A DATA MINING APPROACH TO FUNDRAISING: AN EXPLORATRY ANALYSIS

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 data mining techniques to fund-raising by utilization of data-mining methods that are appropriate for fund-raising applications

INTRODUCTION

Fundraising has gone way past professional. Organizations and universities are trying to find ways of making their efforts more efficient and lucrative for their organizations. Gone are days of randomly targeting individuals for donations and being happy with whatever is collected. Fundraising has become an organized project for most organizations which utilize every resource available to them to achieve a target goal. Non-profit organizations are specially depended on donations and charities.

Colleges, universities, and other nonprofit organizations amass a great deal of information about the people they serve. What many organizations don't realize is that they can use this information to gain valuable insight to improve their advancement efforts. Applying effective fundraising analytics, such as data mining and predictive modeling can yield significant benefits in cost savings and more productive projects.

With information about which prospect relationships will pay off, you can move immediately to:

- Create precisely targeted lists for annual fund and membership campaigns
- Identify new prospects for major and planned gifts
- Find new alumni volunteer candidates
- Save money and generate more revenue on specific appeals
- Incorporate data-driven decision making into your organization

Followings are some of the benefits of applying data mining to a university (name has been removed) foundation database:

1. The use of data mining saves time and expenses of reaching out to those who are unlikely to donate and also of having to implement a multi-day training program. By using different tools in STATISTICA, a data mining software, we are able to predict donations accurately.
2. University foundation project teaches skills that can be used immediately to achieve lower appeals costs and more donation revenue.
3. The interactive table of contents, index, and glossary make this project a great reference for each new data mining project.
4. The demonstration of this project can be helpful among university staff to target potential alumni students who are willing to help our university by way of donation.

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 data mining techniques to fund-raising by utilization of data-mining methods that are appropriate for fund-raising applications.

In this project we use statistical and data mining tests that show us 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 their 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.

METHODOLOGY

Data Mining may be defined as the process of finding potentially useful patterns of information and relationships in data. More and more healthcare organizations are storing large amounts of data about patients and their medical conditions. As the quantity of clinical data has accumulated, domain experts using manual analysis have not kept pace and have lost the ability to become familiar with the data in each case as the number of cases increases. Data visualization techniques can assist in the manual analysis of data, but ultimately the human factor becomes a bottleneck as an organization using a large database can receive hundreds or even thousands of matches to a simple query [1,2,4].

Improved data and information handling capabilities have contributed to the rapid development of new opportunities for knowledge discovery. Interdisciplinary research on knowledge discovery in databases has emerged in this decade. In healthcare, pattern recognition has long been linked with expertise. Data mining, as automated pattern recognition, is a set of methods applied to knowledge discovery that attempts to uncover patterns that are difficult to detect with traditional statistical methods. Patterns are evaluated for how well they hold on unseen cases. Databases, data warehouses, and data repositories are becoming ubiquitous, but the knowledge and skills required to capitalize on these collections of data are not yet widespread. Innovative discovery-based approaches to healthcare data analysis warrant further attention [5,6,7,8].

There are situations where healthcare organizations would like to search for patterns but human abilities are not well suited to search for those patterns. This usually involves the detection of "outliers", pattern recognition over large data sets, classification, or clustering using statistical modeling. Medical data has a lot of information buried within it that will reveal patterns relating to successes and failures in clinical operations. Data mining by discovering these patterns could provide new medical information[9,10, 12].

In this research I used three classification methods plus regression analysis. The following is a brief description of these classification methods.

Decision Trees

Decision trees and rule induction are two most commonly used approaches to discovering logical patterns within medical data sets. Decision trees may be viewed as a simplistic approach to rule discovery because of the process used to discover patterns within data sets.

Decision tree is built through a process known as binary recursive partitioning. This is an iterative process of splitting the data into partitions, and then splitting it up further on each of the branches. Initially, you start with a training set in which the classification label (say, "diabetic" or "non-diabetic") is known (pre-classified) for each record. All of the records in the training set are together in one big box. The algorithm then systematically tries breaking up the records into two parts, examining one variable at a time and splitting the records on the basis of a dividing line in that variable (say, BMI > 30 or BMI <=30). The object is to attain as homogeneous set of labels (say, "diabetic" or "non-diabetic") as possible in each partition. This splitting or partitioning is then applied to each of the new partitions. The process continues until no more useful splits can be found. The heart of the algorithm is the rule that determines the initial split rule[14].

