

chen und die Merkmale, die bei einer Vielzahl von Kunden übereinstimmen, als wesentliche Merkmale für einen potenziellen Neukunden herangezogen. Jeder wird an dieser Stelle übereinstimmen, dass die Frage nach den Käufermerkmalen für jedes Unternehmen sehr interessant ist, aber alles andere als einfach beantwortbar.

Im zweiten Schritt wird anhand von Data-Mining-Techniken ein Regelwerk aufgestellt, das basierend auf den Analysen der bestehenden Kunden einen Käufer ausmacht. Dazu wird in diesem Beispiel ein Entscheidungsbaum verwendet. Dieser verdeutlicht das Regelwerk am bildhaftesten.

Der dritte Schritt analysiert die Interessenten, die mit der Marketing-Aktion angesprochen werden sollen. Das erstellte Regelwerk wird auf die Interessenten angewendet. Dies liefert eine Einschätzung zu jedem Interessenten, ob dieser ein potenzieller Käufer ist oder nicht.

Abschließend wird dann noch eine Liste generiert, die den Namen, die Telefonnummer und eventuell die E-Mail-Adresse des Interessenten enthält und die Priorisierung, d.h. die Einschätzung, ob dieser tendenziell eher ein Käufer ist oder nicht. Diese Telefonliste, die dann einem Telemarketing-Unternehmen zur Verfügung gestellt werden kann, ist der eigentliche Grund, diese ganzen Analysen durchzuführen. Angenommen, diese Liste enthält 10.000 Interessenten, von denen 4.000 aufgrund der Analyse als potenzielle Käufer eingestuft werden. Die Telefonaktion wird dann mit diesen 4.000 Interessenten begonnen. Die Chance, frühzeitig ein definiertes Absatzziel zu erreichen, indem diese vorqualifizierten Interessenten kontaktiert werden, ist ungleich höher als ein willkürliches Abtelefonieren beispielsweise in alphabetischer Form. Im Idealfall führt das dazu, dass vielleicht die letzten 6.000 Interessenten gar nicht kontaktiert werden müssen, sondern über einen anderen Kanal adressiert werden können, der kostengünstiger ist. Wenn man einen Telemarketing-Anruf ganz vorsichtig mit 5€ Kosten pro Anruf ansetzt, spart diese Vorgehensweise auf einmal 30.000 € Marketing-Budget! Wenn das kein Grund ist, sich mit Data Mining zu beschäftigen!

4.6.1 Erstellen des Data-Mining-Modells

Das Erstellen des Data-Mining-Modells ist der erste wichtige Schritt zur anvisierten Telefonliste. Wie in den Grundlagen diskutiert, müssen vorab einige grundlegende Fragen geklärt werden. Dies soll exemplarisch in den weiterführenden Beispielen angesprochen werden. An dieser Stelle gehen wir davon aus, dass die Data-Mining-Methode des Entscheidungsbaums die am besten geeignete ist. Dieser kann nun erstellt werden.

Die Erstellung erfolgt über einen Assistenten, der über das Kontextmenü der Mining-Struktur im Projektmappen-Explorer gestartet wird.

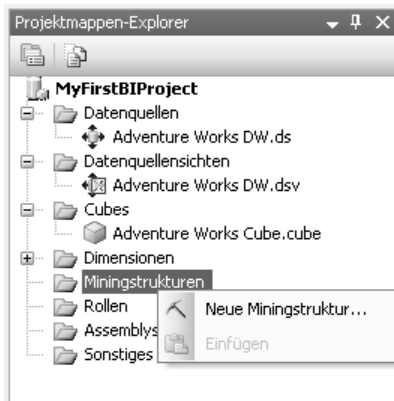


Abbildung 4.45 Start des Data Mining-Assistenten

Nach Bestätigung des Willkommens-Bildschirms durch **Weiter** erfolgt die Auswahl der Definitionsmethode. An dieser Stelle sieht man, dass, wie in den Grundlagen diskutiert, die Verwendung eines Cubes nicht zwingend notwendig ist, um Data Mining durchführen zu können. Dennoch, Cubes bieten Vorteile, und wenn schon einer definiert ist, sollte dieser auch verwendet werden. Dies wird durch die Auswahl des zweiten Optionsfelds „Aus vorhandenem Cube“ eingeleitet.



