## CMPE 255 Significant Paper Report

#### Group 8

Chao Pi 013785265, Xunan Dai 013766441, Ryan Haoyuan Wang 013768209

**What the Article was About**

The purpose of this study is to propose and form an efficient data mining framework for customer relationship management (CRM) systems that can produce prediction on customers’ decision on future campaigns. CRM is a popular technology aim to boost business by improving customer relationships with the firm. Therefore, the proposed framework is set to analyse customers’ behavior, and accordingly to generate predictions on customers’ future decisions on certain campaigns based on their relationship with the firm and their own certain characters.

The given dataset that used to make this analysis and model building is a dataset from a Portugeese bank that includes 17 campaigns between 2008 to 2010. The customers were contacted by phone and given a long-term deposit application. Their decision were recorded along with information like their previous relation with the bank and their current financial status. Dataset is preprocessed with Rattle analysis, manual feature selection and later CRISP-DM iterations leaving 17 out of 58 attributes for model building. The final used dataset only accounts for 10% of the total instances. Two classification methods which are Multilayer Perceptron Neural Networks and Naïve Bayes are used to build models for comparison in this study. 10-fold validation is used to create partitioning of the dataset for a more accurate model building. The prediction accuracy of loyal customers is compared with metrics like processing time, confusion matrix and ROC analysis.

**What the Study Found**

The two classifiers Multilayer Perceptron Neural Networks and Naïve Bayes model produced the accuracy rate with 88.63% and 87.97%, respectively. The accuracy is not significant different, but the report states the MLPNN is performing better than NB model based on the accuracy rate. Other evaluation metrics are used including sensitivity, specificity, TPR, FPR and ROC areas. In the study provided result table of two classifiers, the NB model’s TPR and ROC areas are larger than MLPNN’s result which indicates the better predictions on different metric scales, but the report did not mention this part. The final conclusion the study gives is that the MLPNN is the best model in providing prediction based on the high predictive performance accuracy.

**Opinion on the Study**

The proposed framework gives a general view of how CRM customer prediction can help firms with business decision making. Whereas, the methods implemented in this paper with data preprocessing, model building and evaluation are some basic data mining methods that should be conducted with more depth to reveal the hidden knowledge. The data preprocessing phase can be implemented in a more sophisticated approach with feature reduction analysis like correlation analysis and principal components analysis. Instances with missing value can be predicted or be replaced by mean value instead of discarding them causing information loss. There are more models to be investigated that suits this dataset like support vector machines, decision tree and KNN. The validation and evaluation methods used in the study are the most common methods that are best showing the result of the models, so they will be kept for future investigation of the expansion of the framework.

**Key Learning**

* Basic structure of data mining framework for CRM analysis
  + Six steps including customer identification, business/domain understanding, data preparation, model building, model evaluation and visualization.
* Data preprocess procedures on customer’s information
  + Attributes relevant to the business decision making using Rattle analysis and CRISP-DM iterations to reduce the features.
* Evaluation methods to compare the result
  + Confusion matrix, ROC areas and time to build models are used to evaluate the performance of the models.
* Classification models selection
  + The report implemented two models and compared their results. Later in the conclusion, it suggested Neuro fuzzy classifiers and ensemble of classifiers for future investigations.