The process starts with a training set consisting of pre-classified records. Pre-classified means that the target field, or dependent variable, has a known class or label: "diabetic" or "non-diabetic. The goal is to build a tree that distinguishes among the classes. For simplicity, assume that there are only two target classes and that each split is binary partitioning. The splitting criterion easily generalizes to multiple classes, and any multi-way partitioning can be achieved through repeated binary splits. To choose the best splitter at a node, the algorithm considers each input field in turn. In essence, each field is sorted. Then, every possible split is tried and considered, and the best split is the one which produces the largest decrease in diversity of the classification label within each partition. This is repeated for all fields, and the winner is chosen as the best splitter for that node. The process is continued at the next node and, in this manner, a full tree is generated.

Artificial Neural Networks (ANN)

Artificial neural networks are defined as information processing systems inspired by the structure or architecture of the brain (Caudill & Butler, 1990). They are constructed from interconnecting processing elements, which are analogous to neurons. The two main techniques employed by neural networks are known as supervised learning and unsupervised learnin*g*. In unsupervised learning, the neural network requires no initial information regarding the correct classification of the data it is presented with. The neural network employing unsupervised learning is able to analyze a multi-dimensional data set in order to discover the natural clusters and sub-clusters that exist within that data. Neural networks using this technique are able to identify their own classification schemes based upon the structure of the data provided, thus reducing its dimensionality. Unsupervised pattern recognition is therefore sometimes called cluster analysis [3,16,17].

Supervised learning is essentially a two stage process; firstly training the neural network to recognize different classes of data by exposing it to a series of examples, and secondly, testing how well it has learned from these examples by supplying it with a previously unseen set of data. A trained neural network can be thought of as an "expert" in the category of information it has been given to analyze. It provides projections given new situations of interest and answers "what if" questions.

There are disadvantages in using ANN. No explanation of the results is given i.e. difficult for the user to interpret the results. They are slow to train due to their iterative nature. Empirical studies have shown that if the data provided does not contain useful information within the context of the focus of the investigation, then the use of neural networks cannot generate such information any more than traditional analysis techniques can. However, it may well be the case that the use of neural networks for data mining allows this conclusion to be reached more quickly than might ordinarily be the case.

DATA PREPERATION

The Montana State University Billings Foundation is a non-profit organization founded to enhance the overall quality of the academic foundations at a university. This is accomplished through a combination of solicitation, investment and management of monetary support for university programs.

As primary source of funding for the Foundation is donations made by alumni, it is important to identify which factors will predict the probability of donations to the Foundation. A database of alumni is maintained by the Foundation, tracking such basic information as name, age, marital status, degree, address, and total amount of donations made to the Foundation.

The data used for this study was obtained from a university foundation. The alumni dataset acts as an address bank, and tracks such additional information as gender, birth date, marital status, class, degree, major, and total donations. Additional information regarding current business, position, and salary are also tracked. While these may be predictors of one's ability to contribute to the Foundation, the information collected is sporadic, at best, which may skew the results of the data model. Thus, these factors will not be introduced as independent variables.

One important variable that was not included in the information is age. Because a birth date is provided for the majority of entries, a column calculating was be added. This was done by utilizing the YEAR()

function available in Excel. All entries that did not have birth date information recorded were deleted, along with any entries not containing degree information.

The remaining dataset contained 5,993 lines of data, and included independent variables of gender, age, marital status, and degree to predict the dependent variable of total $ (contributions). The other variable information was not used for modeling in the data application.

Before applying data mining to our data set, we used feature selection to determine those variables or features which are more important or closely related to the dependent variable. Feature selection methods not only help us avoid noise in the dataset, but also helps us figure out which variable is more important to predict.

	Best predictors for categorical dependent var	
	Chi-square	p-value
Years since Grad.	1277.077	0.000000
City	1108.592	0.008604
Major	989.433	0.000000
Degree	500.767	0.000000
BSED	96.405	0.000000
Salary	41.341	0.001366
Gender	33.567	0.000000

Figure 1

The results of applying feature selection of STATISTICA's Data Miner software are listed in Figure 1. They are ordered top to bottom on the basis of highest chi-square to lowest. The most important variables for future prediction in our project are: Years Since Graduation, City, Major, and BSED. These predictors will be further examined using a wide array of data mining and machine learning algorithms available in STATISTICA'S Data Miner.

CLASSIFICATION AND REGRESSION TREE

The basic design is that of a binary decision tree. The C&RT algorithm is a form of decision tree that can be used for either classification or estimation problems. C&RT algorithms continue making splits along the ranges of the predictor variables until some stopping function is satisfied. In our case we have to predict who is willing to donate to the foundation and the split keeps going on until we actually reach our target variable.