Abbildung 4.46 Auswahl der Definitionsmethode für die Mining-Struktur

Eine Bestätigung mit **Weiter** führt zur Auswahl der Data-Mining-Technik. Im Vorfeld wurde entschieden, den Entscheidungsbaum zu verwenden. Dieser wird nun ausgewählt („Microsoft Entscheidungsstrukturen⁶“) und mit **Weiter** bestätigt.

Im nächsten Schritt muss die Quellcubedimension ausgewählt werden. Da eine Analyse der bestehenden Kunden durchgeführt werden soll, wird hier die Dimension „v Target Mail“ ausgewählt.

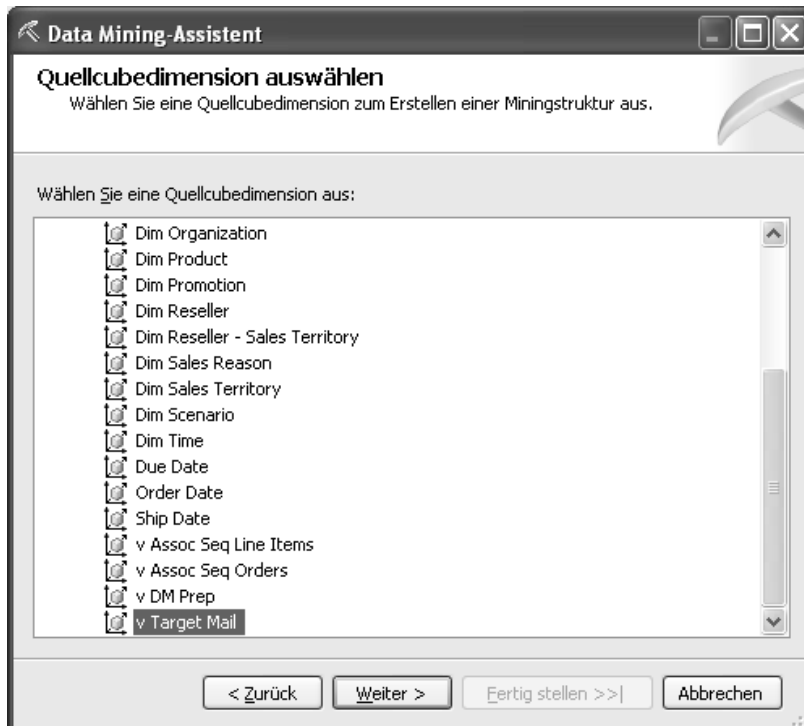


Abbildung 4.47 Auswahl der Quellcubedimension

⁶ Dieser Begriff ist leider sehr unglücklich gewählt. Hier verbirgt sich in der Tat die Microsoft-Implementierung eines Entscheidungsbaums.

Nun muss der entsprechende Fallschlüssel ausgewählt werden. Der Fallschlüssel ist das Attribut, das als Schlüsselattribut für das Mining-Modell verwendet werden soll. Dies ist in diesem Fall „v Target Mail“.

