ENTERPRISE MINER: 4 – CREDIT SCORECARD DEVELOPMENT

DR IAIN BROWN, ANALYTICS & INNOVATION PRACTICE, SAS UK
15 NOVEMBER, 2016
• SAS – 15\textsuperscript{th} November 2016 at 10:00am
• Enterprise Miner: Credit Scorecard Development
• The session looks at:
  - Variable Classing and Selection
  - Scorecard Modelling
  - Reject Inference
  - Scorecard Assessment
THE ANALYTICS LIFECYCLE

PREDICTIVE ANALYTICS AND DATA MINING

IDENTIFY / FORMULATE PROBLEM

DATA PREPARATION

DATA EXPLORATION

TRANSFORM & SELECT

BUILD MODEL

VALIDATE MODEL

DEPLOY MODEL

EVALUATE / MONITOR RESULTS

Domain Expert
Makes Decisions
Evaluates Processes and ROI

BUSINESS MANAGER

Data Exploration
Data Visualization
Report Creation

BUSINESS ANALYST

Model Validation
Model Deployment
Model Monitoring
Data Preparation

IT SYSTEMS / MANAGEMENT

Exploratory Analysis
Descriptive Segmentation
Predictive Modeling

DATA MINER / STATISTICIAN
SAS® ENTERPRISE MINER™

SEMMA IN ACTION – REPEATABLE PROCESS
• Build Scorecards to assign risk scores to customers:
  • Application Scoring
  • Behavior Scoring
  • Probability of Default Scoring (rating)
  • Collection scoring
• Lot’s of control over the scorecard construction. Not a black box.
• Broad set of tools to support the complete data mining process
• Easy-to-use GUI to develop more –and better – scorecards
• Easy to distribute and scalable system
• An open, extensible design for ultimate flexibility
  • GUI + SAS Code Node + Extension Nodes
  • Batch processing
  • Extensive model scoring alternatives
  • SAS, Java, PMML, Scoring Accelerator for Teradata
CREDIT SCORING

OVERVIEW

• Application scoring: Likelihood that an applicant will not repay a loan and will therefore fall into default
  • Applicant characteristics (for example, age, income, employment status, time at address, …)
  • Credit bureau information
  • Application information of other applicants
  • Repayment behaviour of other applicants
• Develop models (scorecards) estimating the probability of default of a customer
• Typically, assign points to each piece of information, add all points, and compare with a threshold (cutoff)
CREDIT SCORING

OVERVIEW

• Behavioural scoring: Through the cycle risk of default based on an existing customer’s behaviour
• Update risk assessment taking into account recent behaviour
• Example of behaviour:
  • Average/Max/Min/trend in checking account balance, bureau score, …
  • Delinquency history (payment arrears, …)
  • Job changes, home address changes, …
• Dynamic
1. Variable Classing and Selection
   - Calculation of Weight of Evidence statistics
   - Calculation of Information Value or Gini statistics
   - Automatic and Interactive
     - Enhances productivity
     - Incorporate business knowledge

2. Scorecard construction
   - Fitting a logistic regression model
   - Scaling and calculating score points

3. Assessing Scorecard quality
   - KS, GINI, ROC and Trade-off charts

4. Reject Inference
   - Model for scoring through-the-door population
EM Credit Scoring nodes:

- **Automatic and interactive variable grouping**
  - Computes Weights of Evidence
  - GINI and Information Values for variable selection

- **Scorecard construction**
  - Logistic regression based using WOE or dummy variables as inputs
  - Parameterized score points scaling
  - Assessment statistics and charts

- **Reject inference**
  - Through the door impact analysis
Let cutoff = 500 points

A new customer applies for credit.

**AGE** 32 120 points

**HOUSE** OWN 225 points

**INCOME** $45K 200 points

Total 545 points

**ACCEPT FOR CREDIT**
CREDIT SCORING IMPROVEMENTS

At constant approval rate:
Reduce bad rate from 2.5% to 1.9%

At constant bad rate:
Increase approval rate from 70% to 87%
• Automatic and interactive variable grouping
• Computes Weights of Evidence

Auto-updating IV and Gini

Fine/Coarse Detail
Metrics for Characteristic Analysis

- Measures commonly used in a characteristic analysis
  - Weights of Evidence (WOE) – measures the strength of the attribute of a characteristic in separating good and bad accounts
  - Information Value (IV) – measures the overall predictive power of the characteristic; that is, its ability to separate good and bad loans. (weighted sum of WOE)
  - Gini Statistic – alternative approach to IV to assess the overall predictive power of a characteristic.
Grouping options

- The Interval and Ordinal Grouping Method fields determine how the grouping algorithm groups the pre-binned interval and ordinal variables:
  - **Optimal Criterion** – groups the variables based on the criterion property
  - **Quantile** – enables the algorithm to generate groups with approximately the same frequency in each group.
  - **Monotonic Event Rate** – generates groups that result in a monotonic distribution of event rates across all levels
  - **Constrained optimal** – generates groups based on predefined constraints
CREDIT SCORING

SCORECARD = LOGISTIC REGRESSION + SCALING

• In order to determine the relative contribution of each characteristic logistic regression is used.

• One regression co-efficient per characteristic is produced if Weight Of Evidence coding is used.

• Dummy variable (1-of-n coding) would instead lead to one co-efficient per attribute.
• An attribute’s points reflect:
  • Risk of an attribute relative to other attributes of the same characteristic; determined by an attribute’s WOE, Weight of Evidence, value
  • Relative contribution of each characteristic to the overall score; determined by a characteristic’s co-efficient in a logistic regression model
In order to obtain more ‘friendly’ scores, the points per attribute are linearly scaled with a factor and an offset.

The score is proportional to the good/bad odds and not the bad/good odds that are modeled in the logistic regression, thus a negative sign is introduced.

Smaller scores thus correspond to higher risk.
CREDIT SCORING

SCORECARD NODE SAMPLE RESULTS
Good / bad Information is only available for past accepts, but not for past rejects.

In order to arrive at models for the through-the-door population, it is common practice to perform ‘Reject Inference’: apply the scorecard to the rejects and classify rejects as inferred good / inferred bad.

An ‘augmented’ input data set then is created by adding the inferred good / bad to the actual good / bad. The scorecard is re-adjusted using this data set.
• The reject inference node enables three types of augmentation technique to be applied:
  • Hard Cutoff
  • Parcelling
  • Fuzzy
Hard Cutoff Augmentation

1. Build a scorecard model using the known good/bad population (that is, accepted applicants).
2. Score the rejected applicants with this model to obtain each rejected applicant’s probability of default and their score on the scorecard model.
3. Create weighted cases for the rejected applicants.
4. Set a cutoff score level above which an applicant is deemed good. All applicants below this level are deemed bad.
5. Add the inferred goods and bads back in with the KGB’s and rebuild the scorecard.
CREDIT SCORING

REJECT INFEERENCE – TECHNIQUES

Parcelling Augmentation

1. Build a scorecard model using the known good/bad population.
2. Score the rejected applicants with this model to obtain each rejected applicant’s probability of default.
3. Create weighted cases for the rejected applicants.
4. The inferred good/bad status of the rejected applicants will be assigned randomly and proportional to the number of goods and bards in the accepted population within each score range.
5. If desired, apply the event rate increase factor to P(bad) to increase the proportion of bards among the rejects (rule of thumb 2-8 times that of the accepted applicants)
6. Add the inferred goods back in with the known goods and bards and rebuild the scorecard.

<table>
<thead>
<tr>
<th>Score Range</th>
<th># Bad</th>
<th># Good</th>
<th>% Bad</th>
<th>% Good</th>
<th>Reject</th>
<th>Rej - Bad</th>
<th>Rej - Good</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-169</td>
<td>290</td>
<td>971</td>
<td>23.0%</td>
<td>27.0%</td>
<td>1,649</td>
<td>379</td>
<td>1,267</td>
</tr>
<tr>
<td>170-179</td>
<td>530</td>
<td>2,414</td>
<td>18.0%</td>
<td>82.0%</td>
<td>1,732</td>
<td>312</td>
<td>1,420</td>
</tr>
<tr>
<td>180-189</td>
<td>366</td>
<td>2,242</td>
<td>14.0%</td>
<td>86.0%</td>
<td>3718</td>
<td>521</td>
<td>3,198</td>
</tr>
<tr>
<td>190-199</td>
<td>131</td>
<td>1,790</td>
<td>10.0%</td>
<td>90.0%</td>
<td>1,734</td>
<td>735</td>
<td>6,407</td>
</tr>
<tr>
<td>200-209</td>
<td>211</td>
<td>2,427</td>
<td>8.0%</td>
<td>92.0%</td>
<td>1,176</td>
<td>94</td>
<td>1,082</td>
</tr>
<tr>
<td>210-219</td>
<td>213</td>
<td>4,047</td>
<td>5.0%</td>
<td>95.0%</td>
<td>3,519</td>
<td>176</td>
<td>3,342</td>
</tr>
<tr>
<td>220-229</td>
<td>122</td>
<td>2,928</td>
<td>4.0%</td>
<td>96.0%</td>
<td>7211</td>
<td>286</td>
<td>6,923</td>
</tr>
<tr>
<td>230-239</td>
<td>139</td>
<td>6,814</td>
<td>2.0%</td>
<td>98.0%</td>
<td>3,871</td>
<td>77</td>
<td>3,794</td>
</tr>
<tr>
<td>240-249</td>
<td>98</td>
<td>10,912</td>
<td>0.8%</td>
<td>99.2%</td>
<td>4773</td>
<td>36</td>
<td>4,735</td>
</tr>
<tr>
<td>250+</td>
<td>94</td>
<td>18,706</td>
<td>0.5%</td>
<td>99.5%</td>
<td>9622</td>
<td>45</td>
<td>9,927</td>
</tr>
</tbody>
</table>
Fuzzy Augmentation

