PREDICTING DONORS IN FUND RAISING USING ARTIFICIAL NEURAL NETWORK

Nafisseh Heiat, College of Business, Montana State University-Billings, 1500 University Drive, Billings, MT 59101. 406-657-2224, nheiat@msubillings.edu

ABSTRACT

Data mining tools make it possible to apply mathematical models to the historical data to manipulate and discover new information. In this study, we are applying MultiLayer Perceptron (MLP) and Radial Basis Function (RBF) neural networks to classify donor and non-donors in a national charitable organization.

INTRODUCTION

Fund raising has definitely come a long way from the pre-World War I spontaneous giving, and it has gone way past professional. Organizations and universities are trying to find ways of making their efforts more efficient and lucrative for their organizations.

The goal of this study is to use data mining to show how certain attributes relate to past patterns. Using attributes that strongly correlate with fund raiser's contribution, help us in developing scoring systems that are specific to organizations. Once a group has assigned a score to each person in the database, it can decide on whom to focus appeals and can limit mailings to donors with high probabilities of giving donations. Two different data mining techniques, MultiLayer Perceptron (MLP) and Radial Basis Function (RBF) neural networks were used in this study to classify donors and non-donors

DATA PREPARATION

The data used in this research provided by IBM Watson Analytics Community- Fund Raising for a National Charity. It consists of 17 usable variables including the dependent variable Donors and Non-Donors. The categorical variables were coded as follow:

HOMEOWNER 1 = homeowner, 0 = not a homeowner

NUMCHLD Number of children

INCOME Household income

GENDER Gender

0 = Male

1 = Female

WEALTH Wealth Rating

Wealth rating uses median family income and

population statistics from each area to

index relative wealth within each state

	The segments are denoted 0-9, with 9 being the highest wealth group and zero being the lowest. Each rating has a different meaning within each state.
HV	Average Home Value in potential donor's neighborhood in $ hundreds
ICmed	Median Family Income in potential donor's neighborhood in $ hundreds
ICavg	Average Family Income in potential donor's neighborhood in hundreds
IC15	Percent earning less than 15K in potential donor's neighborhood
NUMPROM	Lifetime number of promotions received to date
RAMNTALL	Dollar amount of lifetime gifts to date
MAXRAMNT	Dollar amount of largest gift to date
LASTGIFT	Dollar amount of most recent gift
TOTALMONTHS	Number of months from last donation to July 1998 (the last time the case was updated)
TIMELAG	Number of months between first and second gift
AVGGIFT	Average dollar amount of gifts to date
TARGET	
	1 = Donor
	0 = Non-donor

To analyze the data categorical variables needed to be pre-processed for data mining. Certain variables had to be taken into account and others excluded. The excluded variables did not have any likely impact on the employee attrition. The data was prepared and run through exploratory analysis which in Modeler is called Feature Selection in order to find the most influential variables.
The target or dependent variable is Donor following table shows the variables included in the dataset. However, using Feature Selection in SPSS Modeler the total months, average gift, last gift, number of promotions, and number of children are displayed as most important Features.

METHODOLOGY

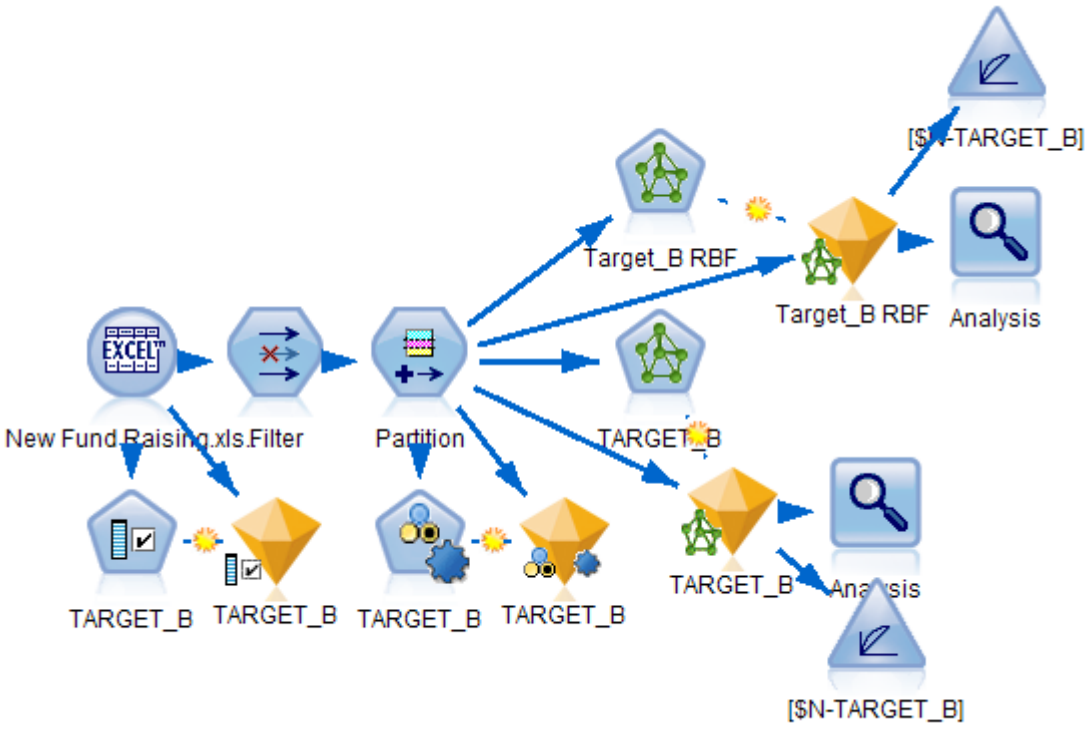
Data Mining may be defined as the process of finding potentially useful patterns of information and relationships in data. In this study I used MultiLayer Perceptron (MLP) and Radial Basis Function (RBF) neural networks. According to Rejane B. Santos the RBF and MLP networks are usually used in the

same kind of applications (nonlinear mapping approximation and pattern recognition), however their internal calculation structures are different (Santos and et.al, 2013)

FINDINGS/DISCUSSION

The Auto-Classifier tool was used to arrive at the best model for determining donors. The most efficient algorithms with highest accuracy rates is neural network. Figure 2 displays the diagram for the models created by SPSS Modeler. The diagram starts with selecting the data set for the analysis. It follows with a Filter node that selects the appropriate inputs and assigns the appropriate data type to the target and input variables. Next, the cleaned dataset was partitioned to training and testing sets (70%, 30%). Once the Auto-classifier determined the best model, that algorithm was applied to the dataset and analysis and evaluations nodes were added to analyze the results.

Figure 2. Study Models



MultiLayer Perceptron ANN Results- Based on the MLP algorithm the most important Variables are shown in Figure 3. According to MLP network analysis Number of Children, Last Gift, and Average Gift are the most important factors determining a donor. Other less important variables are listed in Figure 3. Confusion matrix in Figure 4 shows that the ANN model using the test (validation) dataset predicts donors correctly 54.35% of the time.

Figure 3. Important Variables

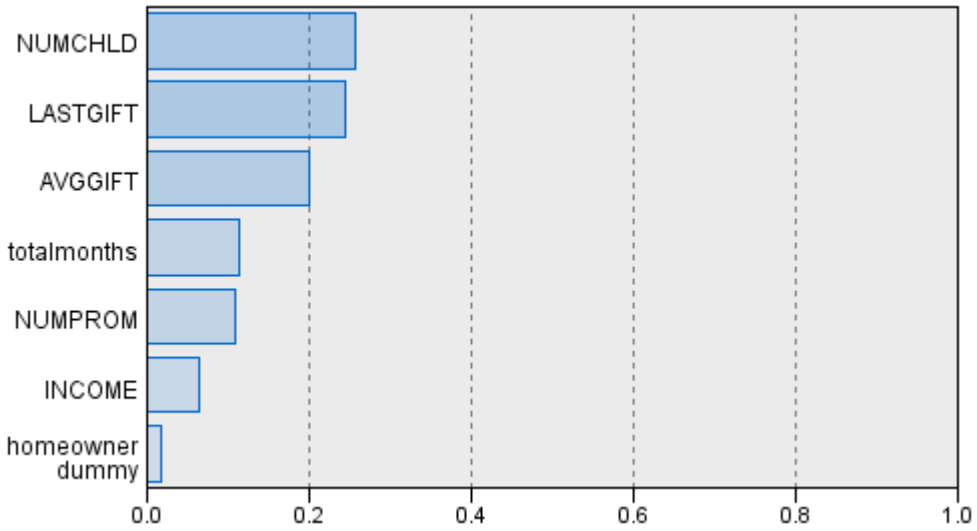


Figure 4. MLP Confusion Matrix

Comparing $N-TARGET_B with TARGET_B

'Partition'	1_Training		2_Testing	
Correct	1,211	57.1%	543	54.35%
Wrong	910	42.9%	456	45.65%
Total	2,121		999	

