Intelligent Miner

Using the Intelligent Miner in a Simple Scenario

Version 2.1.1

Note

Before using this information and the product it supports, be sure to read the general information under “Notices” on page 33.

First Edition (April 1998)

This edition applies to Version 2 Release 1 Modification Level 1 of the IBM Intelligent Miner for Data, 5697–IM2, and to any subsequent releases until otherwise indicated in new editions or technical newsletters. Make sure you are using the correct edition for the level of the product.

Order publications through your IBM representative or the IBM branch office serving your locality.

When you send information to IBM, you grant IBM a nonexclusive right to use or distribute the information in any way it believes appropriate without incurring any obligation to you.

© Copyright International Business Machines Corporation 1998. All rights reserved.
Note to U.S. Government Users — Documentation related to restricted rights — Use, duplication or disclosure is subject to restrictions set forth in GSA ADP Schedule Contract with IBM Corp.

Contents

© Copyright IBM Corp. 1998

About this book

This book contains a simple scenario that will help you learn how to use the wizards (sometimes referred to as TaskGuides) in the Intelligent Miner.

Who should read this book

Any Intelligent Miner user who wants to gain a better understanding of the tasks involved in data mining should read this book. The business problem is provided to give you a framework for learning and is not meant to imply that the methods described are the most appropriate way to solve the problem presented. As you work with the Intelligent Miner and develop an understanding of the mining functions, you will find the methods that best suit your business needs.

We'd like your comments

You feedback is important to us in providing you with accurate and high-quality information. If you have any comments about this book or any other Intelligent Miner documentation or product, please visit our Web site at:

`http://www.software.ibm.com/data/iminer`

There you will find a feedback page where you can enter comments and send them to us.

© Copyright IBM Corp. 1998

Chapter 1. Introduction

This Intelligent Miner scenario consists of a simple mining run comprising several tasks. The scenario starts with raw data in a flat file, details the process of defining Intelligent Miner data objects, creating objects, running Intelligent Miner functions, and viewing results using the Intelligent Miner's visualizers. At the end of this scenario, you will know how to use the Intelligent Miner main window and wizards.

One of the first tasks for new users of Intelligent Miner is to learn how to use the wizards to connect to your data and to begin building objects. This Intelligent Miner scenario consists of an abbreviated data mining scenario with five phases: Defining data, building a model, applying the model, automating the process, and analyzing the results. By following the steps in this scenario, you will learn how to define data objects, run mining functions, and view results in the Intelligent Miner.

This scenario and the sample data used in this scenario are designed to support the learning objectives. As such, they do not represent actual or recommended methods for using the Intelligent Miner. To shorten the time it takes to complete the scenario, the data file is small and can be processed quickly. Additionally, the five phases represent an important subset of the activities at the core of many mining projects. Finally, the scenario uses the Clustering – Demographic function to accomplish its goals. There are other functions within the Intelligent Miner that can be used to accomplish the same end. Typical mining investigations would compare the results of the different functions.

Once you master the mechanics of using the Intelligent Miner main window and wizards, there is much to learn about the information content of your data and of the iterative nature of the data mining process. There are a number of books, some of which can be found at your local book store, that can be very helpful in understanding data mining at a conceptual level.

Before you start

Before you work with this scenario,

- Read ″Using the Intelligent Miner″ in Chapter 1 of the *Intelligent Miner Getting Started* book to understand how to use the Intelligent Miner main window
- Read the readme.txt file provided with the Intelligent Miner client installation

To use this scenario, you need:

- The Intelligent Miner server installed on AIX, OS/400, OS/390, or Sun Solaris
- The Intelligent Miner client installed on AIX, OS/2, or Windows NT or 95

The business problem

Imagine that you work for a bank that sells several products, including Regular Checking, Premier Checking, and Exclusive Checking accounts and option packages for each account. The bank already knows that Premier Checking is their most profitable product, and wants to increase the number of customers who have this type of checking account. The marketing department wants to identify different groups based on demographic data, such as age and income, within the Premier Checking customers so that the department can prepare different ad campaigns for each of the groups. Additionally, the department wants to identify customers who

© Copyright IBM Corp. 1998

are not currently Premier Checking customers who have similar demographics to the customers who are Premier Checking customers.

You have obtained some raw customer data from corporate headquarters to solve this business problem. This data is named banking.txt. It contains information about customers from all branches of the bank. You can use the Intelligent Miner to mine this data and provide demographic information to the marketing department. Your raw customer data includes information about customers who already have the Premier Checking account, so you can use the Clustering – Demographic mining function to find the similarities in the demographic data of customers who already have Premier Checking.

To learn more about demographic clustering, see a description of the Clustering – Demographic mining function in Chapter 2 of the *Intelligent Miner Getting Started* book.

The mining run tasks

This scenario will demonstrate five phases of data mining tasks:

Defining the data

Define a data object that points to a flat file containing your raw customer data file banking.txt. The data object will be named **Customers**. The Intelligent Miner data objects simply point to the location of your data, so that the Intelligent Miner can process this data. You will not actually be changing the contents of the banking.txt file. See “Chapter 2. Defining a data object” on page 7 for instructions on how to complete this step.

Building the model

Define a Clustering – Demographic mining function named **Build model**. This mining function uses the Customers data object as the input data. It will run in training mode, and produce a result object named **Model**. This model information that describes the clusters identified during the mining run. You will use this model in a subsequent step to find other customers who have similar characteristics to the Premier Checking customers. See “Chapter 3. Building a model” on page 11 for instructions on how to complete this step.

Applying the model

Define a Clustering – Demographic mining function named **Apply model**. This mining function uses the Customers data object as the input data. It will run in application mode using the Model result object, and produce an output data named **Scored customers** and a flat file named **scored.txt**. The output data contains information about all potential customers and how similar they are to the Premier Checking customers. See “Chapter 4. Applying the model” on page 19 for instructions on how to complete this step.

Automating the process

To automate the process you create a sequence object **Target Marketing** containing the Build model mining function and the Apply model mining function. A sequence is an object containing several other objects in a specific sequential order. You can run a sequence, which runs each of the objects within the sequence in the order that you specified. This allows you to combine several mining tasks into one step. See “Creating a sequence” on page 23 for instructions on how to complete this step.

Analyzing the results
Define a Bivariate Statistics function named **Analyze**. This statistical function analyzes the data object **Scored customers** and produces an output data object **Target customers**, a flat file **target.txt**, and a result object **Target customer demographics**. These objects contain information about the customers that have a high level of similarity to the Premier Checking customers. See “Chapter 5. Creating a statistics function” on page 27 and “Chapter 6. Interpreting the results” on page 31 for instructions on how to complete this step.

Starting the Intelligent Miner in demo mode

The Intelligent Miner includes a sample flat file named banking.txt that contains raw customer data. To access this sample data, you must start the Intelligent Miner in demo mode.

Running the Intelligent Miner in demo mode provides all the same functionality as running Intelligent Miner in regular mode, but also includes sample mining bases and sample data. The interface that you use to create objects, define data, and perform other data mining tasks for this scenario is the same interface that you use in regular mode.

To start the Intelligent Miner in demo mode, first determine if you are running in AIX standalone mode or if you are using remote access to an AIX server.

If you are running in AIX standalone mode: Start the Intelligent Miner server using these commands:

1. `cd /usr/lpp/IMiner/bin`
2. `./idmstart`
3. `./imdemo`

If you are using remote access to an AIX server running the Intelligent Miner in demo mode: Start the Intelligent Miner server using these commands:

1. `cd /usr/lpp/IMiner/bin`
2. `./idmstart`
3. `./idmstartdemo`

Then start the Intelligent Miner client as usual. All clients connecting to the server running the Intelligent Miner in demo mode will be able to share the sample mining bases. Each client must sign on to the Intelligent Miner server using the appropriate user ID and password on the Servers page of the Preferences notebook.

Restarting the Intelligent Miner in demo mode: To remove all additional mining bases and reset them to their initial state, issue the command `idmstop` and `idmstartdemo` again.

The Intelligent Miner main window

The Intelligent Miner main window helps you manage mining bases and perform data mining tasks. A mining base is a collection of the mining objects needed to conduct a mining run.

After starting the Intelligent Miner client on AIX or Windows,

1. Select **Preferences** from the **Options** menu. The Preferences notebook opens.
2. Click the **Save after creating** radio button under Autosave Mining Base.
3. Check the **After settings object is run, automatically visualize a result** check box.
4. Click **OK** to save your preferences.

