Set) wenden Sie sich an die für geistiges Eigentum zuständige Abteilung von IBM in Ihrem Land. Schriftliche Anfragen können Sie auch an folgende Adresse senden:

*Intellectual Property Licensing, Legal and Intellectual Property Law, IBM Japan Ltd., 1623-14,
Shimotsuruma, Yamato-shi, Kanagawa 242-8502 Japan.*

Der folgende Abschnitt findet in Großbritannien und anderen Ländern keine Anwendung, in denen solche Bestimmungen nicht mit der örtlichen Gesetzgebung vereinbar sind: INTERNATIONAL BUSINESS MACHINES STELLT DIESE VERÖFFENTLICHUNG IN DER VERFÜGBAREN FORM OHNE GARANTIEN BEREIT, SEIEN ES AUSDRÜCKLICHE ODER STILLSCHWEIGENDE, EINSCHLIESSLICH JEDOCH NICHT NUR DER GARANTIEN BEZÜGLICH DER NICHT-RECHTSVERLETZUNG, DER GÜTE UND DER EIGNUNG FÜR EINEN BESTIMMTEN ZWECK. Manche Rechtsprechungen lassen den Ausschluss ausdrücklicher oder implizierter Garantien bei bestimmten Transaktionen nicht zu, sodass die oben genannte Ausschlussklausel möglicherweise nicht für Sie relevant ist.

Diese Informationen können technische Ungenauigkeiten oder typografische Fehler aufweisen. An den hierin enthaltenen Informationen werden regelmäßig Änderungen vorgenommen. Diese Änderungen werden in neuen Ausgaben der Veröffentlichung aufgenommen. IBM kann jederzeit und ohne vorherige Ankündigung Optimierungen und/oder Änderungen an den Produkten und/oder Programmen vornehmen, die in dieser Veröffentlichung beschrieben werden.

Jegliche Verweise auf Drittanbieter-Websites in dieser Information werden nur der Vollständigkeit halber bereitgestellt und dienen nicht als Befürwortung dieser. Das Material auf diesen Websites ist kein Bestandteil des Materials zu diesem IBM-Produkt und die Verwendung erfolgt auf eigene Gefahr.

IBM kann die von Ihnen angegebenen Informationen verwenden oder weitergeben, wie dies angemessen erscheint, ohne Ihnen gegenüber eine Verpflichtung einzugehen.

Lizenznahmer dieses Programms, die Informationen dazu benötigen, wie (i) der Austausch von Informationen zwischen unabhängig erstellten Programmen und anderen Programmen und (ii) die gegenseitige Verwendung dieser ausgetauschten Informationen ermöglicht wird, wenden sich an:

IBM Software Group, Attention: Licensing, 233 S. Wacker Dr., Chicago, IL 60606, USA.

Derartige Informationen stehen ggf. in Abhängigkeit von den jeweiligen Geschäftsbedingungen sowie in einigen Fällen der Zahlung einer Gebühr zur Verfügung.

Das in diesem Dokument beschriebene lizenzierte Programm und sämtliche dafür verfügbaren lizenzierten Materialien werden von IBM gemäß dem IBM-Kundenvertrag, den Internationalen Nutzungsbedingungen für Programmpakete der IBM oder einer anderen zwischen uns getroffenen Vereinbarung bereitgestellt.

Jegliche hier enthaltene Daten zur Leistung wurden in einer überwachten Umgebung ermittelt. Aus diesem Grund können in anderen Betriebsumgebungen gewonnene Ergebnisse stark davon abweichen. Einige Messungen wurden unter Umständen auf Systemen im Entwicklungsstadium durchgeführt und es kann nicht garantiert werden, dass diese Messungen auf allgemein verfügbaren Systemen zum gleichen Ergebnis führen. Darüber hinaus wurden einige Messungen unter Umständen durch Extrapolation bestimmt. Die tatsächlichen Ergebnisse können hiervon abweichen. Die Benutzer dieses Dokuments sollten die entsprechenden Daten für die jeweils vorliegende Umgebung prüfen.

Informationen zu Produkten von Drittanbietern wurden von den Anbietern des jeweiligen Produkts, aus deren veröffentlichten Ankündigungen oder anderen, öffentlich verfügbaren Quellen bezogen. IBM hat diese Produkte nicht getestet und kann die Genauigkeit bezüglich Leistung, Kompatibilität oder anderen Behauptungen nicht bestätigen, die sich auf Drittanbieter-Produkte beziehen. Fragen bezüglich der Funktionen von Drittanbieter-Produkten sollten an die Anbieter der jeweiligen Produkte gerichtet werden.

Alle Aussagen bezüglich der zukünftigen Ausrichtung von IBM oder der Absichten des Unternehmens können ohne vorherige Ankündigung geändert oder zurückgenommen werden und stellen lediglich Ziele und Vorgaben dar.

