

Leseprobe

Holger Schrödl

Business Intelligence mit Microsoft SQL Server 2008

BI-Projekte erfolgreich umsetzen

ISBN: 978-3-446-41210-1

Weitere Informationen oder Bestellungen unter

<http://www.hanser.de/978-3-446-41210-1>

sowie im Buchhandel.



## 5 Einsatzszenarien und Fallbeispiele

Mit dem ersten BI-Projekt hat man einen kleinen Einblick in die Mächtigkeit der Technologie gewonnen. In dem Beispielprojekt wurde eine der im Microsoft SQL Server 2008 implementierten Data-Mining-Methoden verwendet: der Entscheidungsbaum. Nun sind aber noch weitere Methoden implementiert, die ganz unterschiedlich je nach Aufgabenstellung eingesetzt werden können.

In größeren Projekten ist es häufig so, dass nicht nur eine Methode zum Einsatz kommt, sondern die gesamte Analyse aus verschiedenen Methoden zusammengesetzt ist. Da jedes Verfahren seinen spezifischen Einsatzbereich aufweist, werden für die einzelnen Teilbereiche die am besten geeigneten Verfahren eingesetzt. Die erzielten Teilergebnisse werden dann entweder konsolidiert oder dienen selbst wieder als Eingangsinformationen für eine andere Analyse. Beispielsweise könnte eine Assoziationsanalyse eingesetzt werden, um Kundengruppen zu identifizieren. Dieses Ergebnis könnte als Eingangsinformation für eine Zeitreihenanalyse genutzt werden, welche die weitere Entwicklung der Kundensegmente analysiert.

Im Folgenden sollen einige weit verbreitete Anwendungen betrachtet werden.

### 5.1 Warenkorbanalyse

---

Die Warenkorbanalyse ist eine der frühen Anwendungen im Data-Mining-Umfeld. Sie ist mittlerweile ein etabliertes Verfahren und wird häufig genutzt, um Produktsortimente zu bearbeiten und die verschiedenen Werbemaßnahmen hierfür zu optimieren. Gerade der Einzelhandel ist eine Branche, die sehr stark auf solche Analysen setzt. Die große Produktmenge auf der einen Seite und die mittlerweile strukturierte Erfassung der Verkaufsdaten durch die bestehenden Kassensysteme ermöglicht detaillierte Analysen auf sehr hohem Niveau. Eine der Hauptforderungen nach Datenqualität und Datenmenge ist hier ausreichend gut erfüllt.

Das Prinzip der Warenkorbanalyse lässt sich sehr kurz zusammenfassen. Jeder Verkaufsvorgang wird als Datensatz betrachtet, der aus verschiedenen Einzelpositionen besteht

(bildhaft gesprochen der Kassenzettel, den der Kunde bekommt). Diese Datensätze werden alle nebeneinander gelegt, und es wird analysiert, ob es Einzelpositionen gibt, die mit einer gewissen Häufigkeit immer gemeinsam auftreten (Nudeln und Nudelseife werden tendenziell immer gemeinsam gekauft). Aus dieser Analyse entstehen dann Gruppen von Artikeln, die in einer mehr oder weniger ausgeprägten Beziehung zueinander stehen.

Die Kenntnis über solche Artikelbeziehungen können dann vielfältig genutzt werden:

- Artikel, die häufig zusammen gekauft werden, können nebeneinander platziert werden<sup>1</sup>, um bei den Kunden, die bisher nur einen der Artikel gekauft haben, ein Kaufinteresse für die anderen Artikel zu wecken.
- Es können Artikelempfehlungen ausgesprochen werden („Kunden, die dieses Produkt gekauft haben, haben sich auch für dieses und jenes interessiert.“).
- Artikel können gebündelt und so neue Angebote geschaffen werden, die für viele Kunden attraktiv sein können.
- Verkäufer können Kunden gezielt Empfehlungen geben und so aktiv den Verkaufsprozess unterstützen.

Und deren noch vieles mehr. Bei all den Möglichkeiten dennoch der Hinweis auf einen wichtigen Punkt: Diese Verfahren stellen statistische Verfahren dar. In dem folgenden Beispiel wird aus einer Vielzahl von Verkäufen auf die mögliche Struktur eines einzelnen, zukünftigen Verkaufs geschlossen. Statistische Verfahren können nur Tendenzen abgeben, sie können keine Garantien abgeben, dass sich eine Tendenz genau so in Realität einstellt, wie die Analyse sie zutage bringt. Es sind viele Parameter an der Analyse beteiligt, welche die Qualität der Prognose essenziell beeinflussen. Darum sollte zwingend bei solchen Analysen immer das Bild des BI-Zyklus im Blickfeld bleiben. Eine einmalige Analyse gibt eine gewisse Richtung vor. Erst die Umsetzung, die Bewertung der erzielten Ergebnisse sowie die erneute Durchführung der Analyse mit veränderten Erkenntnissen aus der Umsetzung liefern einen praktikablen Weg zu mehr Intelligenz im Geschäft: eben Business Intelligence.