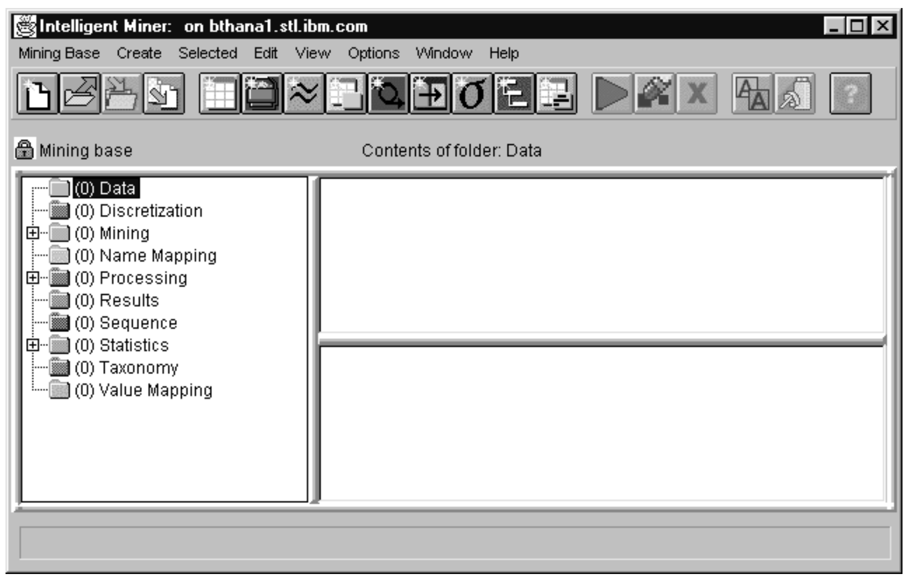

Figure 1. The Intelligent Miner main window

Figure 1 shows the Intelligent Miner main window, which has the following containers that help you work with Intelligent Miner objects:

Mining base container
A container consisting of folders for different mining objects that you create. It is located on the left side of the main window.

Contents container
A container that displays the contents of the folder that you select in the Mining base container. It is located on the top of the right side of the main window.

Work area container
A container that diplays the objects you dragged and dropped from the Contents container. It allows you to collect all objects that you are currently working with in a convenient area. It is located on the bottom of the right side of the main window.

When you select an object, you can click the right mouse button to get a menu of actions to perform on the selected object.

Using the Intelligent Miner help

While you are using the Intelligent Miner, you can use the Intelligent Miner help system to provide additional information that you might need to complete a task. The Intelligent Miner provides several different types of help:

Intelligent Miner wizards: To perform Intelligent Miner tasks, you use wizards that provide instructions on the left side for each page of a wizard, as shown in Figure 2 on page 8. These instructions help you complete the task for that page, and also provide definitions of terms that appear on the interface. To see a term definition, click on any highlighted term on a wizard page. The first page of a wizard is a Welcome page that provides an overview of the task you are performing.

Intelligent Miner online help: You can also access online help by clicking on the **Help** button on each page of a wizard. The online help provides concepts, examples, and valid values and actions for each page of an Intelligent Miner wizard. The online help comes up in a web browser, and you can navigate to the help information for different pages of the current wizard, or help information for other wizards.

Intelligent Miner help contents: You can also access the online help by clicking on the **Help** menu on the main menu bar.

Chapter 2. Defining a data object

The first step in this scenario is to define an Intelligent Miner data object that points to the raw customer data you want to mine. For this scenario, you are using data from a flat file banking.txt that resides on the Intelligent Miner server. To define a data object:

1. Click the **Create Data** icon on the toolbar to start the Data wizard from the Intelligent Miner main window.
2. From the Welcome page of the Data wizard, click **Next** to continue.

Specifying the data format and object name

On the **Data format and settings** page of the Data wizard, you must specify what kind of data you are using, and the name of the Intelligent Miner data object.

1. Select **Flat file** from the list. You can change the default selection in this list by selecting Options from the main menu.
2. Enter the name of the data object in the Settings name field: `Customers`. Optionally, you can also specify a comment associated with this data object, for example, you might type the comment: `Data about banking customers`.
3. Ensure that the **Show the advanced pages and controls check box** is not checked.
4. Click **Next** to continue.

Specifying the location of the data

On the Flat files page of the Data wizard, you specify the name and location of the data for this Intelligent Miner data object.

1. Double-click on the folder **dmtksample.nnnnn** in the list on the left, where nnnnn is a variable number. The contents of this folder are displayed in the list.
2. Double-click on the **Data** folder. The contents of this folder are displayed in the list on the right.
3. Scroll through the list and select **banking.txt**.

   If you do not see the banking.txt file in the list of available files, make sure that you provided the correct user ID and password for the Intelligent Miner server on the Servers page of the Preferences notebook.
4. Click **Add file**.
5. Make sure that the use mode **Read and write** is selected.
6. Click **Next** to continue.

Specifying the field parameters

On the Field Parameters page, you specify the begin and end positions, field names, field types, and optional name mappings for fields in this data object. The flat file display shows you a few lines of the flat file data, to help you enter the begin and end positions for each field. You do not have to specify the field parameters for all fields in the flat file, only the ones that you want to use for the mining run.

Table 1 on page 8 shows the field parameters for the flat file banking.txt. The begin and end positions are the numerical positions of the columns representing each

© Copyright IBM Corp. 1998

field in the flat file. For example, the Flat file display field inFigure 2 shows columns 1–6 contain values for the field gender, which is of the categorical field type.

Table 1. Field parameters for the Customers data object

Begin and end position	Field name	Field type	Name mapping
1-6	gender	Categorical	
10-16	age	Continuous	
25-26	siblings	Continuous	
30-36	income	Continuous	
38-44	type	Categorical	
45-45	product	Categorical	

To enter the field parameters:

1. Type `1–6` in the **Begin and end position** entry field.
2. Type `gender` in the **Field name** entry field.
3. Select **Categorical** in the **Field type** entry field.
4. Click **Add**.
5. Repeat the previous steps define the field parameters for this data object, as shown in Table 1.

   When you finish defining the field parameters, the window should look like Figure 2:

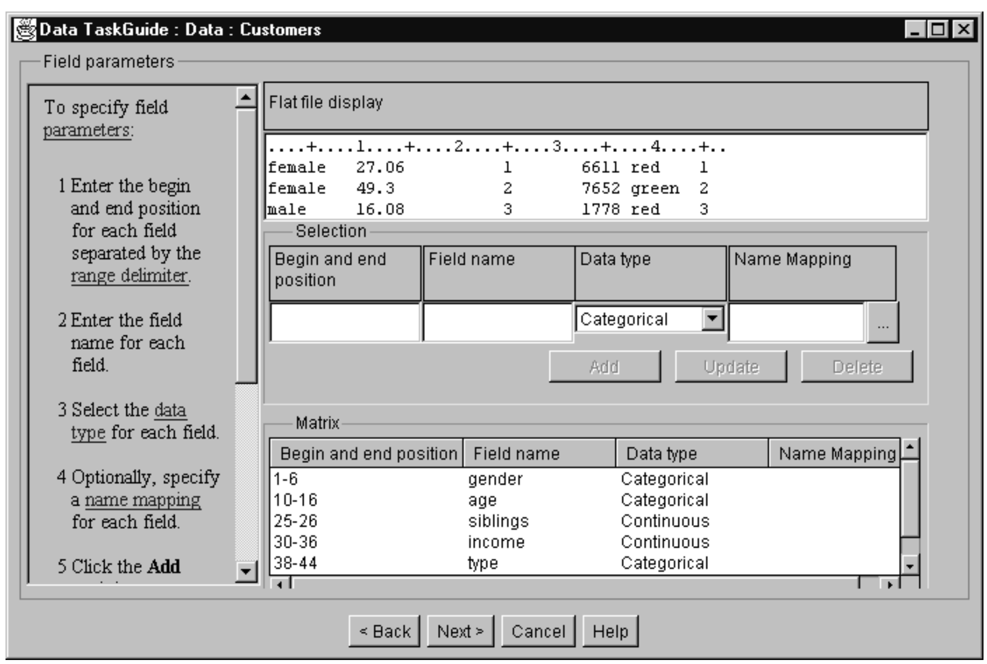

Figure 2. Field parameters page

Tip: You can use a name mapping to substitute a character string for a numeric code. For example, you might find it convenient to use a name mapping for the product field, which has values ranging from 1–8. You can map each numeric value to the name of the product, for example, product 1 is the Premier

Checking account, product 2 is the Regular Checking account, and so on. This scenario does not use name mappings for any fields in this data object.

6. Click **Next** to continue.

Defining computed fields

On the Computed fields page, you can define a computed field by using a mathematical expression, the fields in your data, and constant values to create a field value. For example, you might have a flat file containing values for your customers' year of birth. You might create a computed field named Age by subtracting the year of birth from the current year for each customer record. Because there is no need for such a computed field in this scenario, this scenario does not define any computed fields for this data object.