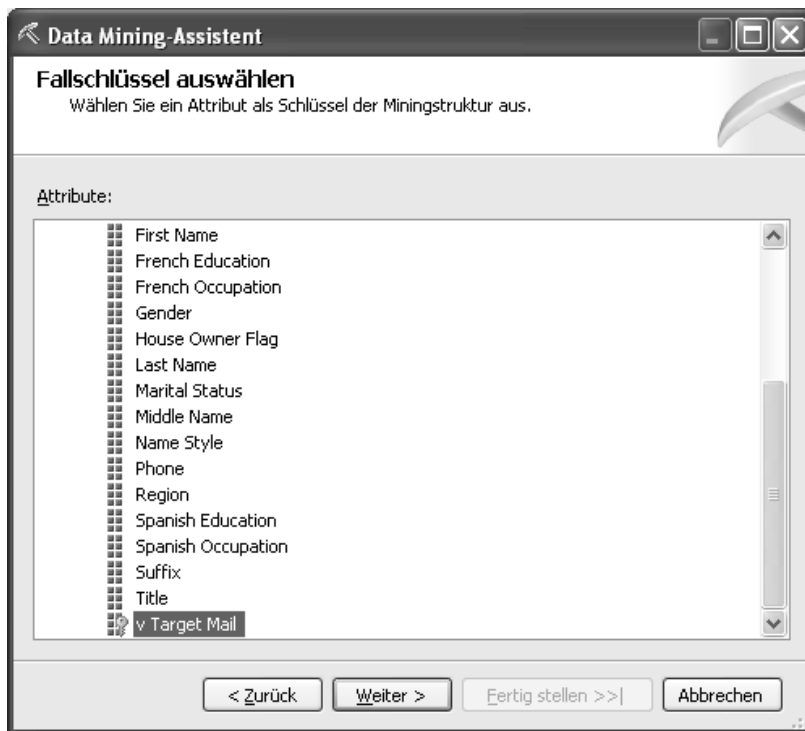


Abbildung 4.48 Auswahl des Fallschlüssels

Eine Bestätigung mit **Weiter** führt zur Auswahl der Attribute und Measures für das Mining-Modell. Hier wird ausgewählt, von welchen Faktoren es wohl abhängen wird, ob jemand als Käufer einzustufen ist oder nicht. An der Stelle gehören auf jeden Fall ein wenig Intuition und genaue Kenntnisse des Geschäftsmodells in die Waagschale. Grundsätzlich könnte man alle Attribute auswählen, aber das würde die Rechenzeiten für die Analyse extrem erhöhen. Von daher wird hier auf folgende Attribute beschränkt:

- *CommuteDistance*
- *Gender*
- *HouseOwnerFlag*
- *MaritalStatus*
- *Age - v Target Mail*
- *Bike Buyer*
- *GeographyKey*