Solch eine Warenkorbanalyse lässt sich mit den Data-Mining-Mitteln des Microsoft SQL Servers 2008 bewerkstelligen. Das folgende Beispiel zeigt den Weg auf, bestehend auf einem Beispieldatenbestand, eine solche Analyse durchzuführen.

Als Beispieldatenbestand werden wieder Teile der mit SQL Server 2008 mitgelieferten Beispieldatenbank AdventureWorksDW2008 genutzt. Leser, die bereits das Beispiel in Kapitel 4 durchgeführt haben, haben die entsprechenden Vorbereitungen getroffen, um direkt ins Data Mining einsteigen zu können. Ansonsten müssten, um das folgende Beispiel durchführen zu können, noch die Teilschritte

---

<sup>1</sup> Wenn hier von „Platzieren“ gesprochen wird, bedeutet das nicht nur einen physischen Platz im Einkaufsregal eines Supermarktes. Auch Produktplatzierungen in Webshops oder allgemein in Werbeangeboten können hiermit optimiert werden.

- Datenquelle erstellen (auf AdventureWorksDW2008)
- Datensicht erstellen

durchgeführt werden.

Das Beispiel gliedert sich in zwei Teilschritte: Zunächst wird das Data-Mining-Modell erstellt und kalkuliert. Im zweiten Schritt erfolgt eine Interpretation der Ergebnisse.

### 5.1.1 Erstellen des Data-Mining-Modells für eine Warenkorbanalyse

Zum Erstellen des Data-Mining-Modells wird das Business Intelligence Development Studio des Microsoft SQL Servers 2008 verwendet.

Bevor das Projekt gestartet werden kann, muss noch eine Konfiguration an dem Beispiel Data Warehouse durchgeführt werden. Die Daten, die für das Modell genutzt werden, liegen in zwei Sichten des Data Warehouse. Es muss noch eine Relation (genauer gesagt eine Many-to-one Relationship) zwischen diesen beiden Sichten hergestellt werden, um das Mining-Modell aufbauen zu können.

Um diese Relation aufzubauen, wird im Business Intelligence Development Studio die Design-Ansicht der Datenquellensicht „Adventure Works DW2008“ geöffnet. Im Tabellenfenster auf der linken Seite werden die Sichten vAssocSeqLineItems und vAssocSeqOrders geöffnet.

