Data Mining: Decision Trees

Applies to:
SAP BI 7.0. For more information, visit the EDW homepage

Summary
This article about the Data Mining and the Data Mining methods provided by SAP in brief. It explains the classification method ‘Decision Tree’ in detail. It also explains the steps for implementation of the Decision tree by creating a Model and an Analysis Process and creating a analysis for prediction.

Author: Vishall Pradeep K.S
Company: Applexus Technologies (P) Ltd
Created on: 20 April 2011

Author Bio
Vishall Pradeep is working as SAP Technology Consultant with Applexus Technologies (P) Ltd. He has experience in SAP ABAP and SAP BI
Table of Contents
Introduction .......................................................................................................................... 3
Data Mining Methods ........................................................................................................ 3
Decision tree .................................................................................................................... 4
    Creating a Model ........................................................................................................ 4
    Creating a Analysis Process for Training ................................................................. 7
Creating a Analysis Process for Executing a Prediction .................................................... 12
Related Content ................................................................................................................ 19
Disclaimer and Liability Notice ....................................................................................... 20
Introduction

Data mining is to automatically determine significant patterns and hidden associations from large amounts of data. Data mining provides you with insights and correlations that had formerly gone unrecognized or been ignored because it had not been considered possible to analyze them. The data mining methods available in SAP BW allow you to create models according to your requirements and then use these models to draw information from your SAP BW data to assist your decision-making.

Data Mining Methods

SAP delivers the following SAP-owned data mining methods,

1. Decision trees:

   Decision trees are used to learn from historic data and to make predictions about the future. Prediction involves establishing rules using historic data and applying these rules to new data. These rules are displayed graphically as a hierarchy.

2. Clustering:

   Clustering allows you to segment data automatically into clusters. In a subordinate dataset, the system groups together associated data by forging formerly unknown links.

3. Association analysis:

   The purpose of association analysis is to find patterns in particular in business processes and to formulate suitable rules, of the sort "If a customer buys product A, that customer also buys products B and C".

4. Scoring:

   The purpose of scoring is to valuate data records and it has three function types as below
   a) Weighted Score tables
   b) Linear regression
   c) Non Linear Regression

5. ABC Classification:

   Displays data grouped into classes of A, B, C and so on, using thresholds and classification rules. The classified results are displayed in the form of ABC chart or list. ABC classification allows you to classify your data based on specified classification rules. The data to be classified is generated by a query in the SAP BW. The classification rules refer to a single key figure value in your data and implicitly specify which absolute or relative key figure values map to which classes.
**Decision tree**

Decision trees are used to learn from historic data and to make predictions about the future. Prediction involves establishing rules using historic data and applying these rules to new data. These rules are displayed graphically as a hierarchy.

**Creating a Model**

- Go to Transaction RSDMWB (Data Mining Workbench)

  1. Data Mining->Expand Classification->Right Click Decision Tree->Create Model

- Choose the Model Name and Description
- The method name for which you are creating a model is displayed. You have three options for model field selection
- To create the model fields manually, select the Manual option.
- If you want to create a model that is similar to an existing model created previously, you can copy it choosing the Use Model as Template option. You can make minor changes to the copied version manually to suit your requirements.
To create a model from a query, choose Model Field Selection and select the query which you want use as a source for model fields. The InfoObjects contained in the selected query are available in the next step as model fields.

The screen shows the list of Fields and we can select and exclude fields in it.

In the step Edit Model Fields, specify the attributes for each field and the description you give the model field does not necessarily have to be identical with that of the InfoObjects.

The Content types valid for a model field are dependent on the method that you are creating the model for and on the data type of the model field. The value type specified for a model field determines which entries can be made as Field Parameters and Field Values.
• Set the *Prediction Variable indicator* for the model field for which the subsequent prediction is to be made. Select as a prediction variable that model field for which you wish to gain more information (via the model).

**Create Model ZPD T_MODE L**

- **Header**
  - Name: ZPD T_MODE
  - Description: Product Model
  - Created by
  - Created on
  - Method

- **Fields**
  - **Name**: Description
  - **OBASE_UOM**: Base Unit of Measure
  - **OBASE_UOM**: UNIT
  - **OD_PO_QTY**: Purchase order quantity (SAP D)
  - **OD_PO_QTY**: QUAN
  - **OD_EBELN**: Purchasing document number
  - **OD_EBELN**: CHAR
  - **OD_EBELP**: Item number of purchasing document
  - **OD_EBELP**: NUMC

- **Parameters**
  - **Training Process**
    - Use Sampling
    - Init. Sample Size (%): 15.00
    - Max. Sample Size (%): 75.00
    - No. of Trials: 2
  - **Stopping Conditions**
    - Minimum Leaf Cases: 10
    - Min. Leaf Node Accuracy(%): 95.00
  - **Advanced Settings**
    - Perform Relevance Check
      - Use Threshold: 0.10
    - Use Top-N Attributes: 10
  - **Pruning**
    - Perform Pruning
    - Use Extended Pruning

• In the *Model Parameters* step, enter the parameters that are valid for the entire model. The model parameters are dependent on the data mining method.

• Save and Activate the Model (we can only train or validate a model or use it for the prediction if the model has been activated.)
Creating a Analysis Process for Training

- Go to Transaction RSANWB (Analysis Process Designer)

- Choose General->Right Click->Create

- Give the description to the APD
From the Data Sources, drag and drop the Query to the work area.

It asks for a Popup and click on Choose Query.

From the Help, Select the query.
• Click on continue and the Query as data Source is added as below

• For the data target, drag the icon for the relevant data mining method in the work area

• Connect the two nodes
- Double click on data mining node to make the settings in the dialog box that appears
- Choose the required model from F4 Help

![Data Mining: Decision Trees Diagram](image)

- To make an explicit field assignment, double click on the data flow arrow that connects the nodes
- Click on Automatic Assignment and choose Same Infoobject

![Field Assignment Diagram](image)
- Save and activate the APD
- While saving it will ask for a Technical Name

![Image of Data Mining Model](image)

- Execute the APD

![Image of Display Logs](image)

- To view the training results, in the context menu of data target, choose **Data Mining Model → View Model Results**
- The data mining model acquires the status **Trained**.

![Image of Display Model](image)
Creating a Analysis Process for Executing a Prediction

- A model that you trained using historic data from a source can now be applied to a different set of data. By doing so, a prediction is made for the model field that you selected as the prediction variable.
- Goto Transaction RSANWB (Analysis Process Designer)

SAP Easy Access - User menu for APPLEXUS DEVELOPER

Choose General->Right Click->Create

Analysis Process Designer

Give the description to the APD

Analysis Process Designer - Create -
• From the Data Sources, drag and drop the Query to the work area

Analysis Process Designer - Create -

- It asks for a Popup and click on Choose Query

• From the Help, Select the query

- GDS - Test by Vishall
- IC for Ekpo
- Purchase Order Test - Vishall
- Purchase Test - Vishall
- Human Resources
- Industry Sectors
• Drag the relevant prediction icon, that is, source for transformation, in the work area

• Connect the two nodes
• Double click on data mining node to make the settings in the dialog box that appears
• Choose the required model from F4 Help

• In select Prediction Output Fields “Select the Fields”
• For the data target, drag the icon for the relevant data mining method in the work area (In this case I am downloading the predicted values to the .CSV file)

• Choose the file name

• In CSV File properties tab, Choose the separators

• Connect the nodes
• Save and activate the APD
• While saving it will ask for a Technical Name

![Image of technical name input]

• Execute the APD

![Image of display logs]

• To display the prediction results, choose Display Data from the context menu

![Image of display data selection]
Data is displayed as below

<table>
<thead>
<tr>
<th>Order quantity</th>
<th>Unit</th>
<th>Part No.</th>
<th>Base Unit</th>
<th>Prediction</th>
<th>Probability</th>
<th>Probability</th>
<th>Probability</th>
<th>Predicted V</th>
<th>Predicted V</th>
<th>Predicted V</th>
</tr>
</thead>
<tbody>
<tr>
<td>250,000</td>
<td>ST</td>
<td>40601969</td>
<td>ST</td>
<td>4</td>
<td>0.1198</td>
<td>0.1198</td>
<td>0.0008</td>
<td>5500000130</td>
<td>5500000131</td>
<td>4600000823</td>
</tr>
<tr>
<td>125,000</td>
<td>ST</td>
<td>40601969</td>
<td>ST</td>
<td>4</td>
<td>0.1198</td>
<td>0.1198</td>
<td>0.0008</td>
<td>5500000130</td>
<td>5500000131</td>
<td>4600000823</td>
</tr>
<tr>
<td>85,000</td>
<td>ST</td>
<td>40601969</td>
<td>ST</td>
<td>4</td>
<td>0.1198</td>
<td>0.1198</td>
<td>0.0008</td>
<td>5500000130</td>
<td>5500000131</td>
<td>4600000823</td>
</tr>
<tr>
<td>61,000</td>
<td>ST</td>
<td>40601969</td>
<td>ST</td>
<td>4</td>
<td>0.1198</td>
<td>0.1198</td>
<td>0.0008</td>
<td>5500000130</td>
<td>5500000131</td>
<td>4600000823</td>
</tr>
<tr>
<td>127,000</td>
<td>ST</td>
<td>40601969</td>
<td>ST</td>
<td>4</td>
<td>0.1198</td>
<td>0.1198</td>
<td>0.0008</td>
<td>5500000130</td>
<td>5500000131</td>
<td>4600000823</td>
</tr>
<tr>
<td>70,000</td>
<td>ST</td>
<td>40601969</td>
<td>ST</td>
<td>4</td>
<td>0.1198</td>
<td>0.1198</td>
<td>0.0008</td>
<td>5500000130</td>
<td>5500000131</td>
<td>4600000823</td>
</tr>
<tr>
<td>100,000</td>
<td>ST</td>
<td>40601969</td>
<td>ST</td>
<td>4</td>
<td>0.1198</td>
<td>0.1198</td>
<td>0.0008</td>
<td>5500000130</td>
<td>5500000131</td>
<td>4600000823</td>
</tr>
</tbody>
</table>

To display the summary of the prediction results, choose Calculation Summary

Calculation Summary will be displayed as below
To display the stored results, choose Intermediate Results

**Intermediate result is as below**

- The results are downloaded into the flat file as well (We can use another Decision tree as a target to have the values)
Related Content

http://help.sap.com/saphelp_erp2004/helpdata/en/72/e5293b5fdebb1ce10000000a114084/content.htm

For more information, visit the EDW homepage
Disclaimer and Liability Notice

This document may discuss sample coding or other information that does not include SAP official interfaces and therefore is not supported by SAP. Changes made based on this information are not supported and can be overwritten during an upgrade.

SAP will not be held liable for any damages caused by using or misusing the information, code or methods suggested in this document, and anyone using these methods does so at his/her own risk.

SAP offers no guarantees and assumes no responsibility or liability of any type with respect to the content of this technical article or code sample, including any liability resulting from incompatibility between the content within this document and the materials and services offered by SAP. You agree that you will not hold, or seek to hold, SAP responsible or liable with respect to the content of this document.