	Predictor importance (Split Input Data into Tr Response: Donation	
	Variable Rank	Importance
Years since Grad.	41	0.410567
BSED	4	0.038896
Degree	50	0.500191
Major	100	1.000000
Gender	4	0.042370

Figure 2

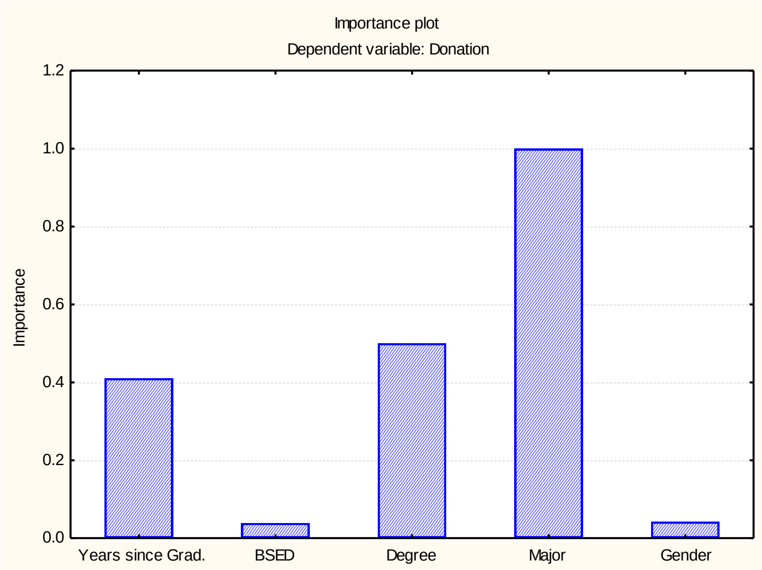

Figure 3

According to Figure 2 and Figure 3, which show the results of applying the classification tree, most important variables for prediction are:

1. Major
2. Degree
3. Years since graduation.
4. Gender
5. BSED

	Classification matrix 1 (Split Input Data into Training and Testing Samples (Classification)) Dependent variable: Donation Options: Categorical response, Analysis sample			
	Observed	Predicted 1	Predicted 0	Row Total
Number	1	866	1053	1919
Column Percentage		68.14%	25.73%	
Row Percentage		45.13%	54.87%	
Total Percentage		16.14%	19.63%	35.78%
Number	0	405	3040	3445
Column Percentage		31.86%	74.27%	
Row Percentage		11.76%	88.24%	
Total Percentage		7.55%	56.67%	64.22%
Count	All Groups	1271	4093	5364
Total Percent		23.70%	76.30%	

Figure 4

The Classification (Confusion) Matrix in Figure 4 indicates that the classification model is more accurate in classifying non-donors (76.30 %) compared to classifying donors (23.70 %).
In addition, from the decision tree we can induce the following general rules to predict who is most likely to donate to the Foundation in the future:

1. The longer a graduate has been graduated, the more they are willing to donate.

2. Major, Degree and BSED are also most important variables to use in our prediction.

NEURAL NETWORK ANALYSIS

Neural networks function relatively similar to the human brain. Both the human brain and neural networks are operated by neurons, the only difference being that the human brain is controlled by biochemical processes whereas the neural networks that are binary digits. The activation process of the neural network can be either linear or logistic, depending on whether it is a numerical estimation problem (linear) or a classification problem (logistic). The neural net architecture may contain one or more middle layers that help model nonlinear relationships by assigning weights to the connections between the middle layer and the input and output nodes.

The results of Classification with neural network for the Foundation dataset are listed in the Figure 5. 2429 cases were predicted accurately as being donors, 555 were predicted accurately as non-donors, 362 cases were predicted to be donors and were not, and 997 were predicted to be non-donors and ended up donation. One more time the prediction for non-donors is better than the prediction for donors weighing in at 87% predicted correctly for non-donors versus 35% predicted correctly for donors.

	Donation (Classification summary Samples: Train	
	Donation-0	Donation-1
Total	2791.000	1552.000
Correct	2429.000	555.000
Incorrect	362.000	997.000
Correct (%)	87.000	35.000
Incorrect (%)	12.000	64.000

Figure 5

CONCLUSION

The performance results of classification and regression tree are very close with artificial neural network, the latter yielding the best performance. However, the 35% correct classification of donors is not good enough. One reason for this outcome is that the ration of donors to non-donors is small in the data set. In future studies, once more data is collected, equal number of donors and non-donors should be used in the analysis. This would lead to much better results in accurately classifying the donors.

REFRENCES

References are available upon request.