Click **Next** to continue.

Saving the data object

The next page of the Data wizard is the Summary page. The Summary page provides a summary of the parameters you defined for the current object you are creating. For all wizards in the Intelligent Miner, you can review the Summary page, and click **Finish** to create the object, or click **Back** to make changes to the existing parameters.

Click **Finish** to complete the wizard. If you receive an error message after clicking **Finish** in any wizard, you can use the **Back** button to return to any page and make corrections based on the message. At this point, you have defined the data object for your mining run. By selecting the Data folder in the mining base container, you can see the icon representing this data object.

Now that you have defined the first object for your mining run, you should save the mining base.

1. To save the mining base, click on the **Save mining base as** icon from the main window toolbar.
2. Enter `Target Marketing` as the mining base field. Optionally, enter a descriptive comment about the purpose of that mining base.
3. Click **Save** to save the mining base.

Chapter 3. Building a model

You can use the Clustering – Demographic mining function to obtain information about customers who already have the Premier Checking account. You run this function in training mode to produce a model as a result object.

This function generates clusters from your input data. These clusters are in the form of a result object that you can view using one of the Intelligent Miner visualizers. The model can then be used to create a score that measures how similar other customers are to the customers who already have the Premier Checking account.

To create the mining function for building the model:

1. Click on the **Create mining** button from the Intelligent Miner main window tool bar. The Intelligent Miner displays the Welcome page of the Mining wizard.
2. Click **Next** to continue.

Specifying the mining function and name

On the Mining functions and settings page, you select the type of mining function that you want to create and specify the name and an optional comment for this mining function:

1. Select Clustering – Demographic from the list of mining functions.
2. Type the settings name `Build Model`. Optionally, you can specify a more descriptive comment associated with this mining function.
3. Check the **Show the advanced pages and controls** check box.

   The advanced pages and controls of this wizard allow you to use additional options when defining your mining function. For example, using the advanced pages and controls, you can specify a filter condition for the records of the input data. The Intelligent Miner filters the records based on the condition you specify while it is running the mining function.
4. Click **Next** to continue.

Specifying the input data

On the input data page of the wizard, you specify the input data for this mining function:

1. From the list of available input data, select the **Customers** data object that you created in a previous step.
2. Select **Disk space** under **Optimize mining run for**.
3. Under the Advanced options, click the **...** button that is next to **Filter records condition**. You can filter the records of the input data for this mining function.

The Intelligent Miner displays the Expression Builder for filtering records.

For this scenario, you want to include all input data records for customers who have the Premier Checking account. Because the code for the Premier Checking account is 1, the filter condition you want to specify is product=1.

1. Click the **AND** push button. The expression builder creates a template for the expression, which displays as `((Arg1 = Arg2))`.
2. In the **Category** list, click on **Field Names**. The **Value** list displays all the available fields that you can include in this expression.

© Copyright IBM Corp. 1998

3. From the **Value** list, select the field **product**.
4. Click on the **Arg1** button. This sets the field product as the first argument in the expression.
5. Select **Constants** from the **Category** list.
6. Double-click on **<new constant>** in the **Value** list.
7. Type in the new constant value 1.
8. Press Enter. The new constant is added to the list of constants.
9. Select the constant **1** from the **Value** list.
10. Click the **Arg2** button. This sets the constant value of 1 as the second argument in the expression. The Expression Builder window looks like this:

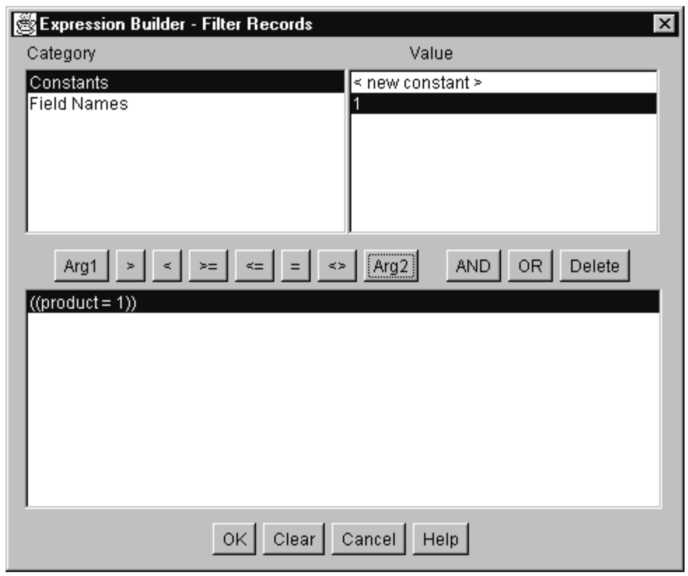

Figure 3. Expression Builder

11. Click **OK** to return to the Input Data page of the Mining wizard.
12. Click **Next** to continue.

Setting the mode parameters

On the mode parameters page of the wizard, you specify the mode parameters for the mining function. You can run the Clustering – Demographic mining function in two modes: clustering mode or application mode. For more information on the mode parameters, see the online help for the Clustering – Demographic mining function.

For this scenario, you will run this mining function in clustering mode.

Clustering mode: In clustering mode, the function creates groups of records called clusters. The function has parameters that you can use to control the results, including the maximum number of clusters, maximum number of passes, accuracy, and similarity threshold.

Use the default values for these parameters for this scenario. Refer to the *Intelligent Miner Getting Started* book and the online help for more information about these parameters.

Maximum number of passes
The default value is 2.

Maximum number of clusters
The default value is 9.

Accuracy
The default value is 2.

Similarity threshold
The default value is 0.5.

To set the mode parameters, ensure that the **Clustering mode** radio button is selected.

Click **Next** to continue.

Specifying the input fields

On the Input fields page of the Mining wizard, you specify the active fields and supplementary fields for this mining function.

The Clustering – Demographic mining function searches the input data for records with similarities, and places similar records into clusters. The active fields you specify are used to determine whether or not the input data records are similar. The supplementary fields you specify are included in the result, but are not used to determine whether or not the input data records are similar.

For example, one of the fields in the customer data is gender. If you specify gender as an active field, the Clustering – Demographic mining function uses this as a criteria in determining whether or not two customers are similar. You do not use gender as an active field because this will probably result in two large clusters, one consisting of male customers, and another consisting of female customers. Instead, you use demographic information like income, age, and siblings because that will generate more meaningful clusters in the results. To see how gender is distributed within the clusters, you add it as a supplementary field.

Because you have filtered out all customers who do not have the Premier Checking account, you do not want to use the field product as an active field in the clustering process.

1. Select **age** from the list of **Available fields**. Click the **>** push button to add it to the list of **Active fields**.
2. Select **siblings** from the list of **Available fields**. Click the **>** push button to add it to the list of **Active fields**.
3. Select **income** from the list of **Available fields**. Click the **>** push button to add it to the list of **Active fields**.
4. Select **type** from the list of **Available fields**. Click the **>** push button to add it to the list of **Active fields**.
5. Select **gender** from the list of **Available fields**. Click the **>** push button to add it to the list of **Supplementary fields**.

   If you want to add all the items in the **Available fields** list, you can use the **>>** push button.

   After you have specified the active fields and supplementary fields, the input fields page should look like Figure 4 on page 14:

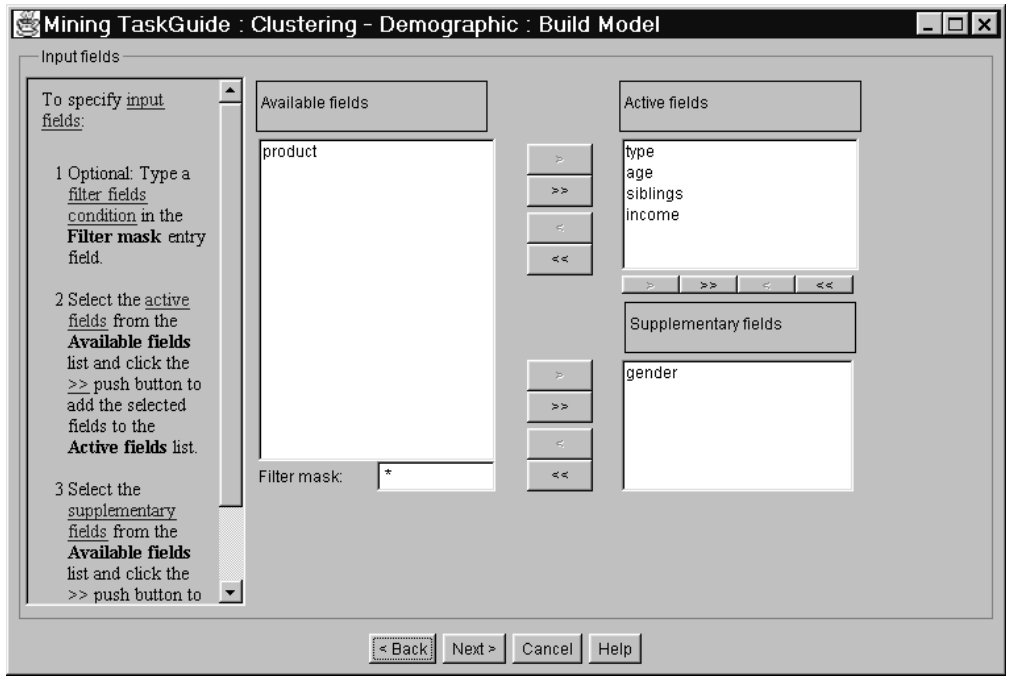

Figure 4. Input fields page

6. Click **Next** to continue.

Specifying advanced parameters

For each mining function, you can choose to see the advanced pages and controls. These advanced pages and controls allow you to specify advanced parameters for the mining function, such as field weights, value weights, compensation for active fields, distance factors for numeric fields, and outlier treatment. If you do not use the advanced pages and controls, the mining function uses the default values for these advanced parameters.

The next few pages of the wizard are advanced pages. For this scenario, you accept the default values. For more information about these parameters, see the online help for the Clustering – Demographic mining function.

1. On the Field parameters page of the wizard, click **Next** to continue.
2. On the Additional field parameters page of the wizard, click **Next** to continue.
3. On the Outlier treatment page of the wizard, click **Next** to continue.
4. On the Similarity matrix page of the wizard, click **Next** to continue.

Specifying other parameters

The next two pages of the wizard allow you to specify parallel parameters and whether or not you want the mining function to create output data.

You may be able to improve the performance of your mining runs by running the Intelligent Miner in parallel mode. By default, all mining functions are run in serial mode, on the Intelligent Miner server node. For this scenario, you can run this mining function in serial mode. For more information about the parallel parameters, see the online help for the Clustering – Demographic mining function.

You can specify whether you want the mining function to create output data. In addition to the fields in the input data, the output data also contains new fields generated by the mining function, including the cluster ID, record score, confidence. These fields provide information on the clusters generated by the mining function. For example, the cluster ID field tells you which cluster a particular customer record is in. Because you are using this mining function as a model, you are more concerned with the result object generated by the mining function than the output data. For this scenario, you do not create output data.

Depending on the options that you purchased with Intelligent Miner, your wizard may not contain a Parallel parameters page. (If your wizard does not contain a parallel parameters page, proceed to step 4 in the following procedure.)

1. From the Parallel parameters page of the wizard, ensure that the **Run the serial mode of the function** radio button is selected.
2. Click the **On the Intelligent Miner server node** radio button.
3. Click **Next**.
4. From the Output fields page of the wizard, ensure that the **Create output data** radio button is not selected.
5. Click **Next** to continue.

Specifying the result object name

Each time that you run this mining function, it creates a result object and saves it with the name you that specify on this page. Because the mining process is an iterative one, you will probably run a mining function more than once. You can choose to allow this mining function to replace a result object with the same name, because an existing result object with the same name may have been generated by a previous iteration of this mining function.

To specify the result object name:

1. Type `Model` as the name of the result object. Optionally, specify a comment associated with this result object.
2. Check the **If a result with this name exists, overwrite it** check box.
3. Click **Next**.
4. On the Summary page of the Mining wizard, review the parameters for the mining function that you are defining. Select the **Run this settings immediately** check box.

   If you do not want to run the current mining function immediately, you can run it later from the Intelligent Miner main window. For this scenario, run the mining function now.
5. Click **Finish** to complete this task.

After you complete each step in the mining wizard, the Intelligent Miner runs the mining function and displays a progress indicator that allows you to monitor the status of the mining function. After successfully running the mining function, the Intelligent Miner displays the result object generated by this mining function.

This mining function generates a result object named Model, which describes clusters of customers who have the Premier Checking account. Each cluster contains customers with similar characteristics for the fields income, age, type, and siblings.

Whether the Intelligent Miner displays the result object immediately after generating it depends on a preference setting on the Miscellaneous page of the Preferences notebook. You can specify whether you want to visualize results immediately after they are generated. If the Intelligent Miner does not display the result object immediately, you can view the result object by double-clicking on the result object in the Results folder from the Intelligent Miner main window.

Interpreting the results generated

The results generated by the mining function are shown in Figure 5. The multiple rows of graphs are designed to give you an understanding of your clustering results at a high level.

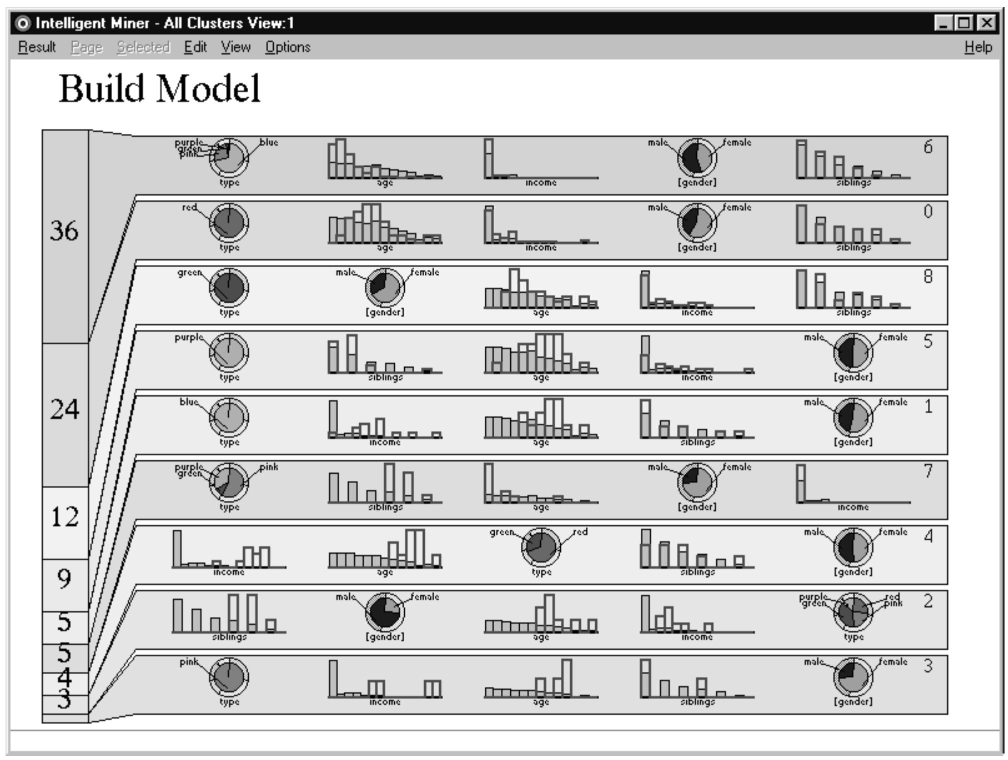

Figure 5. Results of the build model mining function

The display shows nine rows of graphs each representing one of the nine clusters identified by the mining run. Within each cluster, the pie charts and the bar charts represent active and supplemental fields used in the cluster, ordered by influence. Fields that had the greatest influence on forming the cluster are displayed on the left, while fields with the least influence are displayed on the right. The numbers down the left side represent a cluster as a percentage of the whole; for example, the top cluster represents 36%, the next lower cluster represents 24 percent, and so on. The numbers down the right side identify the cluster ID number.

The top row is the cluster with the largest number of customers. The cluster accounts for 36% of the customers. Each bar or pie chart shows the distribution of the field for the cluster and for the entire set of Premier Checking (product 1) customers. Supplementary fields are indicated with square brackets around the field names.

You can display more detail about a cluster by double-clicking on the row. Figure 6 shows the display of the top cluster, the largest cluster. This cluster contains the account type, age, gender, and siblings fields. Assume that account type indicates the options package the customer purchased, specified at the bank by colors. Double-clicking on any of the graphs shows a single graph.