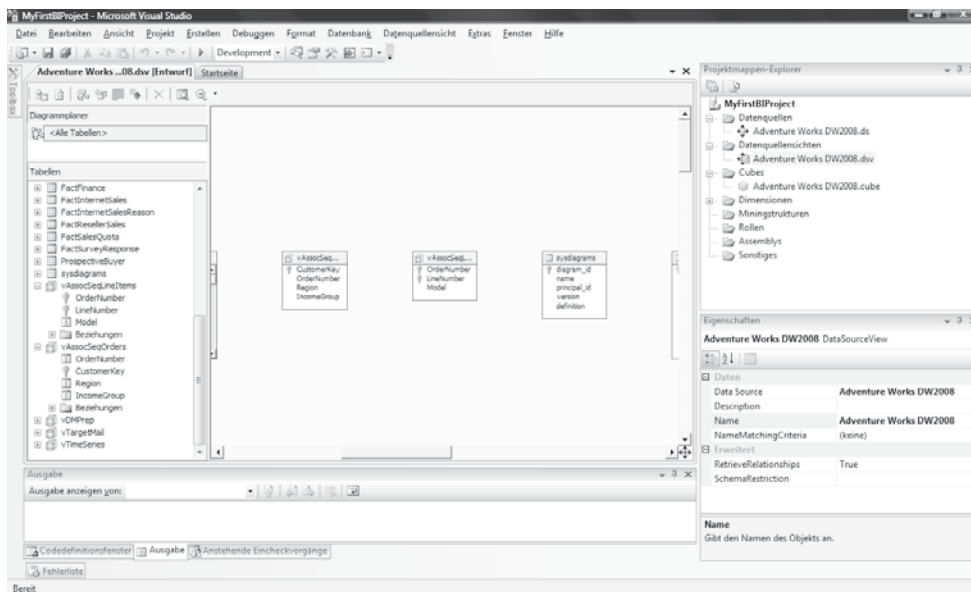
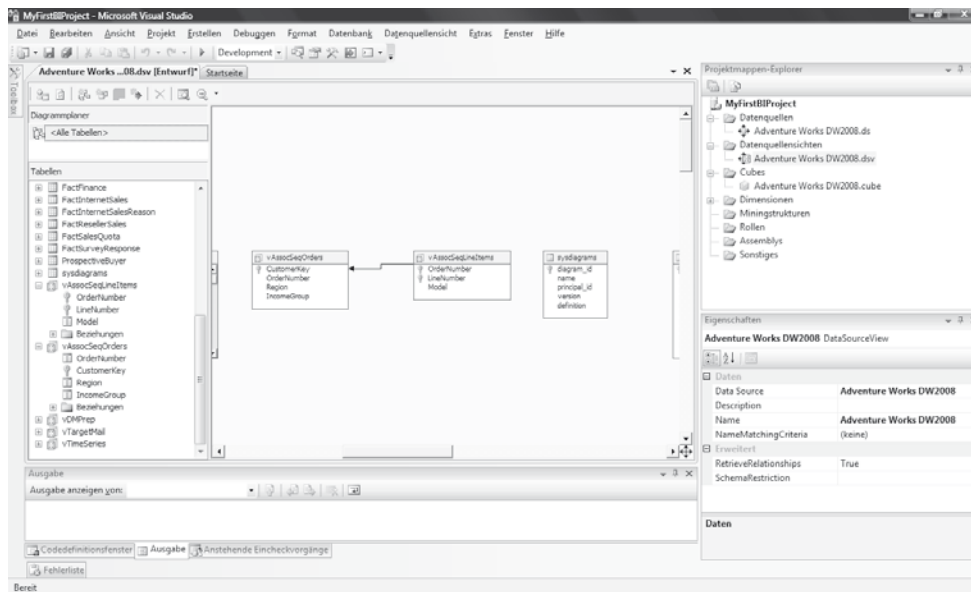


Abbildung 5.1 Design-Ansicht der Datenquellensicht „Adventure Works DW2008“

Nun wird mit der Maus die Spalte „OrderNumber“ in der Sicht „vAssocSeqOrders“ ausgewählt und auf die Spalte „OrderNumber“ der Sicht „vAssocSeqLineItems“ gezogen. Mit

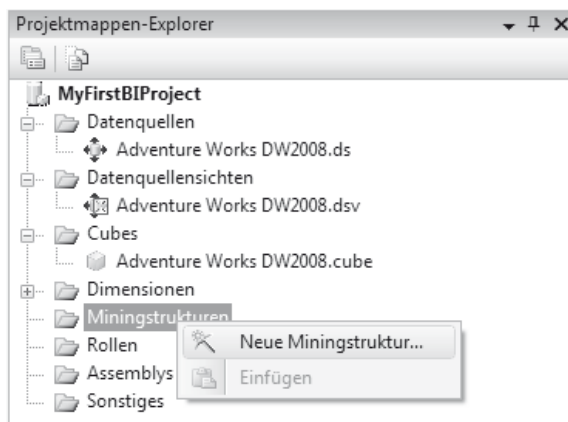
diesem Vorgang wird die gewünschte Relation zwischen den beiden Sichten hergestellt. Das Ergebnis sollte wie folgt aussehen:



**Abbildung 5.2** Ergebnis der neuen Relation zwischen „vAssocSeqLineItems“ und „vAssocSeqOrders“

Nun kann die Erstellung des Mining-Modells starten.

Im Projektmappe-Explorer wird im Kontextmenü der Mining-Strukturen eine neue Mining-Struktur erstellt.



**Abbildung 5.3** Neue Mining-Struktur im Projektmappe-Explorer erstellen

Nach Bestätigung des Willkommens-Bildschirms durch **Weiter** öffnet sich der Data-Mining-Assistent mit der Abfrage nach der Definitionsmethode. Für dieses Projekt wird

kein Cube verwendet, sondern die normale relationale Struktur des Data Warehouses. Es wird daher die Methode „Aus vorhandener relationaler Datenbank oder aus vorhandenem Data Warehouse“ ausgewählt.

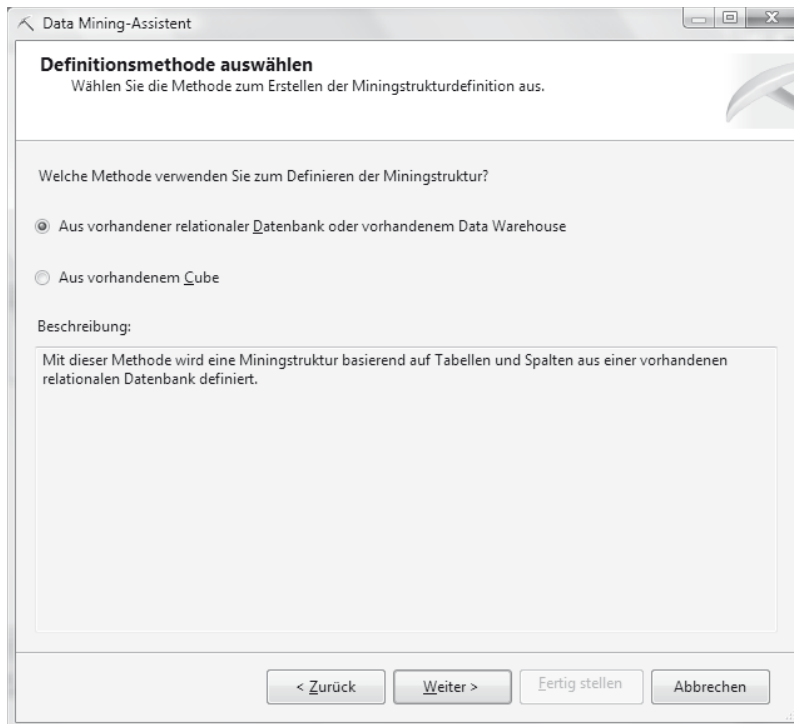
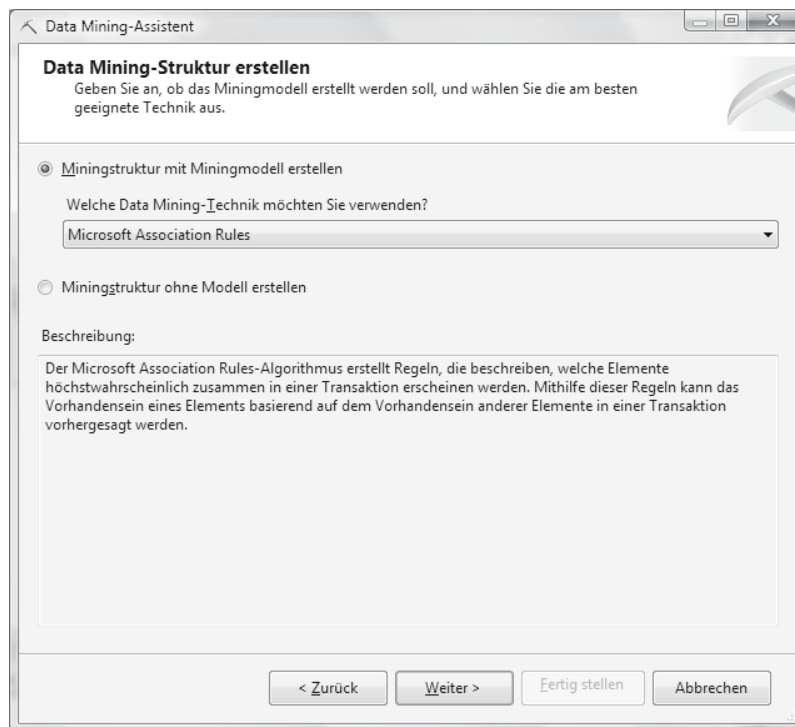


Abbildung 5.4 Auswahl der Definitionsmethode „Aus vorhandenem Data Warehouse“

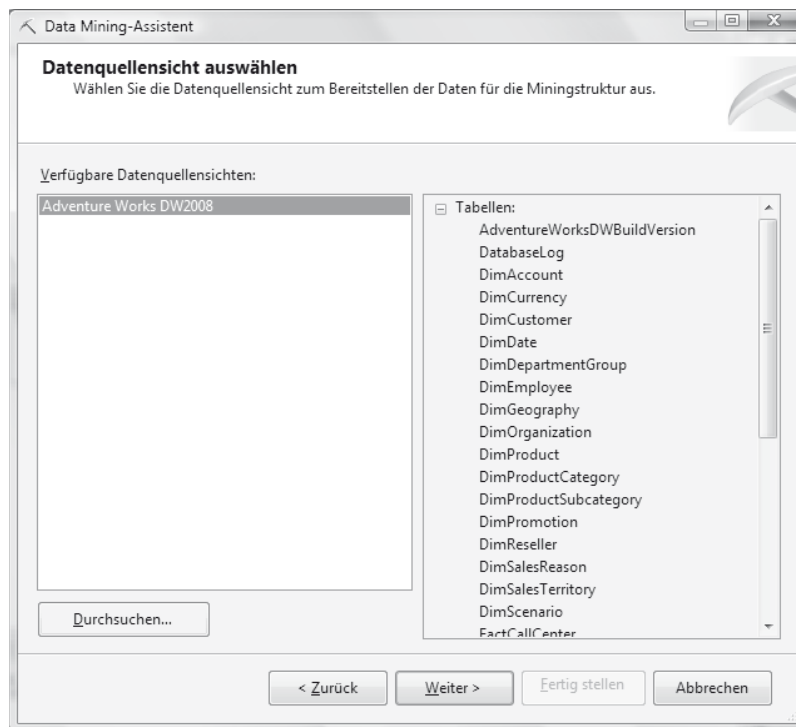
Die Bestätigung der Auswahl durch **Weiter** führt zur Abfrage der Data-Mining-Technik. Erinnern wir uns kurz an die Fragestellung: Es geht darum, Artikelabhängigkeiten zu erkennen und Gruppen zu identifizieren. Um dies zu erreichen, werden in der Regel Assoziationsmethoden eingesetzt, so auch in diesem Beispiel. Es wird die Data-Mining-Technik „Microsoft Association Rules,<sup>2</sup> ausgewählt.

<sup>2</sup> Mit der Übersetzung der Data-Mining-Technik hat Microsoft kein so ganz glückliches Händchen bewiesen. In SQL Server 2005 wurde an dieser Stelle die Methode mit „Zuordnungsregel“ übersetzt. Mein damaliger Hinweis an dieser Stelle, dass die Beibehaltung der englischen Begriffe besser gewesen wäre, wurde erhört: Nun sind alle Techniken mit dem auch in der Literatur verwendeten Begriff bezeichnet.



**Abbildung 5.5** Auswahl der Data-Mining-Technik „Microsoft Zuordnungsregeln“

Die Auswahl wird mit **Weiter** bestätigt. Der nächste Schritt beinhaltet die Auswahl der Datenquellensicht, die verwendet werden soll. Hier wird die bereits erstellte Datenquellensicht **Adventure Works DW2008** verwendet.



**Abbildung 5.6** Auswahl der Datenquellensicht **Adventure Works DW2008**

Auch diese Auswahl wird durch einen Klick auf **Weiter** bestätigt.

Nun erfolgt die Angabe der Falltabellen. Dies ist die Auswahl der Daten, die verwendet werden sollen, um das Mining-Modell aufzubauen. Die Daten über die Verkaufstransaktionen sind im AdventureWorks-Data Warehouse unter zwei Sichten zusammengefasst:

- **vAssocSeqLineItems**
- **vAssocSeqOrders**

Diese beiden Sichten werden verwendet, um die Daten für das Mining-Modell zur Verfügung zu stellen. Dabei wird die Sicht **vAssocSeqOrders** als Falltabelle verwendet (hier sind die ganzen Bestellungen enthalten) und die Sicht **vAssocSeqLineItems** als geschachtelte Tabelle (hier stehen die Unterinformationen zu den einzelnen Bestellungen).

In der Auswahl sieht das wie folgt aus:



