Evaluation of the performance of 3 DM techniques in the creation of an Credit Card Attrition Model

Why?

Attrition → Big Problem

Causes:
- Open Commerce
- High competition
- Access to Information
- Bad Management
- Aggressive Marketing
- ...

Possible Solutions:
- Better Products
- Loyalty Programs
- Retention Campaigns
- Customer Relation Management
- ...

Retention Campaigns

Customers

Identification of the Customers with the highest probability of leaving
Where to Start?

Identification of Customers

Classification Model

Possible definitions of a Customer that leaves:
- Customer that cancelled all his products
- **Customer that cancelled one of his products**
- Customer that stopped doing transactions for a period of time
- ...

Customer that cancelled one of his products

X Credit Card

Identify the Customers with the highest probability of cancelling their X Credit Card

Where to Start?

Create a Data Mining Model

SAS Enterprise Miner

This software structures Data Mining processes in 5 stages:

- **Sample***
- Explore
- **Modify***
- **Model***
- Assess

Let’s Start!
Sample

Population: Customers that had a X Credit Card in February 2003
- 248,057 -

and in March 2003
1,104 Cancelled (0.45%)
246,953 Didn’t Cancelled

Create a random sample of the non-attrition to have a proportion of 10% - 90%. Why?

The initial Data Set has 2 variables:
- Customer ID
- Target (=1 if cancelled and =0 otherwise)

Selection of the Model inputs

Information merged in to the Data Set selected from the Bank Datamart:
- Customer socio-demographic characteristics (February 2003);
- Customer transactional behaviour (September 2002 - February 2003);
- Customer score in other models (February 2003);
- Product characteristics (February 2003);
- Product transactional behaviour (September 2002 - February 2003);
- etc.
Sample

Creation of a Enterprise Miner diagram

Inserting a Input Data Source

A Target Profile was created a to give weight 9 to a True Positive. Why?

Splitting the initial Data Set in 3 different Data Sets: Why?

- Training
- Validation
- Test

Explore

Exploration of the distribution of the several inputs

Some of the problems found:

- Values very concentrated and outliers
Explore

- Values not meaningful

Minimum - 0
Maximum - 1

Minimum - 6.551
Maximum - 24.002

Explore

- Very different scales

Sex
Minimum - 0
Maximum - 1

Age (in number of day)
Minimum - 6.551
Maximum - 24.002
Modify

The transformation used has:
Interval variables
- Logarithm
- Square
- Inverse
Class variables (ordinal)
- Exponential
- Square root
- Percentile

This transformations were performed with SAS macro's that:
- Creates the code that transforms the variables
- Applies that code to the variables in the initial Data Set

Why not use the Transform Variables node in Enterprise Miner?

The range of all the interval variables (original and transformed) was transformed into [0,1] range. Why?

The final Data Set has:
- 1 target variable
- 1 id variable
- 245 input variables (that after transformation become 2.237)
Model

Data Mining techniques used:

- Decision Tree
- Regression
- Neural Networks

How it works

- Recursive method
  - In each iteration it tries to identify which variable alone does the best job in separating the different classes according to the target variable
  - Creates a set of if-then rules, using one variable at a time, dividing the input space in rectangular subspaces
**Model**

**Strengths**
- Creates rules that can be easily explained in current language
- The rules are computationally simple to create and fast to apply to new data
- It accepts both interval and class variables
- Doesn’t require complex data preparation on missing values, outliers, scales, etc.
- Consistent with the existence of rare values in the validation and test data sets
- Provides indication of which variables are most related with the target

**Weaknesses**
- Doesn’t consider relation between variables
- May give very complex solutions for linear problems
- Very instable to changes in the inputs of the training data set
- Bad results if the event to predict is rare in the training data set
- Needs powerful computation resources
- Transforming interval variables in classes leads to a loss of information
Choosing the best Decision Tree
( maximum Average Profit )

Options initially chosen:
Number of branches = 2
Maximum depth = 6
Observations required for a split search = 110

Splitting Criteria
- Chi-Square test (Default)
- Entropy reduction (Default)
  - obs. required for a split search = 2
- Gini reduction (Default)

How it works
- Creates a linear equation that returns the predicted value of the event in analysis given a set of inputs
- Estimates the weight of each input in the calculation of that value
- These weights are chosen so the difference between the predicted value and the real one is minimum
- In a Logistic Regression the value that we want to predict is the probability of the event, so that value has to be in [0,1]

\[ \pi(x) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ...}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ...}} \]
Model

Strengths

- Considers relations between variables
- Accepts both interval and class variables
- The equation created by the regression can be interpreted
- It is stable to changes in the inputs of the training data set

Weaknesses

- Produces only linear (hyperplane) separations between classes
- Requires intensive data preparation
- Ignores inputs with missing values
- Dealing with class variables is not simple
- High computer capacity to develop and apply the regression equation to new data
- Requires statistically significant training data set
Choosing the best Logistic Regression
(maximum Average Profit)

Options:
- Stepwise
- Significance Levels=0.05

Link Function
- Probit (Default)
- Logit (Default)
- Cloglog (Default)

How it works
- Each unit starts by aggregating the inputs in a single value, using the chosen combination function
- Outputs a value using the chosen transfer function
- The outputs of the units of one hidden layer are used as inputs of the units of the next hidden/output layer
- This process goes on until it reaches the output layer that gives the output of the Neural Network
- The weights of each variable in each unit are chosen so that the predicting error is minimum

Model

\[ e^{-1 + 0.3 \text{age} + 0.3 \text{sex} + 0.06 \text{salary}} \]
Model

Strengths
- Works with very complex problems and gives good results
- Accepts both interval and class variables
- Can handle linear and non-linear problems
- Robust
- May give an idea of which inputs are the most relevant in the model

Weaknesses
- Doesn’t easily explain the results
- Requires data preparation
- Ignores inputs with missing values
- Dealing with class variables is not simple
- Sensitive to local minima with the weight space
- Non-deterministic
Choosing the best Neural Networks
(maximum Average Profit)

Options:
- Multilayer Perceptron
- Default training technique
- Hyperbolic Tangent (Activation Function)

Network Structure
- 1 Hidden Layer, 1 Neuron
- 1 Hidden Layer, 5 Neurons
- 2 Hidden Layer, 8 Neurons (5+3 Neurons)

Assess
Insert Data Mining techniques nodes
Assess

Accumulative Response

Non-Accumulative Response
Assess

Accumulative Captured Response

Misclassification

- Decision Tree: 0.0990338164
- Regression: 0.0997886473
- Neural Networks: 0.0999396135

Conclusions

- Even with simple information, Data Mining techniques applied with SAS Enterprise Miner, produce good results in the creation of an Attrition Model.

- The best splitting criteria in the Decision Tree was clearly the Chi-square test.

- The cloglog was the link function with worse results in Logistic Regression.

- The number of neurons, and not hidden layers, improved the performance of the Neural Networks.

- All three DM techniques presented very similar results in the identification of leaving Clients.

- To achieve better results, we first have to focus on the quality of the information used.
What’s next?

- Determine the best transformations of each variable
- Create new information by combining the variables
- Test different structures of Neural Networks
- Create new models that combine the results of the previous ones
- Apply these models to a new Data Set, that will have information relative to 2004
- Compare the performance of the final models applied to this new Data

Thank you!