Coincidence Matrix for $N-TARGET_B (rows show actuals)

'Partition' = 1_Training	0.000000	1.000000
0.000000	631	451
1.000000	459	580

'Partition' = 2_Testing	0.000000	1.000000
0.000000	257	221
1.000000	235	286

Radial Basis Function Results- Based on the RBF algorithm the most important Variables are shown in Figure 6. According to RBF network analysis Total Months, Last Gift, Average Gift, and Number of promotions are the most important factors determining a donor. Other less important variables are listed in Figure 6.

Confusion matrix in Figure 7 shows that the ANN model using the test (validation) dataset predicts donors correctly 55% of the time. The gain charts in Figure 8 demonstrate the improvement gained by using ANN model as compared with a non-model approach like using average attrition.

Figure 6. RBF Important Variables

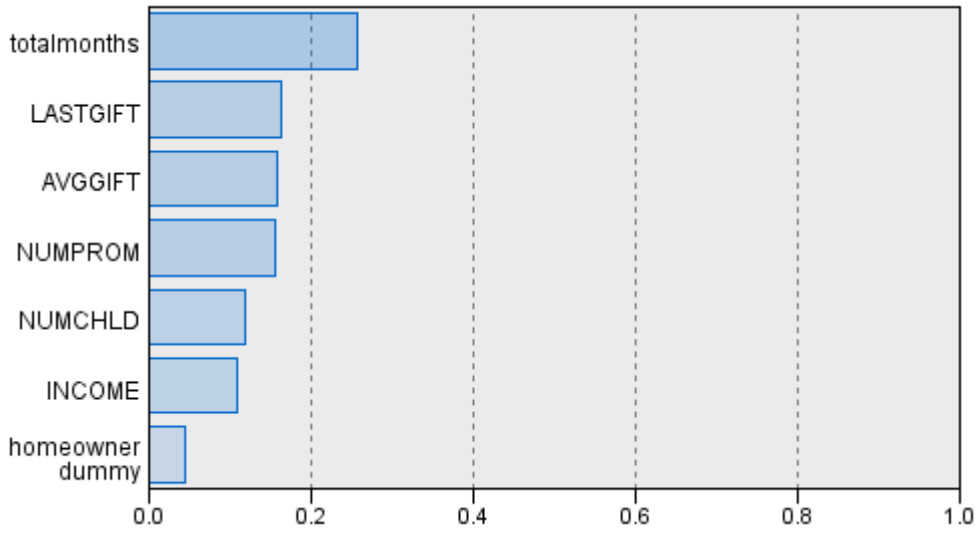


Figure 7. RBF Confusion Matrix

Comparing $N-TARGET_B with TARGET_B

'Partition'	1_Training		2_Testing	
Correct	1,213	57.19%	550	55.06%
Wrong	908	42.81%	449	44.94%
Total	2,121		999	

Coincidence Matrix for $N-TARGET_B (rows show actuals)

'Partition' = 1_Training	0.000000	1.000000
0.000000	670	412
1.000000	496	543

'Partition' = 2_Testing	0.000000	1.000000
0.000000	287	191
1.000000	258	263

CONCLUSIONS

Two classification methods used to develop models for predicting people who donate to a national charity. Both Artificial Neural Network (ANN) models, MLP and RBF predicted the donors accurately at 54.3% and 55%. Of the time. MLP and RBF models indicated Last Gift and Average Gift as important predictors. However, RBF listed the total months as the most important predictor and Number of promotions as an important predictor. Accuracy rate of both models, while showing improvement versus not using a model, are low. This might be due to the specific dataset that I used, or other environmental or organizational factors. In that case separate models based on different types and environments should be developed. Further studies are needed to investigate, confirm or reject the validity of the last statement.

REFRENCES

References are available upon request.