

ENHANCING CUSTOMER RETENTION THROUGH DATA MINING TECHNIQUES

Alexiei Dingli¹, Vincent Marmara² and Nicole Sant Fournier³

¹Department of Artificial Intelligence, University of Malta

² Faculty of Economics, Management and Accounting

³Department of Artificial Intelligence, University of Malta

ABSTRACT

A top priority in any business is a constant need to increase revenue and profitability. One of the causes for a decrease in profits is when current customers stop transacting. When a customer leaves or churns from a business, the business loses the opportunity for potential sales or cross selling. When a customer leaves the business without any form of advice, the company may find it hard to respond and take corrective action. Ideally companies should be proactive and identify potential churners prior to them leaving. Customer retention has been noted to be less costly than attracting new customers. By analysing the data analytics, companies may analyse customer behavioural patterns and gather insight on their customers. These insights will help to identify profitable customers and improvements in their business process thereby increasing customer retention. This paper demonstrates the power and value of data, companies may attain through data analysis and data mining. Two techniques have been implemented Random Forest and Logistic Regression attaining 94% and 76% respectively. Through data analyses and data mining, retail businesses may adopt a proactive approach in identifying possible churners. The novelty of this paper is the concept of implementing deep learning algorithms in addition to data mining techniques. Through this, marketing campaigns may be targeted to specific profitable customers who might leave, therefore increasing profitability and reducing marketing campaign costs.

KEYWORDS

Customer Churn, Data Mining, Data Analysis , Retail, Supermarket

1.INTRODUCTION

Customers are one of the companies most valuable assets. With this in mind, businesses need to provide customer's incentives in order to retain and continue to grow the business. A negative interaction with a customer might mean that customers might churn [1]. An important aspect within the business is to have a good understanding of customer's needs, whereby holistic views of their patterns may be analysed. Across industries, data on customer demographics, historic transactional data and in some cases customer satisfaction (Net Promoter Scores) are stored and may be used for analysis [2]. Researchers discuss that revenues and margins increase if the right customers are retained in comparison to attracting new ones [3]. With data mining techniques, data may be processed and analysed to identify trends and behavioural patterns [4]. Researchers further discuss that it is becoming more fruitful to apply data mining techniques rather than standalone statistical models to predict outcomes [5]. The reason for this, is due to the constant increase in data. This paper discusses the data preparation and exploration required to enhance and optimise business processes to decrease customers with a high propensity to churn

2. BACKGROUND AND LITERATURE REVIEW

2.1. Customer Analysis

2.1.1. Customer Classification

Prior to applying classification techniques, customers are categorised into groups based on similar behavioural patterns. Recency, Frequency and Monetary (RFM) [6], [7], [8] is the main technique used to understand customer patterns.

Within the grocery industry researchers [6], [7], [8] implement RFM within their model to predict churn. When modelling churn, [7] pays attention to the score attained using these three variables. The advantage of utilising RFM is that it takes into perspective behavioural trends.

2.1.2. Churn Definition

Within the *Grocery Retail Industry*, the identification of the exact moment a customer will churn is hard to define, as customers do not stop purchasing from the store abruptly, but rather they partially defect -in other words they gradually move to a competitor [10]. Based on this definition, [7] predicts churn for the grocery industry based on the first product category the customer has purchased. This provides a good indication on whether the customer is loyal towards the store or not. Furthermore, churn may be viewed from a product perspective whereby a customer who has been loyal to a particular brand has stopped purchasing this brand and started to buy a different one. Reasons for this could be the new brand has better marketing schemes and competitive prices compared to other brands [11].

2.2. Data Analytics

Customer demographics and historic purchases are part of the data companies currently collect on customers [2]. At times, although the data is available within sources, it is not used for analysis purposes. [12] refers to this as dark data, as it is data which may have high values for decision making. Various techniques are implemented to extract both known and hidden knowledge. At times, patterns and trends are not easily visible.

To obtain this knowledge data mining is applied [12]. Prior to implementing data mining techniques, it is essential that data is cleaned and transformed into predefined structure to be fed into the model. Keeping the key users in mind, [13] proposes that a data warehouse is implemented so as to have easily access to data. This structure gives users power to 'slice and dice' and analyse their data to assist in their day to day decisions. The process to do this is depicted as Extract, Transform and Load (ETL) [14].

2.3. Data Mining Techniques

The most popular techniques used to predict customer churn within retail industries are Logistic Regression, Decision Trees and Random Forest.

Logistic Regression, as described by [15] is a method used to test a hypothesis pertaining to the relationships between categorical variables. As stated by [16] Logistic Regression is easy to use and provides quick and robust results. This method is used as the outcome for the model is the probability of achieving an output of 1 or 0 (e.g. 1 if the customer is classified as a churner, 0 if the customer is classified as a non-churner).

Random Forest is an ensemble classification method, which may be used to overcome the issues encountered by the decision tree [18]. Multiple trees are built for the training dataset utilising

randomness in the tree learning process. From the classification results attained from each tree, a consensus of all the votes is taken. The randomness is based on bagging, which is a method used to increase accuracy. Here the class with most votes is the prediction class. In this case, multiple trees are generated which predict the class label. This is a robust method and can overcome the issue over fitting as it is dependent on the results attained by the individual classifiers

3. METHODOLOGY

The data collection was a task in itself, as the data used within the model would ideally be realistic data. Therefore, various local supermarkets were contacted to examine the possibility to access their data. One of the local supermarkets responded and accepted to provide their data for this research. The data provided includes:

- Item data - These include: Item, Department, Group, Category code, supplier code and price
- Customer data - Customer Code (this has been converted to retain the customers identification) and Customer Locality.
- Transactional Data - Transaction Number, Time, Customer ID, Item, Quantity and Sales Value.

3.1. Extract Transform and Load

Applying the Kimball methodology, the data available traverses through a process of Extract Transform and Load (ETL). In this process, the data is extracted from the database, transformed into dimensions and measures whereby finally loaded into the data warehouse.

3.1.1. Extract

The data was originally provided in separate Microsoft Access files. For ease of use, each file is converted and imported into a MYSQL database. The database consists of six tables pertaining to details on the Category, Item, Department, Customer, Transaction and Transaction Entry for each year.

This process is responsible for extracting data from the database. This is done by executing scripts which are used to extract the data from the source database. The scripts required are for the fact table and various dimensions.

3.1.2. Transform

Within this process, rules and aggregations are created. Within the transformation, customer names, stock brand, category and department descriptions are eliminated from the dataset for data security. Furthermore, counts for customers frequency and number of receipts are pre-calculated so that on reading data counts are pre aggregated.

3.1.3. Load

The data is loaded into a data warehouse which has three dimensions and one fact table. The schema for this is a star schema whereby the dimensions Customer, Stock and Time have the primary key and indexes set up. The fact table contains the Ids for the stock, customer and time together with values including sales quantity and value.

3.2. Data Exploration

Descriptive and inferential statistics are computed on the dataset to gather an understanding of the available data. Descriptive statistics are carried out based on the customer demographics including Location, District and Age. A similar process is carried out for the stock data.

Within the dataset there are transactions which include a card number whilst others do not. For the purpose of this analysis customers with a card number are required. When analysing sales and frequency by customer locality and age, values included pertain to transactions with a customer card.

Recency, Frequency and Monetary (RFM) values per month and year and analysed as depicted by [6] and [7]. To visualise the frequency and monetary values by various dimensions including district, a histogram is plotted as seen in Figure 1. This depicts the Northern Harbour as the district where most customers originate from. Comparing the frequency across the years of data, this district remains the most popular.

A similar pattern is seen for the monetary values by district in Figure 2, whereby the most popular district is the Northern Harbour district followed by the Western district.

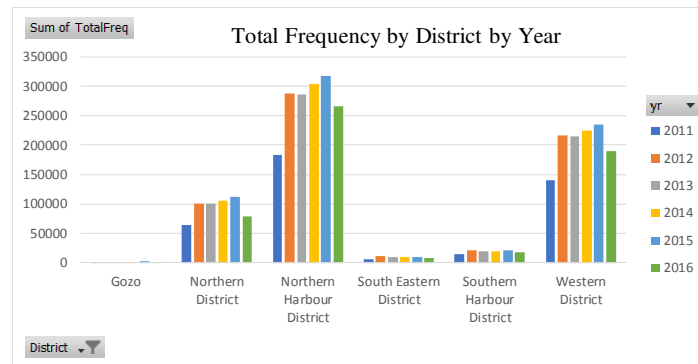


Figure 1: Frequency of visits to Supermarket by District.

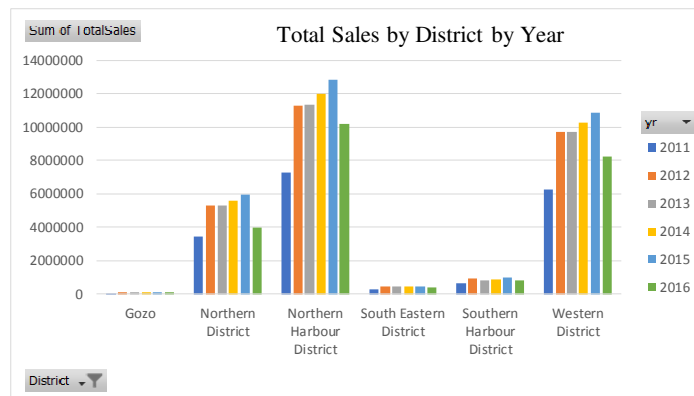


Figure 2: Monetary values by District by year

3.3. Data and Parameter Selection

Having transformed and analysed the data as discussed in the previous section, the next exercise requires the formation of parameters which will be then passed to the various algorithms.

Keeping in mind the objective of the project, the parameters created revolve around the customers buying trends. Two datasets are created based on the RFM values, the first dataset is based on four months of historic RFM data whilst the second dataset includes 11 months of historic data. For the first dataset, four months prior to the predictive month are selected to identify churn. The number of months has been selected as through this period the majority of customer behavioural patterns are captured. The second dataset reinforces the use of historic data, whereby data from January to November is used to predict customers with a propensity to churn. Within the Table 1 the list of all the features available for both datasets are listed. N represents the month within the dataset.

Table 1: List of features available per dataset

Feature Name	Feature Description
Age	Age of Customer.
AgeBucketId	Unique Id for age bucket.
Locality Id	Unique Id for locality.
District ID	Unique Id for district.
MnthNFreq	Frequency for customer for the month.
MnthNDateDiff	Date difference between end of month and last transaction date.
MnthNMonetary	Monetary values spend for the month.
DeptMax Qty	Max quantity for Department.
DeptID	Unique Id for the department
CatMaxQty	Max quantity for the Category
CatID	Unique Id for the department

The first four features parameters relate to the customers descriptive data. These values remain static across the months of available data. The features defined as MnthNDateDiff, refer to the last time a customer has transacted with the supermarket for that month.

Frequency indicates how often they frequent the supermarket. This is computed by counting the number of receipts each customer has per month. The next parameter is the Monetary value which denotes the total value spent per month.

3.4. Churn Defined

For the purpose of this research, Churn is defined as a customer who stops transacting with the supermarket. Customers who decrease in sales will be classified as non-churners. Customers who leave the supermarket in the following period are categorised as churners. So as to identify churners, the dataset is divided into two time frames or time windows. The first window is the predictive window which identifies active customers. Activity is defined by the customers having transactions within this period. Customers which have activity within the first window are tagged as non-churners, whilst the remaining customers are marked as churners. The latter are eliminated from the analysis as they have already churned. The next window, the churn assessment window classify the remaining customers are churner or non-churner. If customers transact in this period they are non-churners, whilst if no transactions are seen then they are churners.

3.5. Implementation

Three algorithms are implemented on the discussed datasets. To abstain the model from over fitting or to avoid creating a bias to the majority class, the dataset is balanced having the same number of examples available in both classes. Furthermore, outliers are removed from the dataset prior to dividing the dataset for training and testing. The apportionment for this is by randomly dividing the dataset into 75:25 for training and testing. Once the model is trained on the training dataset, the unseen data is given to the model excluding the label for testing. The predicted label is appended to the data frame where the actual label is available and compared.

3.5.1. Feature Selection

For each of the datasets, feature selection is applied so that only the statistically significant variables are passed to the algorithms. The feature selection algorithm applied is the greedy backward elimination, whereby statistical measures are applied so as to generate the P value. The features indicating the least statistically significant are removed through each iteration.

The first dataset, pertaining to the four months of historic data contain 28 features together with the class label. Through the implementation of feature selection, the number of features decreased to 10 leaving features with a P value less than 0.05.

The second model contains 48 features referring to 11 months of historic data per customer. Following the same approach adopted for Model 1, the number of statistically significant variables are 11 features and the class label.

3.5.2. Logistic Regression

Since the problem at hand is a two-class binary classification (1 or 0), logistic regression is applied to the dataset. Following the feature selection using the backward elimination method only variables with a p value of 0.05 or lower are retained. As seen in table 2 and 3 each variable is given a P value which indicates importance of the variables to predicting churn. This represents the relationship between the dependent and independent variables.

To fit the logistic regression model, the glm (Generalised Linear Model) function is used. To compute this, the prediction class is given followed by the features identified. Next the training dataset is defined together with the family for this model which is the Binomial. This is selected as the prediction class has an outcome of two: either churn or non-churn.

To test the logistic regression, the predict function is used. This requires three parameters, the first refers to the model which was used for training. The second parameter is the testing dataset and the third parameter is the type. This refers to the type of prediction required, since this is logistic regression the type is set to 'response'. This will generate a probability for each customer.

To classify whether the customer is a churner or not, a threshold of 0.5 is used. If the probability is greater than the threshold the customer is classified as a churner, whilst if the probability is less than 0.5 then the customer is classified as a non-churner

3.5.3. Random Forest

Similar to Logistic Regression, the two datasets are implemented using features identified as significant from implementing Random Forest. To implement the ensemble algorithm the

random forest function is used. The parameters required for this include the name of the class label in this case 'Churn', the training dataset which include all the training examples excluding the class label and the number of trees to be created.

For the first dataset, the significant variables for predicting churn are: frequency of the previous month, monetary values of the 10 and 8th month, date difference for the 8th month, frequency values for the 8th and 9th month, date difference for the 10th month, Locality Id, Customer, date difference for the 9th month and the customer age.

For Model 2, it is evident that the features pertaining to previous months are significant to predicting churn. To evaluate the prediction labels, the test dataset excluding the actual label are passed to the model using the predict function. Once the predicted labels are generated they are placed into a table together with the customer code. These are then merged to the actual label by the customer code and compared.

4.1. Results and Evaluation

4.1.1. Logistic Regression

As depicted in table 2, a comparison of the evaluation metrics for Model 1 and Model 2 are seen. Comparing the results for the two models it is evident that Model 2 has attained greater results than Model 1. The first metric sensitivity, calculates the capability of correctly identifying customers who will not churn. The results obtained for Model 1 and Model 2 are 67% and 72% respectively.

Similarly, specificity tests the ability of the model in correctly classifying customers who will churn. The results acquired is of 79% for both models. Precision defines the classifiers exactness, as this calculates the number of predicted non churners division by the total number of actual non churners. Model 1 obtained 0.67 whilst Model 2 attained 0.72. To understand the completeness, recall is used whereby the results achieved are 0.76 and 0.77 for Model 1 and Model 2 respectively. The F1 measure is computed by taking into consideration the customer misclassified as churners and non churners. Model 1 resulted in a 0.71 whilst Model 2 attained 0.74.

Table 2: Logistic Regression Results

Evaluation Metric	Model 1 Result	Model 2 Result
Sensitivity	0.67	0.72
Specificity	0.79	0.79
Accuracy	0.73	0.76
Precision	0.67	0.72
Recall	0.76	0.77
F1	0.71	0.74

4.1.2. Random Forest

Sensitivity for Model 1 and Model 2 attained 90% and 93 %indicating that non-churners are being correctly classified for this algorithm. The classification of churners is depicted in the results achieved for Specificity, both model obtained similar values of 93% and 94% for Model 1 and Model 2. Reviewing the values attained for Precision, Recall and F1 indicate that both models can correctly classify churners and non-churners. The results obtained for the two models with reference to the three-evaluation metrics discussed are of above 0.90.

Table 3: Random Forest Results

Evaluation Metric	Model 1 Result	Model 2 Result
Sensitivity	0.90	0.93
Specificity	0.95	0.94
Accuracy	0.93	0.94
Precision	0.90	0.93
Recall	0.95	0.94
F1	0.93	0.94

Analysing the evaluation metrics by dataset, Model 2 for both Logistic Regression and Random Forest obtained higher results than Model 1. Evaluating the results, it is evident that Random Forest achieved higher results in correctly classifying churners and non-churners.

5. CONCLUSIONS

This paper discusses the importance of available data supermarkets retain together with the best parameters required to predict churn. From the techniques implemented it is evident that analysing customers previous behaviour patterns is pivotal to identify churn. This is due to the fact that customers within the retail industry are gradual churners. So as to test this model, data from a local supermarket was provided. This was meaningful as the results attained from the model are realistic attaining 94% with Random Forest and 76% when Logistic Regression was implemented.

The customers who have been identified as churners will be passed on to the Sales and Marketing team of the local supermarket. A deeper understanding as to why these customers might churn will be carried out. Following this, using the feedback attained from the customers, a retention campaign will be proposed to target the probable churning customers.

Future work includes the implementation of RBM to predict customers with a high propensity to churn. Furthermore, compare the results attained by the machine learning techniques to the results of the RBM. By using the parameters and model designed in this paper, companies within the retail industry may produce focused marketing campaigns to plausible churners. Therefore, reducing marketing costs and the possibility of retaining customers which if successful will increase sales and profits.

REFERENCES

- [1] 'The Science behind Customer Churn'[Online] Available: <http://financeinbusinesslife.info/thescience-behind-customerchurn/>
- [2] G.D. OlleOlle, S. Cai, "A Hybrid Churn Prediction Model in Mobile Telecommunications Industry", International Journal of eEducation, e-Business, e-Management and eLearning, Vol 4, No 1, 2014
- [3] C.Jie,Y.Xiaobing,Z.Zhifei,"Integrating OWA and Data Mining for Analyzing Customers Churn in E-Commerce", The Editorial Office of JSSC and Springer- Verlag Berlin Heidelberg, Vol 28, pp 381-391 2015.
- [4] <https://www.ibm.com/developerworks/library/ba-data-miningtechniques/>, Developer works, Accessed: November 2016
- [5] E. Siegel, Predictive Analytics, The power to predict who will click, buy, lie or die. Wiley , 2013
- [6] K. Coussement, K W. De Bock, Customer churn prediction in the online gambling industry: The beneficial effect of ensemble learning, Journal of Business Research, No 66, pp 16291636, 2013.

- [7] V. Migueis, D. Van den Poel, A. Camanho and J. Falcao, