Credit Scoring for SAS® Enterprise Miner™

Integrated scorecard development, deployment and monitoring for better decisions

What does Credit Scoring for SAS® Enterprise Miner™ do?
Credit Scoring for SAS Enterprise Miner is used to build, validate and deploy credit risk models. It helps create credit scorecards using in-house expertise and resources to decide whether to accept an applicant (application scoring); to determine the likelihood of defaults among customers who have already been accepted (behavioral scoring); and to predict the likely amount of debt that the lender can expect to recover (collection scoring).

Why is Credit Scoring for SAS® Enterprise Miner™ important?
Credit Scoring for SAS Enterprise Miner enables you to make accurate and timely default predictions to streamline credit approval processes; improve customer acquisition, retention and collections; and reduce exposure to business risk in an organization’s consumer lending portfolio. It helps balance between too much credit exposure – leading to high default rates and charge-offs – and not enough, which can result in lost business and revenue.

For whom is Credit Scoring for SAS® Enterprise Miner™ designed?
Credit Scoring for SAS Enterprise Miner is designed for scorecard developers to perform the statistical analyses needed to create credit scorecards and execute credit risk models. It is also suitable for the credit scoring manager who oversees the portfolio and scorecard usage.

For best performance, detailed knowledge about credit risk cannot be locked up in complex spreadsheets. This approach doesn’t accommodate advanced, predictive analytics and business intelligence to build a risk-aware culture. Organizations need a complete view of credit worthiness of the lending portfolio to take quick action against market changes, competitive pressures and population instabilities. But how do you get there?

Tedious, manual tasks of data preparation and characteristic screening and binning can be nearly impossible to carry out with large data sets. Besides the inability to understand relationships, you may not select the best risk prediction variables and attributes. Software can automate much of these processes, as well as adapt and test the three leading methods of reject inference.

Having software that develops credit scorecards allows organizations to bring the process in-house, lowering overall costs, decreasing cycle time and reducing missed information. By owning the process, you gain more insight into your data. And the added benefit of streamlined reporting allows managers to disseminate vital information in a timely manner.

Credit Scoring for SAS Enterprise Miner enables you to make accurate and timely default predictions to streamline credit approval processes, improve collection management and reduce exposure to business risk. SAS allows firms to develop, deploy and track credit risk scorecards in-house, instead of relying on external credit modeling services that can increase outsourcing expenses, lengthen development cycle times and decrease internal dissemination of the vital information.

Key Benefits

• Comprehensive data preparation. Save time and resources through the ability to access, transform, cleanse and prepare all prerequisite data – including third-party bureau, application, bill-payment and collections data – with SAS. Data sets, no matter the size, can be examined quickly and easily for patterns, anomalies and missing values though built-in, interactive nodes with many options for exploration, transformation, missing value imputation, outlier analysis and correlation analysis.

• Efficient scorecard development. SAS provides a fast, flexible and economical option to create and deploy credit scorecards for virtually all types of consumer lending products – accounts, cards, loans, mortgages – leading to better credit decisions and reduced losses. You can compute scorecard points for each attribute using either the WOE variables or the group variables that are exported as inputs for the logistic regression model, and you can manually assign scorecard points to attributes.

• Variable selection and treatment to quickly understand relationships and behaviors. Credit Scoring for SAS Enterprise Miner purposely censors the data, making it easy to understand relationships and allowing nonlinear dependencies to be modeled with linear models. This gives the user control over the development process and provides insights into the behavior of risk predictors. The node also screens characteristics so that potentially predictive variables are used while other variables are not.
• Automatic creation of the target variable for the rejects data set. As a necessary step in applying a remedy for selection bias, unrealistic expectations and model overconfidence, SAS offers three industry-accepted ways to infer the rejected data – fuzzy augmentation, parceling and hard cutoff. More robust estimates can be quickly made on how the model performs on both the known population and the entire “through-the-door” population.

• Better outcomes and improved customer portfolios performance. Assess and control risk within existing consumer portfolios and improve acquisition strategies using SAS advanced predictive analytics techniques. This approach provides a better understanding of the specific risk characteristics and subsequent attributes that lead to delinquency, default and bad debt.

Product Overview

Flexible data preparation and management capabilities

SAS data access, integration and management capabilities make it easier to prepare data across disparate systems and sources for analysis, and build credit risk models that generalize well and produce superior outcomes. SAS comprehensive variable selection techniques can lead to better credit risk modeling.

SAS® Enterprise Miner™: award-winning predictive analytics software

SAS Enterprise Miner provides the most comprehensive set of advanced predictive and descriptive modeling algorithms – including scorecard, decision trees, neural networks, logistic regression, etc. – and assesses scorecard quality to determine the best course of action.

Patented optimal rigorous binning method

Credit Scoring for SAS Enterprise Miner has a patented optimal rigorous binning method that yields true optimal bins based on constraints defined by the user.

Data Partition node

Most data mining projects use large volumes of sampled data. After sampling, the data is usually partitioned before modeling. The Data Partition node, part of SAS Enterprise Miner, partitions your data into training, validation and/or test data sets.

Interactive Grouping node

When creating a credit scorecard and screening and binning univariate characteristics, the Interactive Grouping node ensures there is purposeful censoring of the data and that there are assessments on the strength of each characteristic individually as a predictor of performance.
Scorecard node

The Scorecard node fits a logistic regression model and computes the scorecard points for each attribute. Users can choose either the WOE variables or the group variables that are exported by the Interactive Grouping node as inputs for the logistic regression model. The scorecard points of each attribute are based on the coefficients of the logistic regression model. The Scorecard node also enables you to manually assign scorecard points to attributes. The scaling of the scorecard points is also controlled by the three Scaling Options properties of the Scorecard node.