Diese Informationen enthalten Beispiele zu Daten und Berichten, die im täglichen Geschäftsbetrieb Verwendung finden. Um diese so vollständig wie möglich zu illustrieren, umfassen die Beispiele Namen von Personen, Unternehmen, Marken und Produkten. Alle diese Namen sind fiktiv und jegliche Ähnlichkeit mit Namen und Adressen realer Unternehmen ist rein zufällig.

Unter Umständen werden Fotografien und farbige Abbildungen nicht angezeigt, wenn Sie diese Informationen nicht in gedruckter Form verwenden.

Marken

IBM, das IBM-Logo, ibm.com und SPSS sind Marken der IBM Corporation und in vielen Ländern weltweit registriert. Eine aktuelle Liste der IBM-Marken finden Sie im Internet unter <http://www.ibm.com/legal/copytrade.shtml>.

Intel, das Intel-Logo, Intel Inside, das Intel Inside-Logo, Intel Centrino, das Intel Centrino-Logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium und Pentium sind Marken oder eingetragene Marken der Intel Corporation oder der Tochtergesellschaften des Unternehmens in den USA und anderen Ländern.

Linux ist eine eingetragene Marke von Linus Torvalds in den USA, anderen Ländern oder beidem.

Microsoft, Windows, Windows NT und das Windows-Logo sind Marken der Microsoft Corporation in den USA, anderen Ländern oder beidem.

UNIX ist eine eingetragene Marke der The Open Group in den USA und anderen Ländern.

Java und alle Java-basierten Marken sowie Logos sind Marken von Sun Microsystems, Inc. in den USA, anderen Ländern oder beidem.

Andere Produkt- und Servicenamen können Marken von IBM oder anderen Unternehmen sein.



Index

- 508-Kompatibilität, 264
- Abrufen von Objekten aus dem IBM SPSS Collaboration and Deployment Services Repository, 194
- abs (Funktion), 154
- Aggregatknoten
- performance, 261
- Aktualisieren
- Quellenknoten, 65
- Aktualisieren von Modellen, 213–214
- allbutfirst (Funktion), 159
- allbutlast (Funktion), 159
- alle Kommentare für einen Stream auflisten, 95
- alphabefore (Funktion), 159
- and (Operator), 154
- Ändern der Größe, 28
- Anmeldung bei IBM SPSS Modeler Server, 16
- Anmerkungen
- in Kommentare umwandeln, 96
 - Knoten, 88, 97
 - Ordner, 233
 - project, 232
 - Streams, 88, 97
- Anwendungen, 35
- Anwendungsbeispiele, 4
- Anwendungsbereiche des Data Mining, 36
- Anzeigeformate
- Dezimalstellen, 67
 - Symbol für Zifferngruppierung, 67
 - Währung, 67
 - Wissenschaftlich, 67
 - Zahlen, 67
- arccos (Funktion), 156
- arccosh (Funktion), 156
- arcsin (Funktion), 156
- arcsinh (Funktion), 156
- arctan (Funktion), 156
- arctan2 (Funktion), 156
- arctanh (Funktion), 156
- Attribut, 34
- Aufheben von Feldzuordnungen, 103
- Ausdrücke, 141
- Ausführen von Streams, 86
- Ausführung anhalten, 26
- Ausführungszeiten anzeigen, 75
- Ausgabe, 23
- Ausgabedateien
- speichern, 101
- Ausgabeknoten, 49
- Ausgabeobjekte
- Speichern im IBM SPSS Collaboration and Deployment Services Repository, 193
- Ausschneiden, 26
- Automatisierung, 118
- Baumbasierte Analyse
- Typische Anwendungen, 35
- Bedingungen, 124
- Befehlszeile
- Starten von IBM SPSS Modeler, 16
- Begrüßungsdialogfeld, 247
- Beispiele
- Anwendungshandbuch, 4
 - Übersicht, 6
- Benachrichtigungen
- Festlegen von Optionen, 243
- Benennen von Knoten und Streams, 97
- Benutzer-ID
- IBM SPSS Modeler Server, 16
- Benutzerdefiniert fehlende Werte, 112
- Benutzerdefinierte Funktionen (UDFs), 134
- Benutzerdefinierte Palettenerstellung, 252
- Unterpalettenerstellung, 255
- Benutzeroptionen, 243
- Bereitstellung, 182
- Bereitstellungstyp, 208
- Berichte
- erzeugen, 235
 - Festlegen von Eigenschaften, 235
 - Hinzufügen zu Projekten, 227
 - Speichern der Ausgabe, 101
- Beschriftungen
- Anzeigen, 65
 - value, 221
 - Variablen, 221
- Beschreibungen, IBM SPSS Collaboration and Deployment Services Repository-Objekt, 206
- Bewertung
- Verzweigung, 88, 212–213, 217
- Bildschirm-Lesesysteme, 266, 269, 275–276
- Beispiel, 273–274
- Bitweise Funktionen, 157
- @BLANK-Funktion, 116, 149, 176
- Blättern
- Festlegen von Optionen, 72
- Cache
- Aktivieren, 57, 242, 258
 - Einrichten eines Caches, 56
 - Leeren, 59, 65
 - Optionen für Knoten, 57, 258
 - speichern, 59
- Cache-Dateiknoten
- Laden, 102
- cdf_chisq (Funktion), 156
- cdf_f (Funktion), 156
- cdf_normal (Funktion), 156
- cdf_t (Funktion), 156
- Champion-Challenger-Analyse, 182, 208
- Chi-Quadrat-Verteilung
- Wahrscheinlichkeitsfunktionen, 156

- CLEM, 131
 - Ausdrücke, 124, 141
 - Beispiele, 121
 - Datentypen, 142–143
 - Einführung, 32, 118
 - Erstellen von Ausdrücken, 133
 - Funktionen, 134
 - language, 141
 - Überprüfen von Ausdrücken, 137
- CLEM Ausdrücke
 - Parameter, 76, 125
- CLEM Funktionen
 - Bitweise, 157
 - datetime, 165
 - Fehlende Werte, 116
 - Globalwert, 175
 - Information, 149
 - Leerstellen und Nullen, 176
 - Liste der verfügbaren Funktionen, 148
 - Logisch, 154
 - numerisch, 154
 - probability, 156
 - random, 158
 - sequence, 170–171
 - Sonderfunktionen, 177
 - string, 159
 - Trigonometrisch, 156
 - Umwandlung, 150
 - Vergleich, 151
- CLEM-Ausdrücke
 - performance, 262
 - Suchen und Ersetzen von Text, 138
- Client
 - Standardverzeichnis, 243
- Coordinator of Processes, 19
- COP, 19
- cos (Funktion), 156
- cosh (Funktion), 156
- count_equal, Funktion, 129, 151
- count_greater_than, Funktion, 129, 151
- count_less_than, Funktion, 129, 151
- count_non_nulls function, 151
- count_not_equal, Funktion, 129, 151
- count_nulls-Funktion, 116, 129, 151
- count_substring (Funktion), 159
- CRISP-DM, 25, 225
 - Projektansicht, 226
- CRISP-DM, Prozessmodell, 37–38

- Data Audit-Knoten
 - Verwendung bei der Datenuntersuchung, 34
 - Verwendung beim Data Mining, 36
- Data Mining, 34
 - Anwendungsbeispiele, 47
 - Strategie, 37
- date_before (Funktion), 151

- Daten
 - Vorschau, 60
- Daten-Streams
 - Erstellen, 48
- Datensätze, 34
 - Fehlende Werte, 114
- Datentypen, 123
 - in Parametern, 79
- Datenuzuordnung, 107
- Datenuzuordnungstool, 103–104
- datetime_date (Funktion), 150
- Datetime-Funktionen
 - datetime_date, 165
 - datetime_day, 165
 - datetime_day_name, 165
 - datetime_day_short_name, 165
 - datetime_hour, 165
 - datetime_in_seconds, 165
 - datetime_minute, 165
 - datetime_month, 165
 - datetime_month_name, 165
 - datetime_month_short_name, 165
 - datetime_now datetime_second, 165
 - datetime_time, 165
 - datetime_timestamp, 165
 - datetime_weekday, 165
 - datetime_year, 165
- Datums-/Uhrzeitwerte, 128
- Datumsangaben
 - bearbeiten, 169
 - Konvertieren, 169
- Datumsformate, 66, 144–145
- Datumsfunktionen, 144–145
 - date_before, 151, 165
 - date_days_difference, 165
 - date_in_days, 165
 - date_in_months, 165
 - date_in_weeks, 165
 - date_in_years, 165
 - date_months_difference, 165
 - date_weeks_difference, 165
 - date_years_difference, 165
 - @TODAY (Funktion), 165
- Deployment von Szenarien, 208
- Deployment-Optionen
 - Szenarien, 208
- Dezimalstellen
 - Anzeigeformate, 67
- Dezimaltrennzeichen
 - Zahlenanzeigeformate, 64
- Diagramme
 - Hinzufügen zu Projekten, 227
 - Speichern der Ausgabe, 101
- DIFF (Funktion), 171
 - @DIFF (Funktion), 170–171
- Direktzugriffstasten
 - Allgemeine Verwendung, 108