- *NumberCarsOwned*
- *NumberCildrenAtHome*
- *TotalChildren*
- *YearlyIncome*

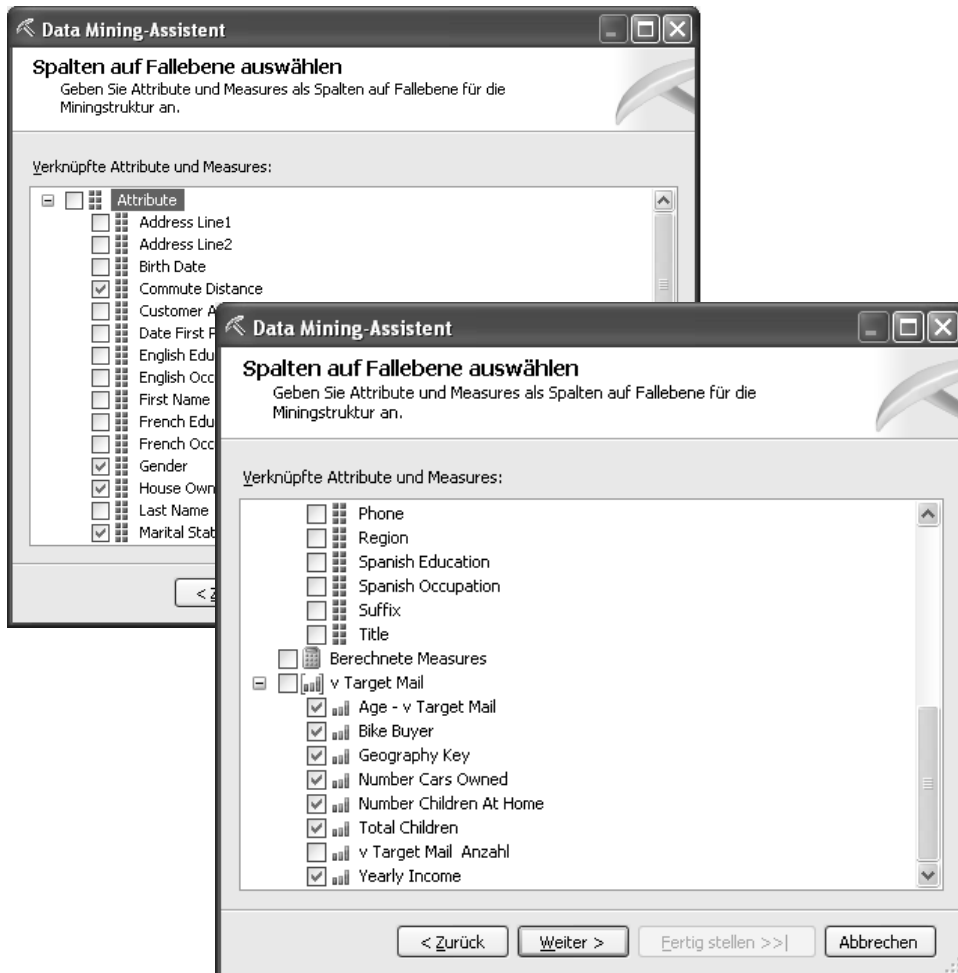


Abbildung 4.49 Auswahl der Attribute und Measures für das Mining-Modell

Die Bestätigung der Auswahl mit **Weiter** führt zum nächsten Schritt des Assistenten. Nun muss für die vorhin ausgewählten Attribute und Measures definiert werden, welche als Eingabeinformationen verwendet werden und für welche eine Vorhersage durch das Modell erarbeitet werden soll. Es geht in diesem Beispiel darum, eine Einschätzung zu bekommen, ob ein Interessent als möglicher Kunde, in dem Fall als Fahrradkäufer, in Frage

kommt. Deshalb wird das Attribut „BikeBuyer“ für das Modell als „Vorhersagbar“ ausgewählt. Da in den vorliegenden Daten zwei Typen von Kundenadressen vorliegen, nämlich echte Kunden, die bereits Käufer sind, und Interessenten, die erst Käufer werden sollen, wird das Attribut „BikeBuyer“ auch noch als „Eingabe“ markiert. Damit werden auch die Adressen im Modell berücksichtigt, die bereits als Kunden registriert sind.

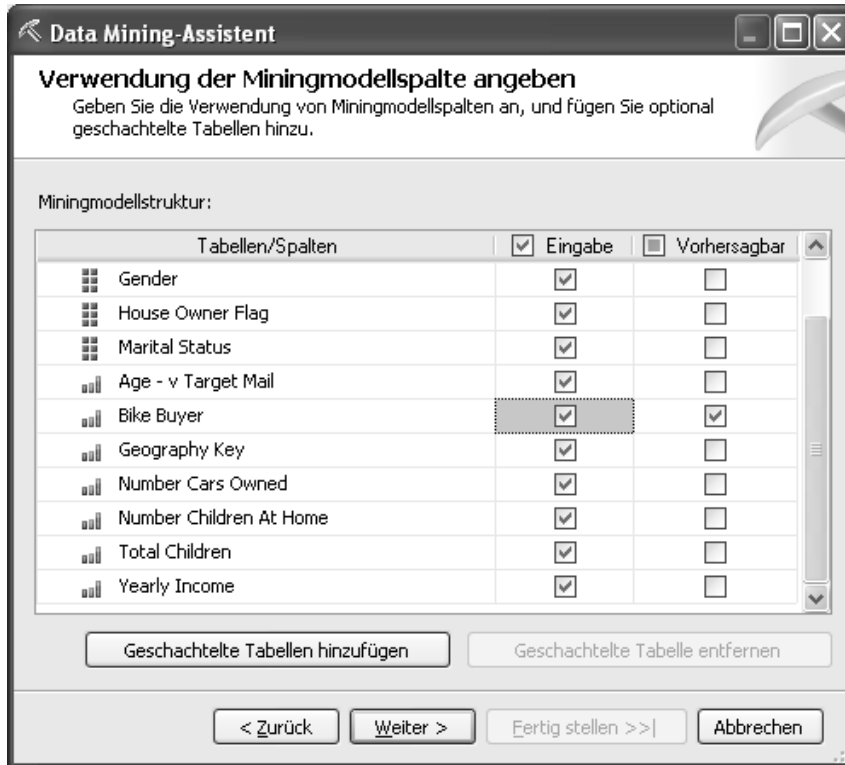


Abbildung 4.50 Verwendung der Merkmale im Mining-Modell

Durch Anklicken von **Weiter** wird nun vom Data-Mining-Assistenten versucht zu ermitteln, welche Datentypen die ausgewählten Merkmale sind. Es wird eine Tabelle erstellt, welche die ermittelte Zuordnung enthält. Diese Übersicht muss nun überprüft werden, ob diese so korrekt ist.