Reject Inference node

The sample data used to develop a credit scorecard is structurally different from the population to which it is applied. The non-event or event target variable that is created for the credit scoring model is based on the records of applicants accepted for credit. However, the population to which the credit scoring model is applied includes applicants who were rejected under the scoring rules that were used to generate the initial model.

The Reject Inference node provides a remedy for this inherent selection bias by using the model that was trained with the accepted applications to score the rejected applications. The observations in the rejected data set are classified as inferred non-event and inferred event. The inferred observations are then added to the accepts data set to form an augmented data set, which represents the “through-the-door” population and serves as the training data set for a second scorecard model.

<table>
<thead>
<tr>
<th>Input Variable</th>
<th>Adverse Reason 1 Count</th>
<th>Adverse Reason 2 Count</th>
<th>Adverse Reason 3 Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGE</td>
<td>356</td>
<td>96</td>
<td>91</td>
</tr>
<tr>
<td>CARDS</td>
<td>402</td>
<td>126</td>
<td>23</td>
</tr>
<tr>
<td>EC_CARD</td>
<td>178</td>
<td>45</td>
<td>38</td>
</tr>
<tr>
<td>INCOME</td>
<td>68</td>
<td>223</td>
<td>87</td>
</tr>
<tr>
<td>PERS_H</td>
<td>4</td>
<td>78</td>
<td>146</td>
</tr>
<tr>
<td>STATUS</td>
<td>4</td>
<td>119</td>
<td>260</td>
</tr>
<tr>
<td>TMJOB1</td>
<td>188</td>
<td>157</td>
<td>103</td>
</tr>
</tbody>
</table>

Figure 3. The Scorecard node provides flexible reporting on adverse characteristics including overall counts per predictor.

Figure 4. Scorecards can be generated as a simple summary or detailed view.

Figure 5. The Scorecard node allows flexible scorecard scaling based on three Scaling Options properties: Odds, Scorecard Points and Points to Double Odds.

<table>
<thead>
<tr>
<th>Grouped: Age</th>
<th>Weight of Evidence: Age</th>
<th>Grouped: Income</th>
<th>Weight of Evidence: Income</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.0</td>
<td>0.3726752853</td>
<td>1.0</td>
<td>0.519053675</td>
</tr>
<tr>
<td>4.0</td>
<td>0.3726752853</td>
<td>2.0</td>
<td>-0.693147101</td>
</tr>
<tr>
<td>3.0</td>
<td>0.2946916154</td>
<td>1.0</td>
<td>0.519053675</td>
</tr>
<tr>
<td>4.0</td>
<td>0.3726752853</td>
<td>3.0</td>
<td>-0.350929566</td>
</tr>
<tr>
<td>7.0</td>
<td>0.3726752853</td>
<td>1.0</td>
<td>0.519053675</td>
</tr>
<tr>
<td>1.0</td>
<td>-1.077858879</td>
<td>3.0</td>
<td>-0.350929566</td>
</tr>
<tr>
<td>4.0</td>
<td>0.3726752853</td>
<td>1.0</td>
<td>0.519053675</td>
</tr>
<tr>
<td>5.0</td>
<td>1.0414030740</td>
<td>2.0</td>
<td>-0.693147101</td>
</tr>
<tr>
<td>3.0</td>
<td>0.2946916154</td>
<td>3.0</td>
<td>-0.350929566</td>
</tr>
</tbody>
</table>

Figure 6. The Interactive Grouping node codes each predictor to Weight of Evidence values to use as inputs in the Scorecard node.

Figure 7. An example Process Flow Diagram representing a subset of a typical credit scorecard flow.
Seamless integration into SAS® Credit Scoring for Banking solution

Credit Scoring for SAS Enterprise Miner software allows seamless integration into the SAS Credit Scoring for Banking solution, a comprehensive solution for developing, deploying and managing scorecards for operational and regulatory compliance that includes data management at the front end and model management at the back end.

Credit Scoring for SAS® Enterprise Miner™ System Requirements

To learn more about Credit Scoring for SAS® Enterprise Miner™ system requirements, download white papers, view screenshots and see other related material, please visit sas.com/creditscoring.

Key Features

Data Partition node
- Training, validation and/or test data sets based on simple random, cluster or stratified random sample.

Interactive Grouping node
- Use predefined groupings.
- Prebinning through quantile or bucket methods.
- Interval binning through optimal criterion, quantile, monotonic event rate and constrained optimal methods.
- Interactive bin changes.
- Variable selection through Gini statistic or information value.
- Response rate plots.
- Gini statistic and information value plots.
- Output of variable mappings.

Scorecard node
- Assignment of scorecard points.
- Intercept-based scorecard.
- Reverse scorecard.
- Scorecard point scaling.
- Use of indeterminate values in gains charts.
- Adverse characteristic analysis through neutral score or weighted average score methods.
- Inputs for trade-off plots.
- Score distribution charts.
- Strategy curve.
- Event frequency chart.
- Scorecard strength chart.
- Selection methods include backward, forward and stepwise.
- Regression criteria include AIC, SBC, validation error, validation misclassification rate, cross-validation error, and cross-validation misclassification rate.

Reject Inference node
- Inference methods include fuzzy, hard cutoff and parceling.
- Classification charts of actual versus inferred.
- Distribution plots of actual versus inferred.
- Summary statistics of actual versus inferred.