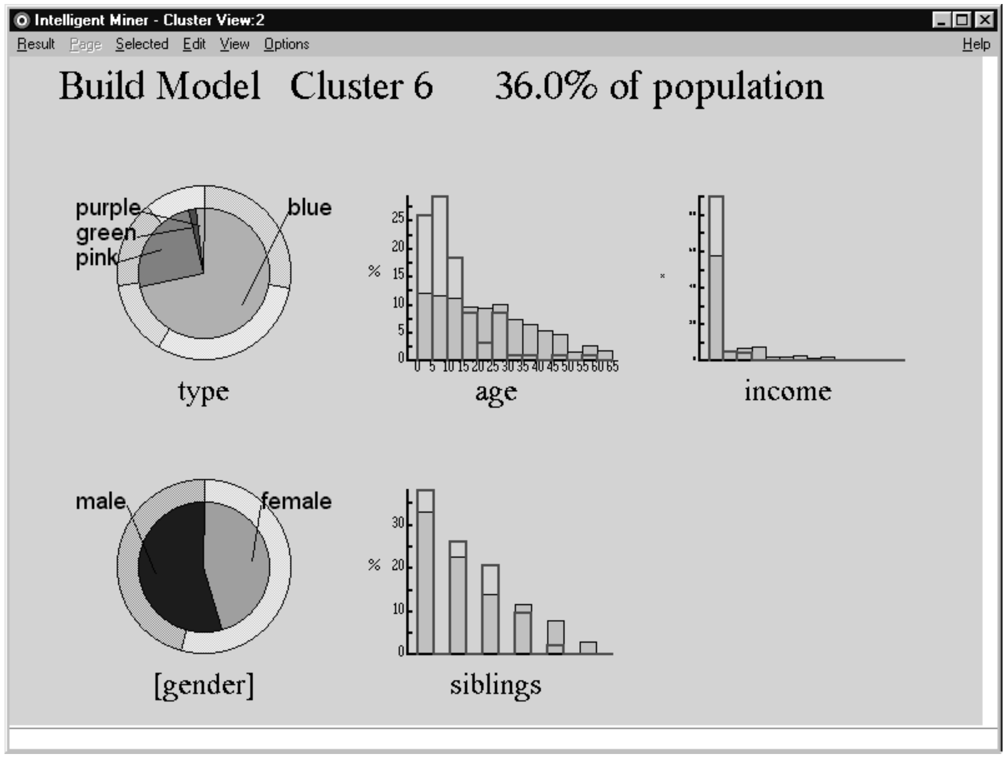

Figure 6. Top cluster of result object

Figure 6 includes a pie chart for gender. Each pie chart produced by the Intelligent Miner shows two distributions. The outside ring shows the distribution for the entire sample. The inside ring shows the distribution for the associated cluster. For example, Figure 6 show a pie chart in which the outside ring represents the distribution of male and female customers for all Premier Checking customers; the inside ring represents the distribution of male and female customers in this cluster. The chart indicates that this cluster has a greater percentage of males in it than the group of Premier Checking customers as a whole. Looking back at Figure 5 on page 16 you can quickly see that cluster 2 is mostly male and cluster 7 is mostly female.

Figure 7 on page 18 shows the distribution of age for the first cluster. The solid bars represent the percentage of people at each age in cluster 6 and the transparent bars represent the distribution for all Premier Checking customers. As shown by the graph, a higher percentage of the customers represented by cluster 6 are below the age of 15 when compared to the entire population of Premier Checking customers.

Figure 7. Age information from the top cluster

Viewing the clustering results can provide insights about the characteristics of the Premier Checking customers, even though the statistical details about the clusters may not be obvious to you when looking at the clusters using the visualizer. The results object contains the detailed statistical information, which you use as a model to apply to a new set of data in a subsequent step of this scenario.

Chapter 4. Applying the model

The next step in the scenario is to apply the clustering model created in the previous step. The result object named Model contains descriptions of nine clusters of customers who already have the Premier Checking account. In this step of the mining run, you will use the result object Model to find the customers who have characteristics similar to the customers who already have the Premier Checking account.

You will create a mining function apply model to create an analysis of customers who do not have Premier Checking. Then you will use the Intelligent Miner to score similarities between the Premier Checking customers and the non-Premier Checking customers.

To create a mining function for applying the model:

1. Click on the **Create mining** button from the Intelligent Miner main window toolbar. The Intelligent Miner displays the Welcome page of the Mining wizard.
2. Click **Next** to continue.

Specifying the mining function and name

For this step in the scenario, you again run a Clustering – Demographic mining function to apply the model created previously.

1. Select **Clustering – Demographic** from the list of mining functions.
2. Type the settings name `Apply Model`. Optionally, specify a more descriptive comment associated with this mining function.
3. Check the **Show the advanced pages and controls** check box.
4. Click **Next** to continue.

Specifying the input data

For this step in the scenario, you use the same customer data used by the previous mining function. In most situations, you will build a model using one set of data, and apply the model to a different set of data. The flat file used in this scenario contains customers with Premier Checking accounts and customers without Premier Checking accounts. In this case, you use the Intelligent Miner filtering feature to separate the customer records into two sets.

1. From the list of available input data, select the **Customers** data object.
2. Ensure that you are optimizing the mining run for disk space.
3. Under the Advanced options, click the **...** button next to **Filter records condition**. You will be filtering the records of the input data for this statistics function.

   The Intelligent Miner displays the Expression Builder for filtering records.

   In this step, you use the filter to select only the customers who do not have premier checking. The expression looks like this: `((product<>1)`. To create this expression:

   a. Click the **AND** push button. The expression builder creates a template for the expression, which displays as `((Arg1 = Arg2))`.
   b. Click on **Field Names** in the **Category** list. The **Value** list displays all the available fields that you can include in this expression.

© Copyright IBM Corp. 1998

c. From the **Value** list, select the field **product**.
d. Click the **Arg1** button. This sets the field product as the first argument in the expression.
e. Click the **<>** button. This sets the operand as ″not equal to.″
f. Select **Constants** from the **Category** list.
g. Select the constant `1` from the **Value** list.
h. Click on the **Arg2** button. The constant value of 1 is set as the second argument in the expression. The expression you defined looks like this: `((product<>1))`.

4. Click **Next** to continue.

Setting the mode parameters

On the Mode parameters page of the wizard, you specify the mode parameters for the mining function. You can run this mining function in clustering mode or application mode. In a previous step, you ran the Clustering – Demographic mining function in clustering mode to create a model. In this step, you apply the model to the customer data by running this mining function in application mode.

In application mode, the mining function scores how similar each customer is to the nine clusters of customers with Premier Checking.

For more information on the mode parameters, see the online help for the Clustering – Demographic mining function.

To set the mode parameters:

1. Click the **Application mode** radio button.
2. Select the result object **Model** from the application mode container under the **Application mode** group box.
3. Click **Next** to continue.

Specifying the input fields

In this step, you score customers using the same fields that were used to define the clusters in the Build model.

1. Select the field **age** from the list of **Available fields**. Click the **>** push button to add it to the list of **Active fields**.
2. Select the field **siblings** from the list of **Available fields**. Click the **>** push button to add it to the list of **Active fields**.
3. Select the field **income** from the list of **Available fields**. Click the **>** push button to add it to the list of **Active fields**.
4. Select the field **type** from the list of **Available fields**. Click the **>** push button to add it to the list of **Active fields**.
5. Select the field **product** from the list of **Available fields**. Click the **>** push button to add it to the list of **Supplementary fields**.
6. Select the field **gender** from the list of **Available fields**. Click the **>** push button to add it to the list of **Supplementary fields**.
7. Click **Next** to continue.

Specifying advanced parameters

The next few pages of the wizard are advanced pages. For this scenario, you accept the default values by clicking the **Next** button on the the next four pages. For more information about these parameters, see “Specifying advanced parameters” on page 14 and the online help for the Clustering – Demographic mining function.

1. On the Field parameters page of the wizard, click **Next** to continue.
2. On the Additional field parameters page of the wizard, click **Next** to continue.
3. On the Outlier treatment page of the wizard, click **Next** to continue.
4. On the Similarity matrix page of the wizard, click **Next** to continue.

Specifying parallel parameters

For this scenario, you run the mining function in serial mode. For more information on running this mining function in parallel mode, see the online help for the Clustering – Demographic mining function.

1. Ensure that the the **Run the serial mode of the function** radio button is selected.
2. Ensure that the **On the Intelligent Miner server node** radio button is selected.
3. Click **Next** to continue.

Specifying output fields

On the output fields page of the mining wizard, you must select the fields that the output data contains. You can include all or some of the input data fields in the output data. The output data will also contain fields generated by the mining function, including cluster ID, record score, and confidence. In this scenario, the record score value is a measure of how similar the customer is to the customer clusters in the model.

1. Select **>>** to add all the available fields to the list of output fields.
2. Type `clusterID` in the **Cluster ID field name** entry field.
3. Type `score` in the **Record score field name** entry field.
4. Type `conf` in the **Confidence field name** entry field.

   The output fields page now looks like this Figure 8 on page 22.