Die korrekte Zuordnung an der Stelle lautet wie folgt:

Tabelle 4.2 Inhalt und Datentyp der gewählten Spalten

| Spalte | Inhaltstyp | Datentyp |
|-------------------------|------------|----------|
| Commute Distance | Discrete | Text |
| Gender | Discrete | Text |
| House Owner Flag | Discrete | Text |
| Marital Status | Discrete | Text |
| Age - v Target Mail | Continuous | Long |
| Bike Buyer | Discrete | Long |
| Geography Key | Discrete | Text |
| Number Cars Owned | Discrete | Long |
| Number Children At Home | Discrete | Long |
| Total Children | Discrete | Long |
| Yearly Income | Continuous | Double |

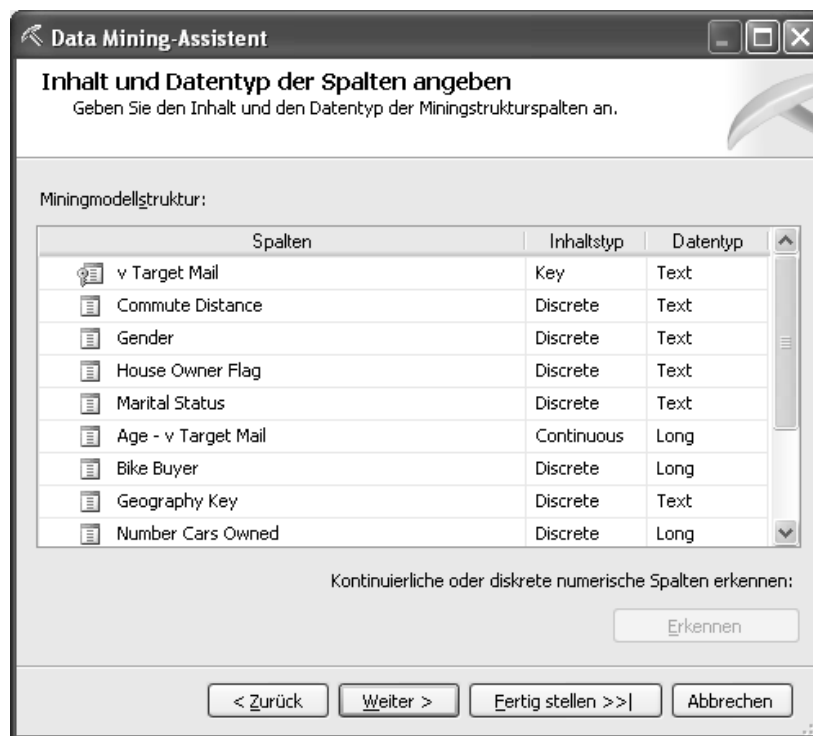


Abbildung 4.51 Inhalt und Datentyp der ausgewählten Spalten

Die Bestätigung der Auswahl erfolgt mit **Weiter**.

Nun erfolgt ein für das Data-Mining-Modell sehr wesentlicher Schritt. Es wird ja anhand von vorhandenen Daten ein Modell aufgebaut. Anhand der Daten wird analysiert, wie die Zusammenhänge zwischen den einzelnen Informationen sind und wie darauf eine Art Regelwerk konstruiert werden kann. Nun muss man, um so ein Modell zu erzeugen, Trainingsdaten zur Verfügung stellen. Diese Trainingsdaten sollen die Zieldaten möglichst gut umfassen, also in gewisser Weise repräsentativ sein, um das Modell verlässlich zu machen. Die Auswahl solcher Trainingsdaten ist alles andere als trivial, und es gibt hierüber eine Vielzahl von Abhandlungen, bei welcher Art von Daten welche Art von Trainingsdaten herangezogen werden sollen. Um die Betrachtung an dieser Stelle nicht zu sehr auszuweiten, soll im Wesentlichen beachtet werden, dass die Daten, die als Trainingsdaten für das Modell herangezogen werden, geeignet sein müssen, ein Regelwerk zu ermitteln, das möglichst generell auf den Anwendungsfall zutrifft, der dann mit der Anwendung realisiert werden soll.

