Data Mining: Scoring (Linear Regression)

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

Summary
This article deals with Data Mining and it explains the classification method ‘Scoring’ in detail. It also explains the steps for implementation of Linear Regression by creating a Model and an Analysis Process.

Author: Vishall Pradeep K.S  
Company: Applexus Technologies (P) Ltd  
Created on: 16 May 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
Scoring .................................................................................................................................................... 3
Regression Analysis ............................................................................................................................... 3
Linear regression .................................................................................................................................... 3
Creating a Model ..................................................................................................................................... 3
Creating a Analysis Process for Training .............................................................................................. 7
Related Content ...................................................................................................................................... 15
Disclaimer and Liability Notice ............................................................................................................ 16
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.

Scoring

The data is displayed using continuous quantities. If required, discretization can then be applied to split the data into classes. The scoring function can either be specified using weighted score tables or be determined by training using historic data as linear or nonlinear regression of a target quantity. The purpose of scoring is to valuate data records.

Regression Analysis

It is used to automatically define valuation functions and thereby determine numeric target values. If you wish to generate the valuation functions, you need to train the analysis process using historic data. After we determined the valuation functions either by defining them directly or by training them on the basis of historic data, you can then apply them to other datasets as part of a prediction.

Linear regression

The system uses training data records with known score values to determine a linear function that approximates the scores of the training data records and the system trains the scoring function using data with known target values. You need to set the value type of the target value as continuous. At least one of the other model fields must also be continuous. The system defines a separate linear function for each combination of values in discrete model fields that occur in the training data.

Creating a Model

- Go to Transaction RSDMWB (Data Mining Workbench)

- Data Mining->Expand Approximation->Right Click Scoring->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:

- **Manual**: To create the model fields manually, select this option.
- **Use Model as Template**: If you want to create a model that is similar to an existing model created previously, you can copy it choosing this option. You can make minor changes to the copied version manually to suit your requirements.
- **Model Field Selection**: To create a model from a query, choose this option and select the query which you want to use as a source for model fields. The InfoObjects contained in the selected query are available 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) and define a Key field and it should not be a Prediction field.

---

**Create Model ZPDT_MODEL_SCORING**

**Fields**

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>InfoObject</th>
<th>Data</th>
<th>Long</th>
<th>Content Type</th>
<th>Pred</th>
<th>Param</th>
</tr>
</thead>
<tbody>
<tr>
<td>001_EBELN</td>
<td>Purchasing document number</td>
<td>001_EBELN</td>
<td>UNIT</td>
<td>3</td>
<td>Discrete</td>
<td></td>
<td></td>
</tr>
<tr>
<td>001_EBELP</td>
<td>Item number of purchasing docu</td>
<td>001_EBELP</td>
<td>QUAN</td>
<td>23</td>
<td>Continuous</td>
<td></td>
<td></td>
</tr>
<tr>
<td>001_EBELN</td>
<td>Base Unit of Measure</td>
<td>001_EBELN</td>
<td>CHAR</td>
<td>10</td>
<td>Key Field</td>
<td></td>
<td></td>
</tr>
<tr>
<td>001_EBELP</td>
<td>Purchase order quantity (SAP D)</td>
<td>001_EBELP</td>
<td>NUGC</td>
<td>5</td>
<td>Continuous</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Choose the Regression type
In the Model Parameters step, to exclude combinations with a minimal amount of data records, we can use the model parameter Minimum Number of Records.
If we select the indicator Skip input outside of trained domain, then no score value is calculated for such data records. If you do not select this indicator, the default score value is assigned to these data records.
With the parameters of the model fields, you can specify for discrete fields whether all values, just special values, or just the most frequent values should be considered. For continuous fields, you can explicitly specify both limits of a value range or have them specified automatically by choosing the option Complete Data Range. With the automatic option, the limits are determined by rounding off the maximum and minimum values of the field in the training data. When the function is applied to other data, values occurring outside of this range are then treated as outliers.

Save and Activate the Model (we can only train or valuate a model or use it for the prediction if the model has been activated.)

Log is Displayed
Creating a Analysis Process for Training

- We can create an analysis process for a data mining method to train a data mining model. The purpose of training a model using historic data is to allow the model to learn from the historic data. The training result can then be used for a prediction or in the operational system.

- Go to 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
• 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
• And Click "OK"

• The Query which is the data Source is added as below

• For the data target, drag the icon for the relevant data mining method in the work area
- Double click on data mining node to make the settings in the dialog box that appears
- Choose the required model from F4 Help

![Image of the dialog box for selecting a data mining model](image1)

- Click on "ENTER" and Connect the two nodes

![Image of the Analysis Process Designer with Data Mining Linear Regression node](image2)
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.

Click on Continue and Save and activate the APD.

While saving it will ask for a Technical Name.

Execute the APD.

The data is written to the data target and a log is displayed.
To view the training results, in the context menu of data target, choose Data Mining Model → View Model Results

The Output will be displayed as below

- The Output will be displayed as below
Scoring Output for Model ZPDT_MODEL_SCORING

Overall Score Descriptive Statistics

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Average</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Invalid</th>
<th>Zero</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score</td>
<td>818</td>
<td>14345.96</td>
<td>77225.61</td>
<td>368333.57</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Click on Other statistics to view Goodness Indicator

Scoring Output for Model ZPDT_MODEL_SCORING

Goodness Indicators

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>I</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Discrete Fields</th>
<th>Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base Unit of Measure</td>
<td>Intercept</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>KG</td>
<td>217.74</td>
</tr>
<tr>
<td>KG</td>
<td>1.04</td>
</tr>
<tr>
<td>M2</td>
<td>16.99</td>
</tr>
<tr>
<td>EA</td>
<td>9113.86</td>
</tr>
<tr>
<td>TO</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Click on General Characteristics and it will ask for a popup as below

Choose the model field and select all to view the output

Scoring Output for Model ZPDT_MODEL_SCORING
Related Content

Data Mining: Decision Tree
Data Mining: Clustering

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.