1. Build a scorecard model using the known good/bad population and score the rejected applicants with this model to obtain the probability of the rejected applicant being a good \( P(\text{good}) \) and the probability of being a bad \( P(\text{bad}) \).

2. Do not assign a reject to a good/bad class. Instead, create two weighted cases for each rejected applicant using \( P(\text{good}) \) and \( P(\text{bad}) \).

3. Multiply \( P(\text{good}) \) and \( P(\text{bad}) \) by the user-specified rejection rate to form frequency variables.

4. Results are in two observations for the rejected applicants. One observation has a frequency variables (rejection weight \(*P(\text{good})\)) and a target variable value of 0. The other observation has a frequency variable (rejection weight*\(P(\text{bad})\)) and a target variable value of 1.
CREDIT SCORING

EM – APPLICATION SCORECARD DEVELOPMENT

Input Data Source:
Contains information on historic Known Good Bads

Data Partition:
Division of raw data into training and validation

Variable Clustering:
Select best subset of independent variables

Interactive Grouping:
Optimal coarse and fine categorisation of inputs

Scorecard:
Logistic Regression model to predict a customer’s probability of default
Data Partition: Division of augmented sample into training and validation

Final augmented Good Bad model
CREDIT SCORING

SCORECARD DEPLOYMENT

• Model Deployment
  • Application of model in business processes for delivery of extracted knowledge in the right time to the right people
  • Realises model value by providing decision support
  • Integration requires scoring code in appropriate language

• Score code
  • includes all data transformation of the scorecard process
  • produced in SAS, C, Java and in-database (Teradata)
  • Enables batch and real-time scoring in SAS, databases and on the web

• Metadata
  • Integration into SAS Business Analytics Framework
• EM – Application Scorecard
Three types of model quality criteria are monitored

- Model stability
- Model performance
- Model calibration

Multiple statistical indicators suggested under Basel Committee Working Paper

Additionally indicators for monitoring LGD and CCF models
MODEL STABILITY

• Important for detecting population shifts, for example for pre-deployment sanity check
  • Is the distribution of scores still similar to when you developed the model?
  • Have you now more high (low) scoring customers than previously?
A measure used to monitor the score distribution:

\[ SSI = \text{Sum} \left( (A-D) \times \log \left( \frac{A}{D} \right) \right) \]

where \( A \) and \( D \) are the No. of Accounts in the Actual and Development sample.
MODEL PERFORMANCE

• Important for ensuring high quality pooling and approval decisions
  • Do the bad customers have low scores?
  • Do the good customers have high scores?
  • To what degree do the score distributions of good and bad customers overlap?
  • How well does the model separate the good from the bad customers?
  • How many of all bad customers can you find within the low scoring customers?
• Validation of the calibration is more complex than the validation of the discriminatory power of internal rating systems
• To transform a PD from a score, direct or indirect methods can be used
• Important for ensuring correct risk assessment, for example Risk Weighted Assets calculation
  • How similar are predicted default rates to actual default rates?
  • Are the differences significant?
• Measures the “closeness” of observed vs. estimated default rates
REPORTING FOR FS REGULATORS BASED ON BASEL 2 REQUIREMENTS

- Probability –of-Default (PD) and Loss-Given-Default (LGD) model performance monitoring
- Calculation of 30 model performance statistics
  - Model Stability
  - Model Performance
  - Model Calibration
- Customizable traffic lighting
- Trend time series charts
SUMMARY HOW TO LEVERAGE SAS ENTERPRISE MINER FOR CREDIT RISK MODELLING

• Make accurate and timely risk calculations to improve
  • Approval processes
  • Risk management
  • Basel compliancy
  • Collection management

• Provides
  • In–house model development
    • A-IRB models
  • Flexible model deployment
  • Out-of-the-box reporting