In diesem Beispiel werden als Testdaten die Informationen verwendet, die über Käufer vorliegen, die aus Deutschland kommen. Die Auswahl erfolgt über einen entsprechenden Filter.

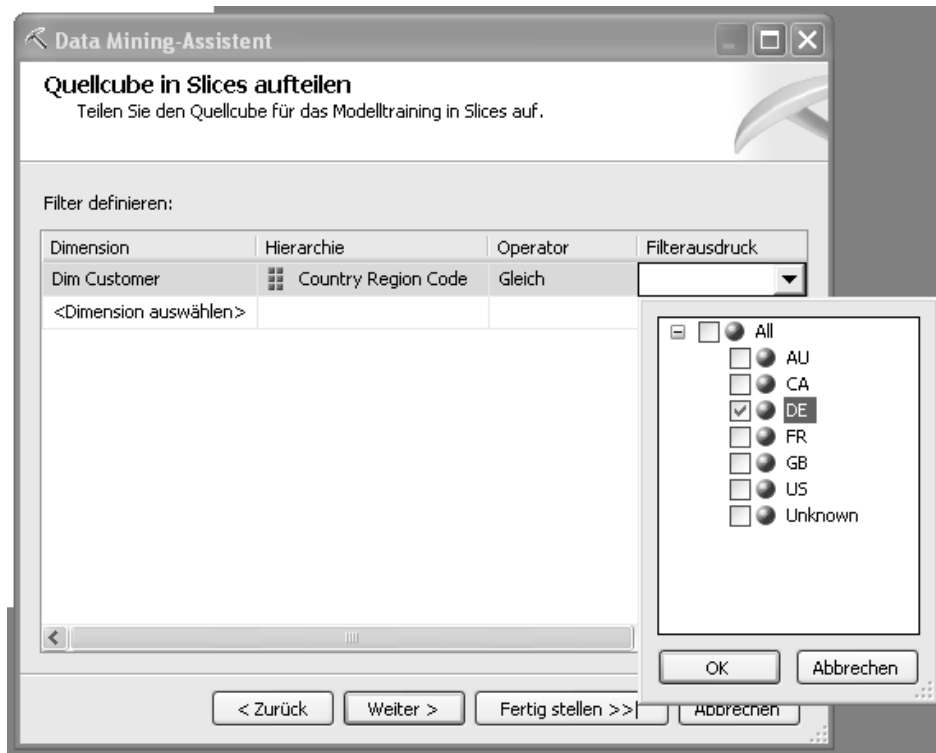


Abbildung 4.52 Auswahl der Cube-Slices für die Testdaten

Nun ist die Konfiguration nahezu abgeschlossen. Durch Klick auf **Weiter** öffnet sich der abschließende Bildschirm. Hier muss noch ein Name für die neue Mining-Struktur vergeben werden. Es wird der Name vorgeschlagen, der sich aus der Quellcube-Dimension ergibt. Um mit dem Modell noch etwas besser arbeiten zu können, wird an der Stelle noch der Haken „Drillthrough zulassen“ gesetzt.

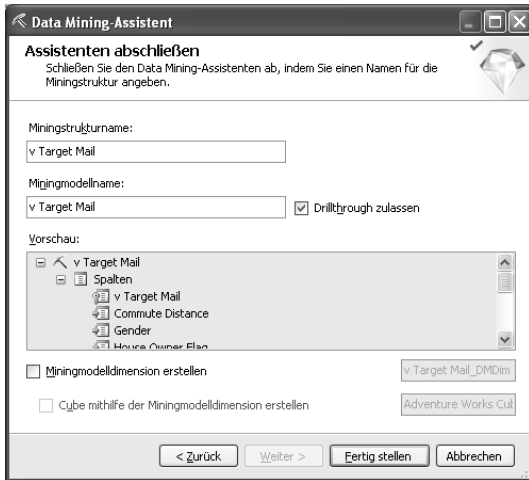


Abbildung 4.53 Abschlussbildschirm des Data-Mining-Assistenten

Der abschließende Klick auf „Fertig stellen“ erzeugt das soeben definierte Mining-Modell. Das Modell steht im Projektmappen-Explorer zur weiteren Verwendung zur Verfügung.