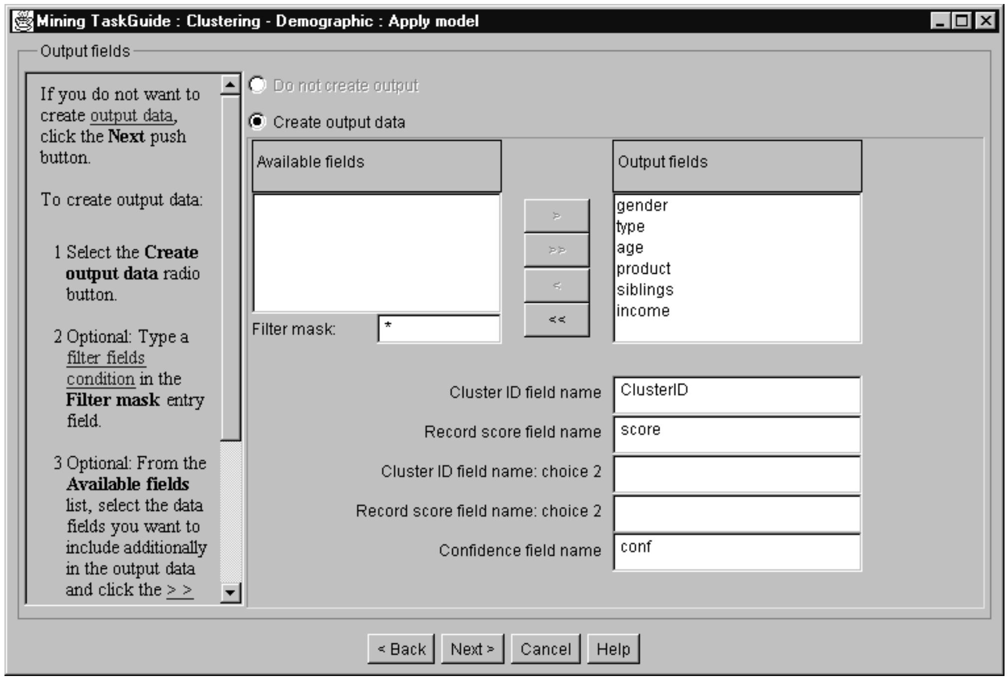

Figure 8. Input fields page

5. Click **Next** to continue.

Specifying the output data object name

On the output data page of the wizard, you specify the name of the output data object for this mining function. This page shows the existing data objects in the current mining base. Because you want to preserve the Customers data object, you need to create a data object that contains the output data named Scored customers. To do this, you can open the Data wizard from this page of the Mining wizard, define the data object, and then return to this page of the Mining wizard.

1. Click on **Create data**. The Data wizard's Welcome page opens.
2. Click **Next** to continue.
3. Select **Flat files**.
4. Type `Scored customers` in the settings name field. Optionally, type a comment to describe this data object.
5. Click **Next**.
6. On the **Flat files page**, change to the directory that contains the file banking.txt. Select **banking.txt**.
7. In the **Path and file name** entry field select **banking.txt** and type `scored.txt`.

   If you do not replace banking.txt with `scored.txt` in the entry field, you will overwrite and lose the original data in banking.txt.
8. Click on **Add file**.
9. Select the **The specified flat file does not yet exist** check box.
10. Click **Next** to continue.
11. On the Summary page of the Data wizard, click **Finish** to continue.

    After defining the output data object, return to the Mining wizard to continue the process of defining the mining object. You will see the data object in the Available output data container, as shown in Figure 9 on page 23.

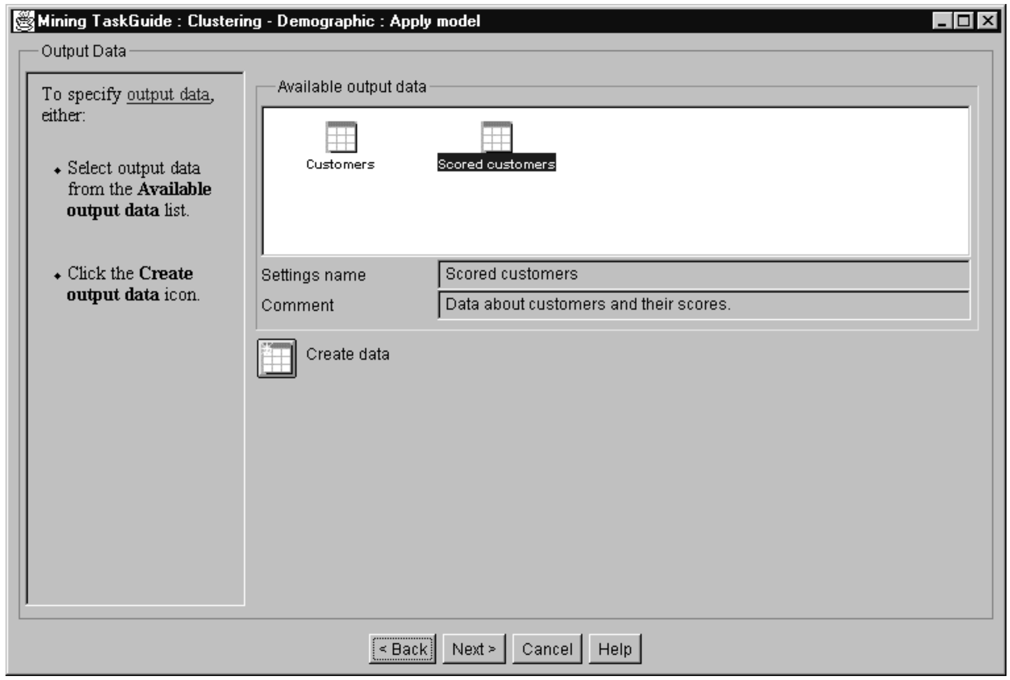

Figure 9. Output data page

1. On the output data page of the Mining wizard, select the data object **Scored customers**.
2. Click **Next**. The Summary page opens.
3. Select the **Run this settings immediately** check box.
4. Click **Finish** to continue. The Intelligent Miner will run the mining function and display a progress indicator that allows you to monitor the status of the mining function.
5. Click **OK** in the progress indicator window after the mining function has stopped running.

You now have a flat file that contains a list of customers and scores of how similar these customers are to the Premier Checking customers.

Creating a sequence

Now that we have created a few of the functions for our mining run, we can create a sequence that runs each of the functions in the order we specify.

To run a series of related data mining functions in sequential order, you can create a sequence object. When you run a sequence object, the Intelligent Miner processes each of the objects in the order you specify. You can use features of the Intelligent Miner sequence object to accomplish a wide variety of mining tasks without creating a specialized sequence for each task. You can run the following Intelligent Miner functions and objects alone, or include them in a sequence:

- Mining functions
- Processing functions
- Statistical functions
- Sequence objects

The benefit of using a sequence object is being able to combine several steps into one step. If you combine several functions into a sequence object, you need to run only the sequence object, which then runs each of the objects within it. You can also selectively enable or disable one or more objects within the sequence object to create specialized mining runs without creating a new sequence object.

To create the sequence for this scenario:

1. Click the **Create sequence** button on the Intelligent Miner main window toolbar.
2. On the **Welcome** page of the Sequence wizard, click **Next**.
3. In the **Setting name** field type `Target Marketing` as the name of this sequence object.
4. Click **Next**.
5. On the Parameters page of the Sequence wizard, you can use the mining base tree view to navigate to the objects you want to include in the sequence. Because the first object you want to run is the Build model mining function, click on the **+** next to the **Mining** folder.
6. Click the **Clustering** folder. The contents of the folder are displayed.
7. Select the mining function **Build model** and drag and drop it to the **Sequence** work area.

   The mining function build model is added to the sequence as the first object to run.
8. In the **Contents of folder** area, select the mining object **Apply model** and drag and drop the object to the **Sequence** work area.

   This adds the mining function **Apply model** to the sequence as the second object to run, after **Build model**.

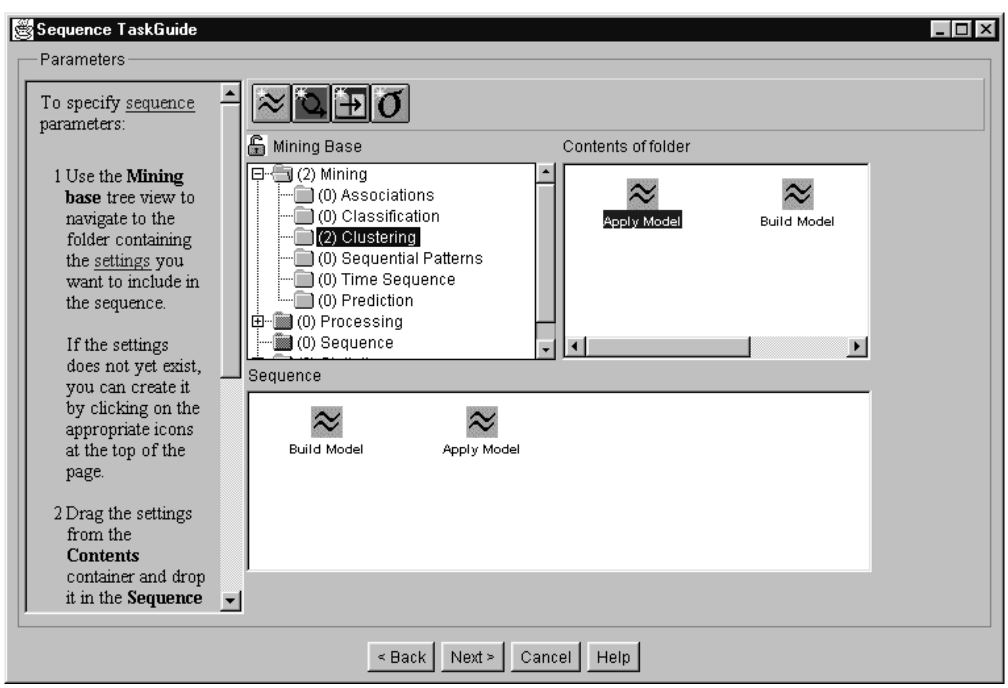

Figure 10. Parameters page of Sequence wizard

9. Click **Next**.

10. On the Additional parameters page of the Sequence wizard, check the **If a settings object in the sequence fails, continue running the sequence** check box.
11. Click **Next**.
12. On the Summary page of the Statistics wizard, select the **Run this settings immediately** check box.
13. Click **Finish**.

Chapter 5. Creating a statistics function

In this step of the mining scenario, you will create a Bivariate Statistics function. This function will produce descriptive statistics about the non-Premier Checking account customers most similar to the Premier Checking account customers.

To create a statistical function,

1. Click the **Create statistics** button from the toolbar on the Intelligent Miner main window. The Statistics wizard opens.
2. Click **Next** to continue.

Specifying the statistics function and name

On the Statistics functions and settings page, you select the type of statistics function that you want to create, and specify the name and comment for this statistics function:

1. Ensure that **Bivariate Statistics** in the list of statistics functions is selected.
2. Type the settings name `Analyze`. Optionally, type a descriptive comment.
3. Check the **Show the advanced pages and controls** check box.
4. Click **Next** to continue.

Specifying the input data for statistics function

To specify the input data for this statistics function:

1. From the list of available input data, select the **Scored customers** data object that you created in a previous step.
2. Under the Advanced options, click the **...** button next to **Filter records condition**. You will be filtering the records of the input data for this statistics function.

   The Intelligent Miner displays the Expression Builder for filtering records.

   In a previous step of this scenario, you created a score for each record in the input data. This score is a measure of how similar each customer is to the customers who already have Premier Checking accounts. The score ranges from 0 to 1, and a higher score means a greater similarity. For this scenario, you include all input data records for customers whose score is greater than 0.7.

   The expression you want to create looks like this: `((score>0.7))`. To create this expression:

   a. Click the **AND** push button. The expression builder creates a template for the expression, which appears as `((Arg1 = Arg2))` on the first line.
   b. Click on **Field Names** in the **Category** list.
   c. Scroll through the list of field names and select the field score.
   d. Click the **Arg1** button. The field score is set as the first argument in the expression.
   e. Click the **>** button.
   f. Click **Constants** in the **Category** list.
   g. Double-click **<new constants>** from the **Value** list.
   h. Type the constant value `0.7`.
   i. Press **Enter**.

j. Select **0.7** from the list of constants.

k. Click on the **Arg2** button.

The expression looks like this: `((score>0.7))`.

3. Click **OK** to return to the Input Data page of the statistics wizard, which now looks like Figure 11.

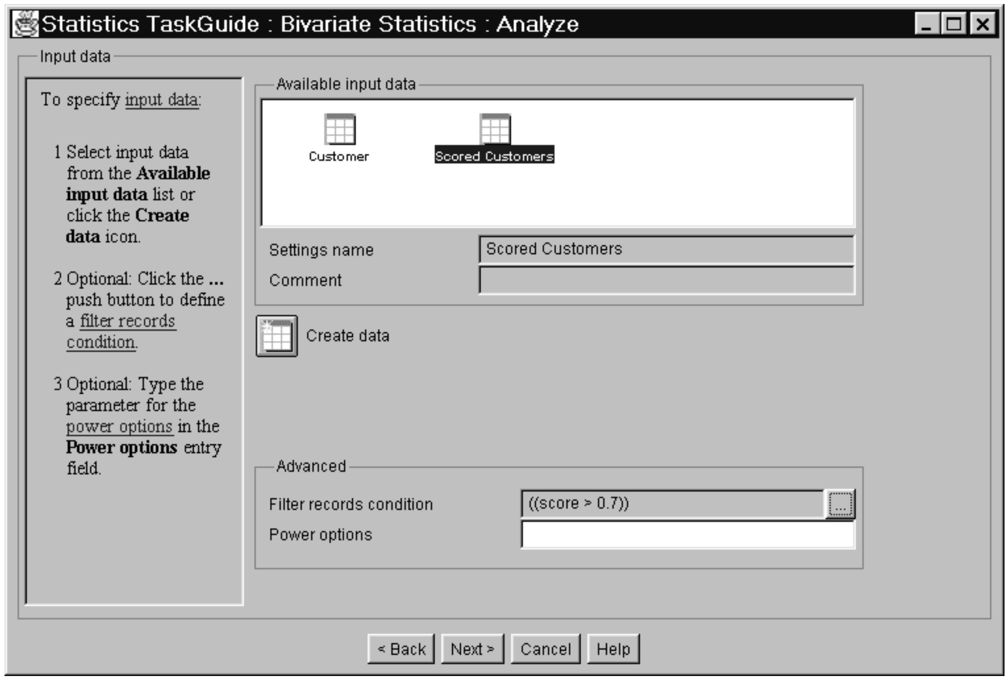

Figure 11. Input data page of the Statistics wizard

4. Click **Next**.
5. On the Parallel parameters page of the statistics wizard, make sure that you are running this function in serial mode, on the Intelligent Miner server node.
6. Click **Next**.

Computing statistics, quantiles, or a sample

On the Statistics page of this wizard, you specify whether you want this statistical function to compute statistics. You can choose to compute univariate statistics for one or more fields in the input data, or chi-square and bivariate statistics for one field in the input data. For this scenario, you will compute univariate statistics for the selected input data fields:

1. Select the **Compute statistics** radio button.
2. Select **conf**, **clusterID**, **age**, **product**, **siblings**, and **income** and click the **>** button to compute the univariate statistics for these fields.
3. Click **Next** to continue.

   On the Quantiles page of the Statistics wizard, you can choose to compute quantiles for the numeric fields in the input data. You will not compute quantiles in this scenario.
4. Click **Next**.

On the Sample page of the Statistics wizard, you can choose to create a sample for the input data fields. You will not create a sample in this scenario.

5. Click **Next**.

Specifying output fields

On this page of the Statistics wizard, you can choose whether to create output data, and what fields to include in the output data. You can include all or some of the input data fields in the output data. Only input data records that meet the filter condition that you specified will be included in the output data. In this case, the output data will contain customers whose scores are greater than 0.7.

1. Click **Create an output table**.
2. Click **>>** to add all the available fields to the list of output fields.
3. Click **Next** to continue.

Specifying the output data object name

On this page of the wizard, you specify the name of the output data object for this statistics function. This output data object must exist before you can proceed. Because you have not yet defined the output data object, you will need to define the output data object from this page of the Statistics wizard:

1. Click on **Create data**. The Data wizard opens.
2. Click **Next** to continue.
3. Select **Flat files**.
4. Enter `Target customers` as the name of the data object. Optionally, type a comment that describes this data object.
5. Click **Next** to continue.
6. On the Flat files page, change to the directory that contains the file banking.txt.
7. In the **Path and file name** entry field, type `target.txt`.
8. Click **Add** file.
9. Click **The specified flat file does not yet exist**.
10. Click **Next**.
11. On the Summary page of the Data wizard, click **Finish**. You now return to the Statistics wizard to continue the process of defining the statistics object.
12. On the output data page of the Statistics wizard, select the data object **Target customers** from the **Available output data** field.
13. Click **Next**.

Specifying the result object name

On the Results page of the Statistics wizard, enter the name of the result object that is generated by this statistical function:

1. Type `Target customer demographics` in the **Results name** field.
2. Check the **If a result with this name exists, overwrite it** check box.
3. Click **Next** to continue to the Summary page of the wizard.

Saving the statistical function

To save the statistical function, check the **Run this settings immediately** check box.

Click **Finish** to continue.

The Intelligent Miner will run this statistical function immediately when you click **Finish**. A progress indicator shows you the status of the function. After completion, the Intelligent Miner automatically displays the results of the statistical function.

Save the mining base from the Intelligent Miner main window. Click on the Save Mining Base icon.

You have produced an output data file named Target customer demographics. This file contains the customers identified as having a high score of similarity with the existing customers of the Premier Checking account. You can now analyze the results.

Chapter 6. Interpreting the results

The result generated by the Analyze statistical function is shown in Figure 12. The multiple graphs show the distribution of the fields that you selected for statistics. The visualizer allows you to display more detail by double-clicking on any of the graphs. The Bivarate Statistics function provides you with statistics about the customers that have been targeted for your ad campaign. A quick look at the product graph shows that the customers similar to the Premier Checking account customers are distributed fairly evenly over several other products.

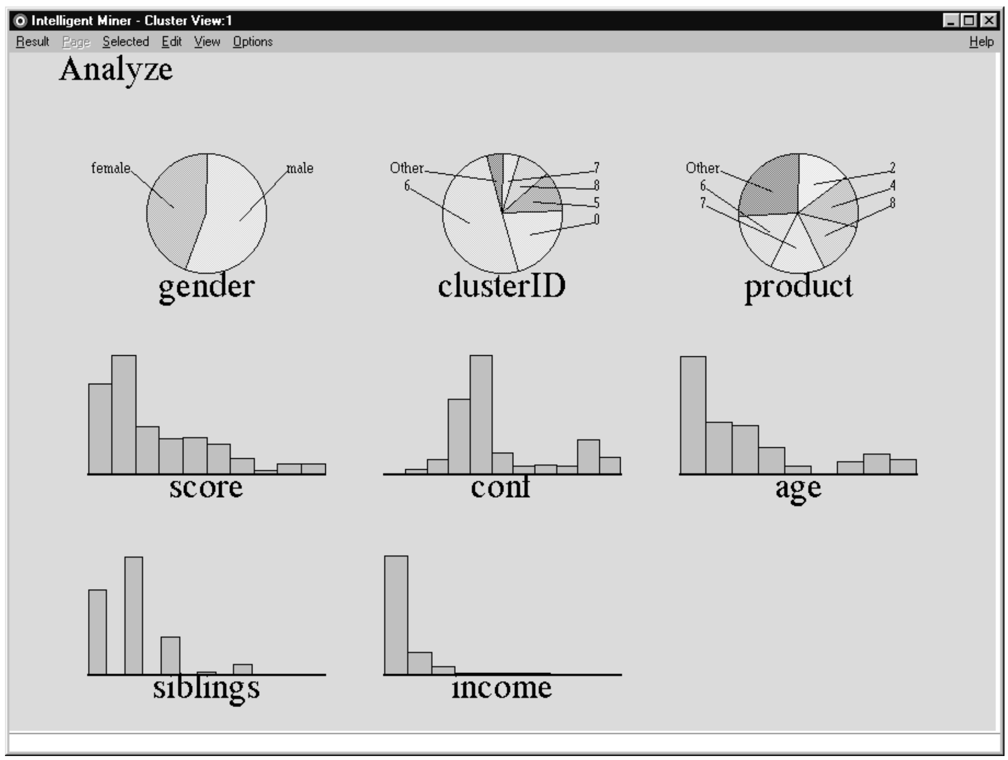

Figure 12. Results of the Analyze statistical function

The statistics computed by mining and statistics object are on the details page. The top portion of the details page for the results generated by the Analyze statistical function is shown in Figure 2.

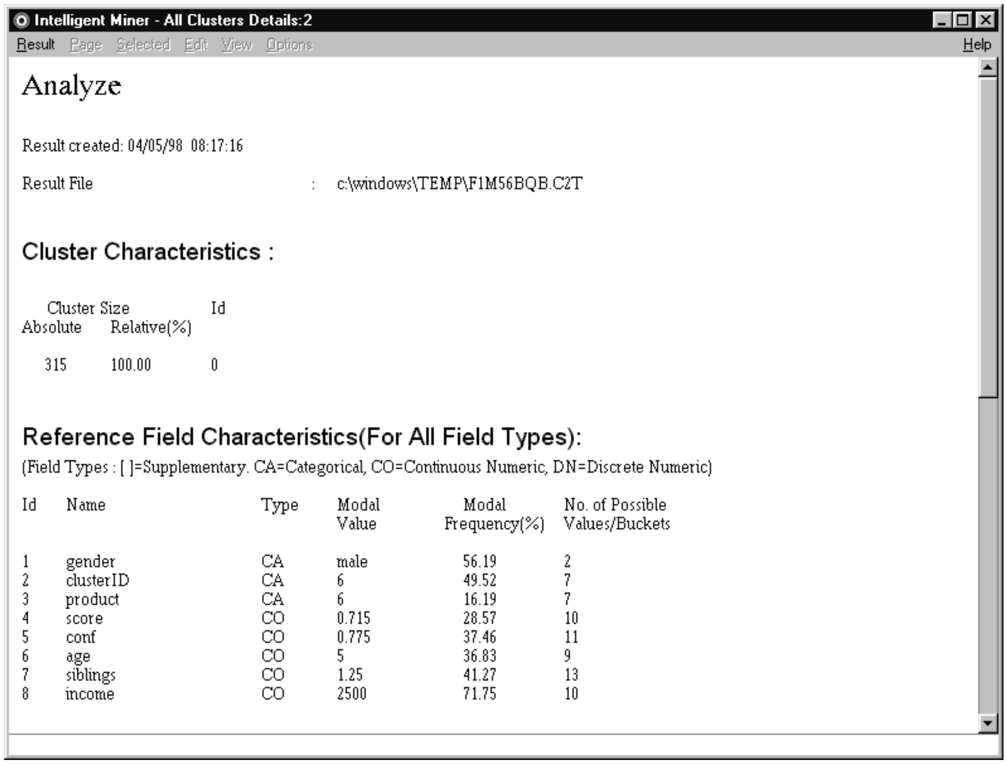

Figure 13. Details for results generated by the Analyze statistical function

To access this page, use the **Details for all partitions** menu item under the View menu on the menu bar. In Figure 13, you can see that of the possible 1792 customers who do not have Premier Checking, 315 had a score of 0.7 or greater. These are the customers who have similar demographics to those who have purchased Premier Checking. You could adjust the filter to 0.6 to include more customers in the analysis.

Notices

This information was developed for products and services offered in the U.S.A. IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing
IBM Corporation
500 Columbus Avenue
Thornwood, NY 10594
U.S.A.

For license inquiries regarding double-byte (DBCS) information, contact the IBM Intellectual Property Department in your country or send inquiries, in writing, to:

IBM World Trade Asia Corporation
Licensing
2-31 Roppongi 3-chome, Minato-ku
Tokyo 106, Japan

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Licensees of this program who wish to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact:

IBM Deutschland Informationssysteme GmbH
Department 3982
Pascalstrasse 100
70569 Stuttgart
Germany

© Copyright IBM Corp. 1998

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

The licensed program described in this information and all licensed material available for it are provided by IBM under terms of the IBM Customer Agreement, IBM International Program License Agreement, or any equivalent agreement between us.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurement may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

The Intelligent Miner incorporates code generated by the Purdue Compiler Construction Tool Set (PCCTS).

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

All statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

All IBM prices shown are IBM's suggested retail prices, are current and are subject to change without notice. Dealer prices may vary.

This information is for planning purposes only. The information herein is subject to change before the products described become available.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

If you are viewing this information softcopy, the photographs and color illustrations may not appear.

Trademarks and service marks

The following terms are trademarks of the IBM Corporation in the United States, or other countries, or both:

ACF/VTAM
AIX
AS/400
Common User Access
DATABASE 2
DB2
MVS
MVS/ESA
OS/2
OS/390
OS/400
Parallel Sysplex

DB2 Universal Database	Power PC
DFSMS	RACF
Distributed Relational Database Architecture	RS/6000
DRDA	SAA
IBM	SP2
Language Environment	System/390

Java and all Java-based trademarks and logos are trademarks of Sun Microsystems, Inc in the United States and other countries.

Microsoft, Windows, and the Windows NT are trademarks of Microsoft Corporation.

UNIX is a registered trademark in the United States and other countries licensed exclusively through X/Open Company Limited.

Other company, product, and service names may be trademarks or service marks of others.