- Tastatur, 31, 266, 269, 271
- div (Funktion), 154
- Dokumentation, 4
- Domänenname (Windows)
 - IBM SPSS Modeler Server, 16
- Drucken, 32
 - Streams, 29, 56
- DTD, 221
- Duplikatknoten
 - performance, 261
- Eigenschaften
 - Berichtphasen, 235
 - für Daten-Streams, 62
 - Projektordner, 233
- Einfügen, 26
- Einführung, 141
 - IBM SPSS Modeler, 15, 241
- Eingabehilfen
 - Funktionen in IBM SPSS Modeler, 264
- Einzelanmeldung, 18, 184
- Einzelanmeldung, IBM SPSS Collaboration and Deployment Services Repository, 180, 183
- encoding, 64, 279
- Endknoten, 49
- endstring (Funktion), 159
- Enterprise-Ansichts-Knoten, 208
- Entscheidungsbäume
 - Zugriffsmöglichkeiten, 276
- Ent sperren von IBM SPSS Collaboration and Deployment Services Repository-Objekten, 201
- Erforderliche Felder, 108
- Ersetzen von Modellen, 245
- Ersetzen von Text, 138
- Evaluationsknoten
 - performance, 262
- exponential (Funktion), 154
- exportieren
 - Stream-Beschreibungen, 86
- Exportieren
 - PMML, 220, 223
- Expression Builder, 269
 - Suchen und Ersetzen von Text, 138
 - Übersicht, 131
 - verwenden, 133
 - Zugriff, 133
- f*-Verteilung
 - Wahrscheinlichkeitsfunktionen, 156
- Faktor, 276
- Fall, 34
- Farben
 - festlegen, 246
- Fehlende Werte, 113–114, 127
 - Behandlung, 112
 - CLEM-Ausdrücke, 116
 - Füllen, 112
- In Datensätzen, 114
- Fehlermeldungen, 73
- Felder, 34, 141, 143
 - Anzeigen von Werten, 136
 - in CLEM-Ausdrücken, 135
- Fenster “Interaktiver Baum”
 - Zugriffsmöglichkeiten, 276
- @FIELD-Funktion, 116, 177
- @FIELDS_BETWEEN-Funktion, 116, 129, 177
- @FIELDS_MATCHING-Funktion, 116, 129, 177
- first_index-Funktion, 130, 151
- first_non_null-Funktion, 130, 151
- first_non_null_index-Funktion, 130, 151
- fracof (Funktion), 154
- Füllerknoten
 - Fehlende Werte, 115
- Funktionen, 144–145, 149, 170
 - Beispiele, 121
 - Benutzerdefinierte Funktionen (UDFs), 134
 - @BLANK, 115
 - @FIELD, 131, 177
 - @GLOBAL_MAX, 175
 - @GLOBAL_MEAN, 175
 - @GLOBAL_MIN, 175
 - @GLOBAL_SDEV, 175
 - @GLOBAL_SUM, 175
 - in CLEM-Ausdrücken, 134
 - @PARTITION, 177
 - @PREDICTED, 131, 177
 - Stufen der, 134
 - @TARGET, 131, 177
 - Umgang mit fehlenden Werten, 116
- Ganze Zahlen, 141–142
- Generierte Modelle, Palette, 23
- gleich (Operator), 151
- Globale Funktionen, 175
- Globalwerte
 - in CLEM-Ausdrücken, 135
- Grad
 - Maßeinheiten, 68
- Grafiken
 - Speichern der Ausgabe, 101
- größer als (Operator), 151
- hasendstring (Funktion), 159
- hasmidstring (Funktion), 159
- hasstartstring (Funktion), 159
- hassubstring, Funktion, 159
- Hauptfenster, 22
- Hinweise
 - Allgemeine Verwendung, 108
 - hinzufügen
 - Zu einem Projekt, 227
 - Hinzufügen von IBM SPSS Modeler Server-Verbindungen, 18–19

- Hostname
 IBM SPSS Modeler Server, 16, 18
- Hotkeys, 31
- HTML-Ausgabe
 Bildschirm-Lesesystem, 276
- IBM InfoSphere Warehouse (ISW)
 PMML-Export, 223
- IBM SPSS Collaboration and Deployment Services, 182
- IBM SPSS Collaboration and Deployment Services Enterprise View, 182, 208
- IBM SPSS Collaboration and Deployment Services Repository, 180, 182
 Abrufen von Objekten, 194
 Durchsuchen, 185
 Einzelanmeldung, 183
 Löschen von Objekten und Versionen, 201
 Objekteigenschaften, 204
 Ordner, 200, 202
 Speichern von Objekten, 186
 Sperren und Ent sperren von Objekten, 201
 Suche in, 197
 Übertragen von Projekten, 230
 Verbindung, 183–184
- IBM SPSS Modeler, 1, 21
 Dokumentation, 4
 Eingabehilfen, 264
 Erste Schritte, 15
 Optionen, 241
 Tipps und Direktzugriffstasten, 108
 über Befehlszeile ausführen, 16
 Übersicht, 15, 241
- IBM SPSS Modeler Advantage, 182, 207
- IBM SPSS Modeler Server
 Benutzer-ID, 16
 Domänenname (Windows), 16
 Hostname, 16, 18
 password, 16
 Portnummer, 16, 18
- IBM SPSS Statistics-Modelle, 47
- if, then, else, Funktionen, 154
- Importieren
 PMML, 221, 223
- INDEX (Funktion), 171
- @INDEX (Funktion), 170–171
- Informationsfunktionen, 149
- integer_bitcount (Funktion), 157
- integer_leastbit (Funktion), 157
- integer_length (Funktion), 157
- intof (Funktion), 154
- is_date (Funktion), 149
- is_datetime (Funktion), 149
- is_integer (Funktion), 149
- is_number (Funktion), 149
- is_real (Funktion), 149
- is_string (Funktion), 149
- is_time (Funktion), 149
- is_timestamp (Funktion), 149
- isalphacode (Funktion), 159
- isendstring (Funktion), 159
- islowercode (Funktion), 159
- ismidstring (Funktion), 159
- isnumbercode (Funktion), 159
- isstartswith (Funktion), 159
- issubstring (Funktion), 159
- issubstring_count (Funktion), 159
- issubstring_lim (Funktion), 159
- isuppercode (Funktion), 159
- Java, 278
- JAWS, 264, 275, 277–278
- K-Means-Knoten
 Große Sets, 65
 performance, 262
- keywords
 Anmerkungen zu Knoten, 97
- Klassen, 25, 225, 227
- Klasserknoten
 performance, 261
- kleiner als (Operator), 151
- Knoten, 15
 Aktivieren, 53
 Anmerkungen, 88, 97
 Anzeigen auf Palette, 252
 Ausführungszeiten, 75
 bearbeiten, 56
 Benutzerdefinierte Palettenerstellung, 252
 Benutzerdefinierte Unterpalettenerstellung, 255
 Datenvorschau, 60
 Deaktivieren, 53, 56
 Deaktivieren in einem Stream, 53
 Duplizieren, 56
 Einführung, 49
 Entfernen aus Palette, 252
 Festlegen von Optionen, 56
 hinzufügen, 50, 54
 Hinzufügen zu Projekten, 227–228
 Kommentare hinzufügen zu, 88
 Laden, 102
 löschen, 50
 Löschen von Verbindungen, 55
 Palettenregisterkarte, Anpassung, 256
 performance, 260, 262
 Reihenfolge, 257
 speichern, 99
 Speichern im IBM SPSS Collaboration and Deployment Services Repository, 193
- Sperren, 61
- Suchen, 81
- Umgehen in einem Stream, 52
- Verbinden in einem Stream, 50
- Vorschau von Daten, 60

- Knoten aktivieren , 53
- Knoten deaktivieren , 53, 56
- Knoten-Caching
 - Aktivieren, 57, 258
- Knotennamen, 97
- Knotenpalettenauswahl, 252
- Knowledge Discovery, 34
- Kohonen-Knoten
 - Große Sets, 65
 - performance, 262
- Komma, 64
- Kommentare
 - alle in einem Stream auflisten, 95
 - Tastaturkombinationen, 271
 - zu Knoten und Streams, 88
- Konventionen, 149
- Konvertierungsfunktionen, 150
- Kopieren, 26

- Laden
 - Knoten, 102
 - Statusangaben, 102
- Ländereinstellung
 - Optionen, 241
- last_index-Funktion, 130, 151
- LAST_NON_BLANK (Funktion), 171
- @LAST_NON_BLANK (Funktion), 170–171, 176
- last_non_null-Funktion, 130, 151
- last_non_null_index-Funktion, 130, 151
- Laufzeit, Eingabeaufforderungen während, 78
- Leerer Bereich
 - Entfernen aus Zeichenketten, 126, 159
- Leerstellenbehandlung
 - CLEM Funktionen, 176
- Leerzeichen, 112–113, 127
 - Entfernen aus Zeichenketten, 126, 159
- length (Funktion), 159
- Lineare Regression
 - Als PMML exportieren, 249
- Listen, 141, 143
- locchar (Funktion), 159
- locchar_back (Funktion), 159
- log (Funktion), 154
- log10 (Funktion), 154
- Logische Funktionen, 154
- Logistische Regression, 276
 - Als PMML exportieren, 249
- lowertoupper, Funktion, 159

- Manager, 23
- Marken, 282
- Maschinenlernen, 34
- matches (Funktion), 159
- Maus
 - Verwenden in IBM SPSS Modeler, 30, 50
- max (Funktion), 151
- MAX (Funktion), 171
- @MAX (Funktion), 170–171
- max_index-Funktion, 130, 151
- max_n (Funktion), 129, 151
- MEAN (Funktion), 170–171
- @MEAN (Funktion), 170–171
- mean_n (Funktion), 129, 154
- Mehrere IBM SPSS Modeler-Sitzungen, 21
- Mehrfachantworten-Sets
 - in CLEM-Ausdrücken, 135
 - In CLEM-Ausdrücken, 130
- Meldungen
 - Anzeigen von erzeugter SQL, 71
- member (Funktion), 151
- Merge, Knoten
 - performance, 261
- Merkmalsauswahlknoten
 - Fehlende Werte, 114
- min (Funktion), 151
- MIN (Funktion), 171
- @MIN (Funktion), 170–171
- min_index-Funktion, 130, 151
- min_n (Funktion), 129, 151
- Minimierung, 28
- Mittlere Maustaste
 - Simulieren, 30, 50
- mod (Funktion), 154
- Modell-Nuggets, 87
- Modellaktualisierung, 208
- Modelle, 87
 - Aktualisieren, 213–214
 - Ersetzen, 245
 - exportieren, 248
 - Hinzufügen zu Projekten, 227
 - Speichern im IBM SPSS Collaboration and Deployment Services Repository, 194
- Modellierung
 - Verzweigung, 88
- Modellierungsknoten, 40, 49
 - Modellierungs-Palettenregisterkarte, Anpassung, 256
 - performance, 262
- @MULTI_RESPONSE_SET-Funktion, 130, 177

- Navigieren
 - Tastaturkombinationen, 266
- negate (Funktion), 154
- Netzwerknoten
 - performance, 262
- Neue Funktionen, 8, 13
- Neuronaler Netzknoten
 - Große Sets, 65
- Normalverteilung
 - Wahrscheinlichkeitsfunktionen, 156
- not (Operator), 154
- Nuggets, 87
 - definiert, 24
- @NULL-Funktion, 116, 149, 176
- Nullen, 112, 127

- Numerische Funktionen, 154
- Objekte
- Eigenschaften, 234
 - Objekteigenschaften, IBM SPSS Collaboration and Deployment Services Repository, 204
 - öffnen
 - Ausgabe, 102
 - Knoten, 102
 - Modelle, 102
 - Projekte, 227
 - Statusangaben, 102
 - Streams, 102
 - OFFSET (Funktion), 171
 - @OFFSET, Funktion, 170–171
 - Erläuterungen zur Leistung, 262
 - oneof (Funktion), 158
 - Operatoren
 - In CLEM-Ausdrücken, 134
 - Verbinden von Zeichenketten, 150
 - Operatorenrangfolge, 145
 - Optimierung, 68
 - Optionen, 241
 - Benutzer-, 243
 - display, 246
 - für IBM SPSS Modeler, 241
 - PMML, 248
 - Stream-Eigenschaften, 62–63, 66–68, 71–73, 75
 - or (Operator), 154
 - Ordner, IBM SPSS Collaboration and Deployment Services Repository, 200, 202
- Palette der Modelle, 194
- Paletten, 22
 - Anpassen, 250
- Palettenregisterkarte, Anpassung, 256
- Parallele Verarbeitung
- Aktivieren, 68
- Parameter
- Bewertung, 211
 - Eingabeaufforderungen während der Laufzeit, 78
 - in CLEM-Ausdrücken, 135
 - Modellerstellung, 211
 - Sitzung, 76, 78–79, 125
 - Stream, 76, 78–79, 125
 - Typ, 79
 - Verwendung in Szenarios, 211
- @PARTITION_FIELD-Funktion, 177
- password
- IBM SPSS Modeler Server, 16
- performance
- CLEM-Ausdrücke, 262
 - Knoten-Caching, 57, 258
 - Modellierungsknoten, 262
 - von Prozessknoten, 260
- pi (Funktion), 156
- PMML
- Exportieren von Modellen, 220, 223
 - Exportoptionen, 248
 - Importieren von Modellen, 221, 223
- PMML-Modelle
- Lineare Regression, 249
 - Logistische Regression, 249
- Portnummer
- IBM SPSS Modeler Server, 16, 18
- power (exponentielle Funktion), 154
- PowerPoint-Dateien, 227
- @PREDICTED-Funktionen, 177
- Predictive Applications, 208
- Projekte, 25, 225
 - Anmerkungen, 232
 - Ansicht “Klassen”, 227
 - CRISP-DM, Ansicht, 226
 - Einrichten eines Standardordners, 226
 - Erstellen, 227
 - Festlegen von Eigenschaften, 231
 - Generieren von Berichten, 235
 - Hinzufügen von Objekten, 228
 - im IBM SPSS Collaboration and Deployment Services Repository, 230
 - Neu erstellen, 228
 - Objekteigenschaften, 234
 - Ordnereigenschaften, 233
 - Schließen, 235
 - Speichern im IBM SPSS Collaboration and Deployment Services Repository, 192
- Protokolldateien
- Anzeigen von erzeugter SQL, 71
- Prozessknoten, 49
 - performance, 260
- Punkt, 64
- Pushbacks, 68
- Qualitätsknoten
- Fehlende Werte, 114
 - Quellenknoten, 49
 - Aktualisieren, 65
 - Datenzuordnung, 104
 - Quickinfos
 - Anmerkungen zu Knoten, 97
- Radian
- Maßeinheiten, 68
- random (Funktion), 158
- random0 (Funktion), 158
- Rangfolge, 145
- Rechtliche Hinweise, 280
- Reelle Zahlen, 141–142
- Regelerstellungsknoten
- Laden, 102
- Regelmengen
- Evaluieren, 64

- regression, 276
- rem (Funktion), 154
- replace (Funktion), 159
- replicate (Funktion), 159
- round (Funktion), 154
- Rückgängig, 26

- SAS-Dateien
 - encoding, 279
- Schriftarten, 246
- SDEV (Funktion), 171
- @SDEV (Funktion), 170–171
- sdev_n (Funktion), 129, 154
- Sequenzfunktionen, 170–171
- Server
 - Anmeldung, 16
 - Hinzufügen von Verbindungen, 18
 - Standardverzeichnis, 243
 - Suchen nach COP für Server, 19
- Set aus dichotomen Variablen
 - In CLEM-Ausdrücken, 130
- set, Befehl, 76, 125
- Sets, 65
- Sets aus kategorialen Variablen
 - In CLEM-Ausdrücken, 130
- sign (Funktion), 154
- sin (Funktion), 156
- SINCE (Funktion), 171
- @SINCE (Funktion), 170–171
- sinh (Funktion), 156
- Sitzungsparameter, 76, 78–79, 125
- Skalierung der Streams für Ansicht, 29
- skipchar (Funktion), 159
- skipchar_back (Funktion), 159
- Skripts, 32, 118
 - Suchen und Ersetzen von Text, 138
- Sonderfunktionen, 177
- Sonderzeichen
 - Entfernen aus Zeichenketten, 126
- Sortierknoten
 - performance, 260
- soundex (Funktion), 164
- soundex_difference (Funktion), 164
- Speicher
 - Verwalten, 241–242
- speichern
 - Ausgabeobjekte, 101
 - Knoten, 99
 - Mehrere Objekte, 100
 - Statusangaben, 99–100
 - Streams, 99
 - Speichern von Objekten im IBM SPSS Collaboration and Deployment Services Repository, 186
 - Sperren von IBM SPSS Collaboration and Deployment Services Repository-Objekten, 201
 - Sperren von Knoten, 61
- Sprache
 - Optionen, 241
- SPSS Modeler Server, 2
- SQL-Erzeugung, 68
 - Protokollierung, 71
 - Vorschau, 71
- sqrt (Funktion), 154
- Standard
 - Projektphase, 226
- Startdialogfeld, 247
- startstring (Funktion), 159
- Statistics-Modelle, 47
- StatisticsDateien
 - encoding, 279
- Statusangaben
 - Laden, 102
 - speichern, 99–100
- Stream, 22
- Stream-Beschreibungen, 83, 86
- Stream-Namen, 97
- Stream-Neuerstellung
 - Aktivieren, 68
- Stream-Parameter, 76, 78–79, 125
- Stream-Sicherungsdateien
 - Wiederherstellen, 99
- Stream-Standardkodierung, 64
- Stream-Zeichenbereich
 - Einstellungen, 72
- Streams, 15
 - Anmerkungen, 88, 97
 - Anzeigen der Ausführungszeiten, 75
 - ausführen, 86
 - Deployment-Optionen, 208
 - Erstellen, 48
 - Hinzufügen von Knoten, 50, 54
 - Hinzufügen von Kommentaren, 88
 - Hinzufügen zu Projekten, 227–228
 - Knoten deaktivieren, 53
 - Laden, 102
 - Optionen, 62–63, 66–68, 71–72
 - Sicherungsdateien, 99
 - Skalierung für Ansicht, 29
 - speichern, 99
 - Speichern im IBM SPSS Collaboration and Deployment Services Repository, 192
 - umbenennen, 83, 97
 - Umgehen von Knoten, 52
 - Verbinden von Knoten, 50
- String-Funktionen, 159
- stripchar (Funktion), 159
- strmember (Funktion), 159
- Stufen der
 - Funktionen, 134
- subscr (Funktion), 159
- substring (Funktion), 159
- substring_between (Funktion), 159

- Suchen
 - Knoten in einem Stream, 81
 - Suchen nach COP für Verbindungen, 19
 - Suchen nach Objekten im IBM SPSS Collaboration and Deployment Services Repository, 197
 - Suchen nach Text, 138
 - SUM (Funktion), 171
 - @SUM (Funktion), 170–171
 - sum_n (Funktion), 129, 154
 - Superknoten
 - Parameter, 76, 125
 - Symbol für Zifferngruppierung
 - Zahlenanzeigeformate, 64
 - Symbole
 - Festlegen von Optionen, 29, 72
 - Symbolleiste, 26
 - System
 - Optionen, 241
 - Systemdefiniert fehlende Werte, 112
 - Szenarien, 207
 - definiert, 182
 - Deployment-Optionen, 208
 - t*-Verteilung
 - Wahrscheinlichkeitsfunktionen, 156
 - Tabellen, 269
 - Hinzufügen zu Projekten, 227
 - Speichern der Ausgabe, 101
 - Tage übertragen, 66
 - tan (Funktion), 156
 - tanh (Funktion), 156
 - @TARGET-Funktionen, 177
 - Tastaturlayout, 266, 269, 271
 - Temp-Verzeichnis, 20
 - testbit (Funktion), 157
 - @TESTING_PARTITION-Funktion, 177
 - Texdatadateien
 - encoding, 279
 - Textkodierung, 64
 - THIS (Funktion), 171
 - @THIS (Funktion), 170–171
 - time_before (Funktion), 151
 - Tipps
 - Allgemeine Verwendung, 108
 - für Eingabehilfen, 276
 - to_date (Funktion), 150, 165
 - to_dateline (Funktion), 165
 - to_datetime (Funktion), 150
 - to_integer (Funktion), 150
 - to_number (Funktion), 150
 - to_real (Funktion), 150
 - to_string (Funktion), 150
 - to_time (Funktion), 150, 165
 - to_timestamp (Funktion), 150, 165
 - @TODAY (Funktion), 165
 - @TRAINING_PARTITION-Funktion, 177
 - Trigonometrische Funktionen, 156
 - trim (Funktion), 159
 - trim_start (Funktion), 159
 - trimend (Funktion), 159
 - Typische Anwendungen, 35
 - Typknoten
 - Fehlende Werte, 115
 - performance, 262
 - Überprüfen von CLEM-Ausdrücken, 137
 - umbenennen
 - Knoten, 97
 - Streams, 83
 - Umgekehrter Schrägstrich in CLEM-Ausdrücken, 143
 - undef (Funktion), 176
 - ungleich (Operator), 151
 - unicode_char (Funktion), 159
 - Unicode-Unterstützung, 279
 - unicode_value (Funktion), 159
 - Unterpalette
 - Anzeigen auf Palettenregisterkarte, 254
 - Entfernen aus Palettenregisterkarte, 254
 - Erstellung, 255
 - uppertolower (Funktion), 159
 - UTF-8-Kodierung, 64, 279
 - @VALIDATION_PARTITION-Funktion, 177
 - value_at-Funktion, 130, 151
 - values, 123
 - Variablen, 34
 - Verbindungen
 - mit IBM SPSS Collaboration and Deployment Services Repository, 183–184
 - mit IBM SPSS Modeler Server, 16, 18–19
 - Server-Cluster, 19
 - Vergleichsfunktionen, 151
 - Verketten von Zeichenfolgen, 150
 - Verrauschte Daten, 36
 - Versionsbeschriftungen, IBM SPSS Collaboration and Deployment Services Repository-Objekt, 206
 - Verteilungsfunktionen, 156
 - Verzeichnis
 - Standard, 243
 - Verzweigungen, Scoring und Modellierung, 88, 212–213, 217
 - Violette Knoten, 68
 - Visuelle Programmierung, 21
 - Vorlagen, 104
 - Vorlagenfelder, 108
 - Vorschau
 - Knotendaten, 60
 - Wahrscheinlichkeitsfunktionen, 156
 - Währungsanzeigeformat, 67
 - Warnungen, 73
 - Festlegen von Optionen, 243

Werte

Anzeigen von einem Data Audit aus, 136

Hinzufügen zu CLEM-Ausdrücken, 136

Wichtige Felder, 103, 107

Wissenschaftliche Notation

Anzeigeformat, 67

Wörterbuchdatei, 275

Zahlen, 128, 142

Zahlenanzeigeformate, 67

Zeichen, 141, 143

Zeichenbereich, 22

Zeichenketten, 141, 143

Abgleichen, 126

Ändern in CLEM-Ausdrücken, 126

Ersetzen, 126

Zeit- und Datumsfunktionen, 144–145

Zeitfelder

Konvertieren , 169

Zeitformate, 66, 144–145

Zeitfunktionen, 144–145

time_before, 151, 165

time_hours_difference, 165

time_in_hours, 165

time_in_mins, 165

time_in_secs, 165

time_mins_difference, 165

time_secs_difference, 165

Zoomen, 26

Zugriffsmöglichkeiten, 264, 278

Beispiel, 273–274

Tipps in IBM SPSS Modeler, 276

Zuordnen von Feldern, 103