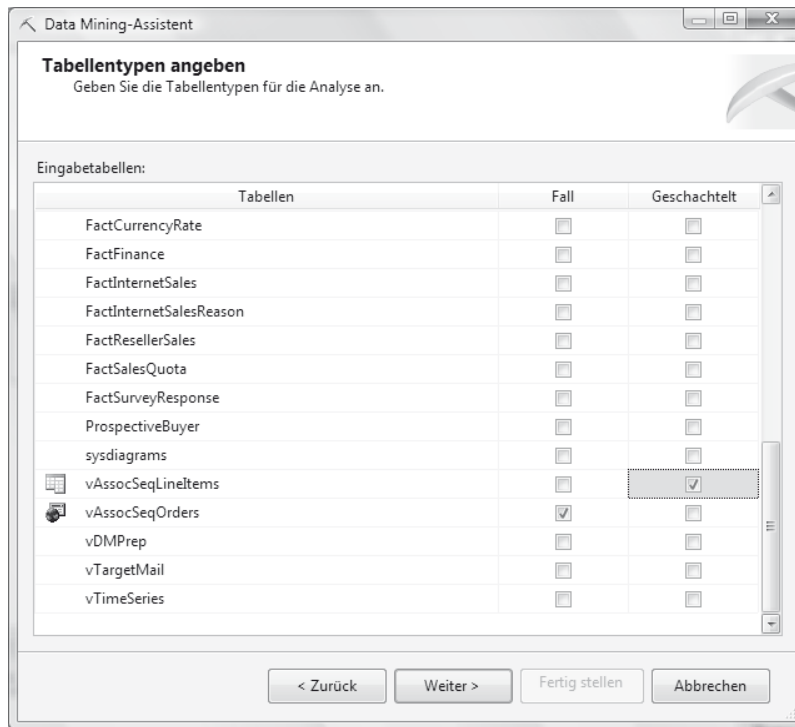


Abbildung 5.7 Auswahl der Tabellentypen

Klicken auf **Weiter** führt zum nächsten Schritt. Nun erfolgt die Auswahl der Trainingsdaten für das Modell. Nochmals zur Erinnerung: Es geht darum, Artikel zu gruppieren. Die Art des Artikels ist in der Spalte „Modell“ enthalten. Deshalb muss diese Spalte zum einen als Schlüssel definiert werden, zum anderen aber auch als Eingabe und Vorhersage. Für das Mining-Modell muss noch die Brücke zwischen den beiden Tabellen geschlagen werden, und das erfolgt über die ganz am Anfang definierte Relation der jeweiligen Spalten „OrderNumber“. Diese Spalte wird auch als Schlüssel verwendet.

Das Auswahlfenster sollte demnach wie folgt aussehen:

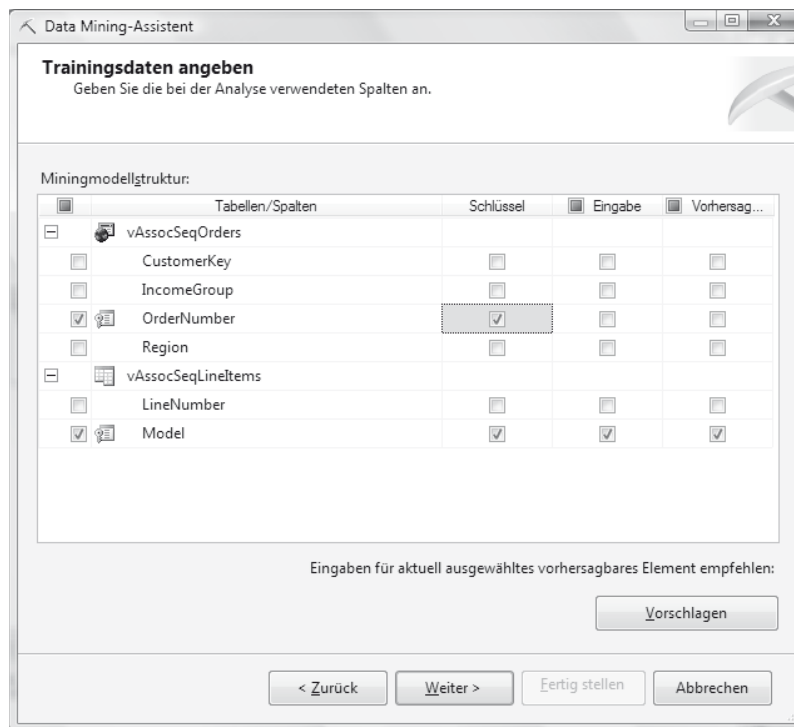


Abbildung 5.8 Auswahl der Trainingsdaten für das Mining-Modell

Über einen Klick auf **Weiter** kommt man auf den nächsten Schritt. Es geht nun darum, den Inhalt und den Datentyp der Spalten anzugeben. Im vorliegenden Fall ist diese Definitionsseite sehr übersichtlich. Für umfangreichere Modelle gibt es hier einen Erkennungs-Assistenten. Hier werden beide Spalten vom Typ „Key“ mit dem Inhalt „Text“ definiert.