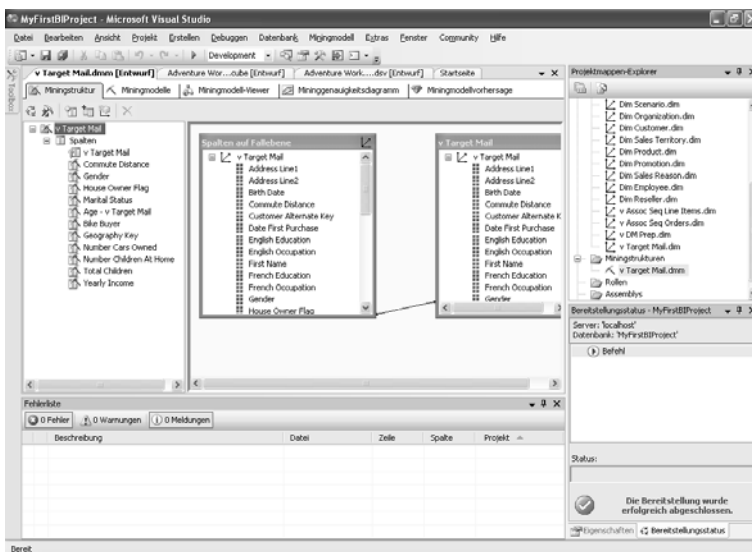


Abbildung 4.54 Neues Mining-Modell im Projektmappen-Explorer

4.6.2 Erstellen des Regelwerks

Mit dem vorliegenden Data-Mining-Modell besteht nun die Möglichkeit, ein Regelwerk zu ermitteln. Ziel der Analyse ist ja, Merkmale zu finden, die einen potenziellen Käufer beschreiben. Dazu wurde bisher eine Data-Mining-Methode ausgewählt, der Entscheidungsbaum. Dieses Regelwerk wird erzeugt, indem das Data-Mining-Modell auf die vorliegenden Daten angewendet wird. Es wird in dem Datenbestand ermittelt, welche Merkmale vorliegen, wie sie ausgeprägt sind und welche Zusammenhänge bestehen. Aus dieser Analyse wird ein Entscheidungsbaum aufgebaut, dessen Äste die unterschiedlichen Regelwerke beschreiben.

Um dieses Regelwerk zu erhalten, wird nun die Mining-Struktur auf den Datenbestand angewendet. Dazu muss im ersten Schritt das aktuelle Projekt auf dem Microsoft Analysis Server bereitgestellt werden. Die Bereitstellung erfolgt wieder über das Kontextmenü des Projektes im Projektmappen-Explorer.

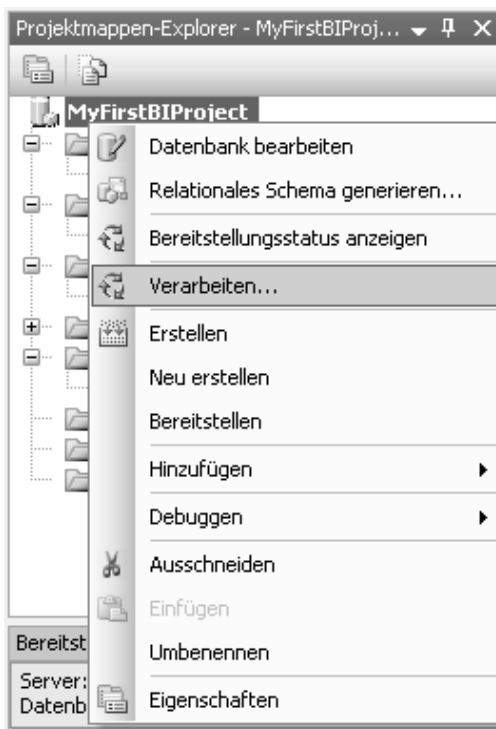


Abbildung 4.55 Bereitstellen des aktuellen Projektes