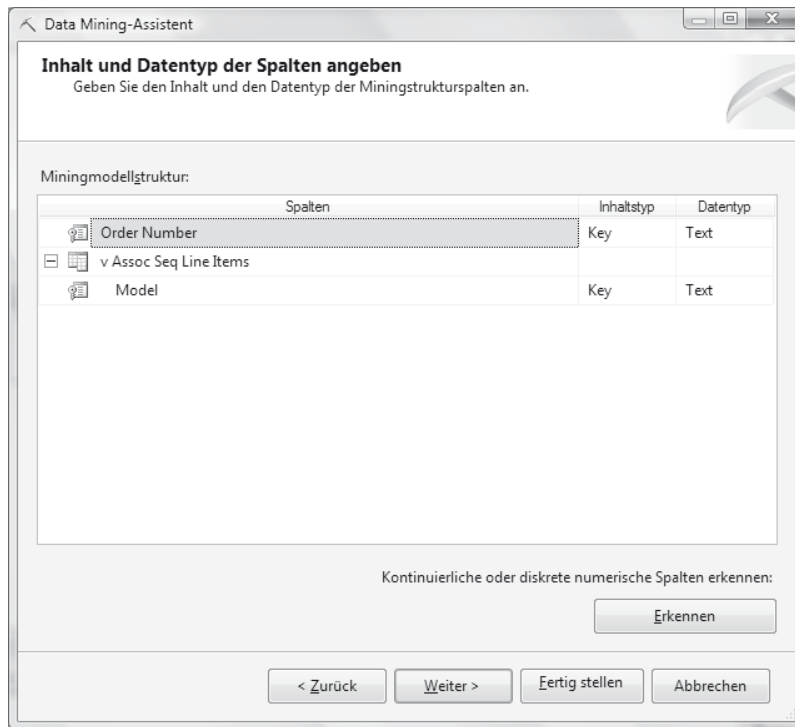


Abbildung 5.9 Definition von Inhalt und Datentyp der Spalten

Weiter geht es nach dem Klick auf **Weiter** mit der Definition der Testdaten. Hier nochmals kurz zur Erinnerung: Dies ist einer der wesentlichen neuen Schritte der Data-Mining-Assistenten. Es gibt die nun recht komfortable Möglichkeit, die Menge der Testfälle zu bestimmen, die für das Training der Modelle verwendet werden. Vorgabe ist dabei der Standardanteil von 30 % der Datensätze, die für das Modelltraining verwendet werden. Um das Beispiel möglichst geradlinig durchführen zu können, wird an der Stelle die Vorgabe übernommen.

The screenshot shows a dialog box titled 'Data Mining-Assistent' with the sub-header 'Testsatz erstellen'. Below the sub-header is the instruction: 'Geben Sie die Anzahl der Fälle an, die für das Testen des Modells reserviert wurden.' The dialog contains two input fields: 'Prozentsatz der zu testenden Daten:' with a value of '80' and a '%' symbol, and 'Maximale Anzahl der Fälle im Testdataset:' with an empty field. Below these fields is a 'Beschreibung:' section containing explanatory text. At the bottom, there are four buttons: '< Zurück', 'Weiter >', 'Fertig stellen', and 'Abbrechen'.

**Testsatz erstellen**  
Geben Sie die Anzahl der Fälle an, die für das Testen des Modells reserviert wurden.

Prozentsatz der zu testenden Daten: 80 %

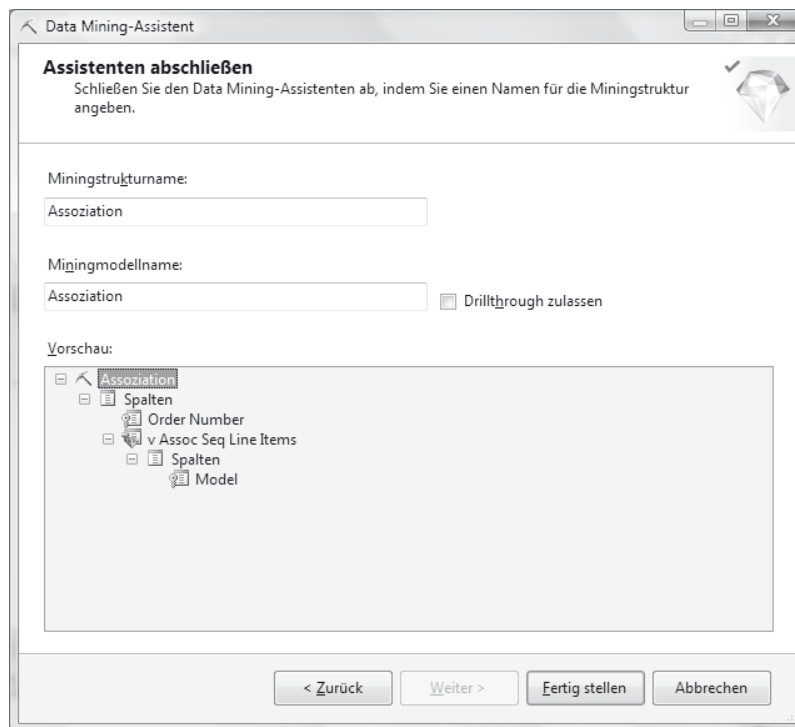
Maximale Anzahl der Fälle im Testdataset:

**Beschreibung:**  
Die Eingabedaten werden zufällig in zwei Sätze aufgeteilt, ein Trainingsatz und ein Testsatz, basierend auf dem festgelegten Prozentsatz der zu testenden Daten und der festgelegten maximalen Anzahl der Fälle im Testdataset. Der Trainingsatz wird zum Erstellen des Miningmodells eingesetzt. Der Testsatz wird zum Prüfen der Modellgenauigkeit verwendet.  
[Prozentsatz der zu testenden Daten] gibt den Prozentsatz der Fälle an, die für den Testsatz reserviert wurden.  
[Maximale Anzahl der Fälle im Testdataset] begrenzt die Gesamtzahl von Fällen im Testsatz. Wenn beide Werte angegeben wurden, werden beide Einschränkungen erzwungen.

< Zurück    Weiter >    Fertig stellen    Abbrechen

**Abbildung 5.10** Auswahl der Menge der Testdaten für das Modelltraining

Ein letzter Klick auf **Weiter** führt zum Abschlussbildschirm des Data-Mining-Assistenten. Es muss noch ein Name für die Mining-Struktur und das Modell angegeben werden. Hier wird für beide Objekte der Name „Assoziation“ verwendet.



**Abbildung 5.11** Abschluss des Data-Mining-Assistenten

Ein Klick auf **Fertig stellen** schließt die Erstellung der neuen Mining-Struktur ab und stellt diese im Projektmappen-Explorer zur weiteren Verwendung zur Verfügung.

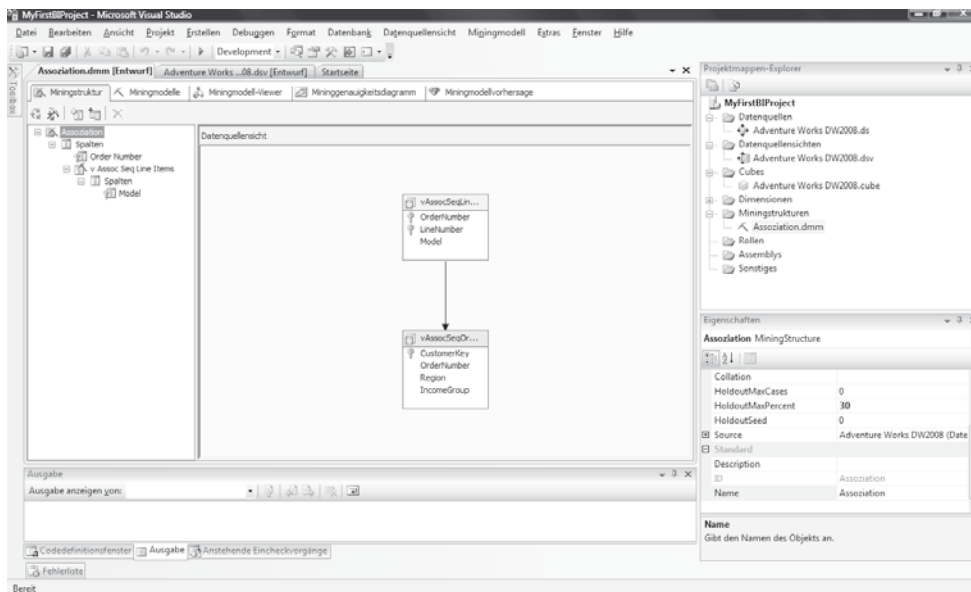


Abbildung 5.12 Die neu erstellte Mining-Struktur im Projektmappen-Explorer

### 5.1.2 Darstellung und Interpretation der Mining-Struktur

Bevor das Modell dargestellt werden kann, muss es verarbeitet werden. Dazu muss zum einen das Projekt aktualisiert auf dem Server bereitgestellt werden, und zum Zweiten muss das Mining-Modell mit den ausgewählten Testdaten verarbeitet werden. Dies erfolgt über den Aufruf der Verarbeitung im Kontextmenü des jeweiligen Objekts (Projekt bzw. Mining-Struktur).

Wenn eine Mining-Struktur neu erstellt ist und das erste Mal verarbeitet werden soll, kann dieser Verarbeitungsprozess auch dadurch gestartet werden, dass direkt der Reiter „Mining-Modell-Viewer“ im Hauptfenster ausgewählt wird. Dadurch werden die einzelnen Verarbeitungsschritte für das Projekt und das Mining-Modell gestartet.

Nach der Verarbeitung wird das Modell im Mining-Modell-Viewer angezeigt.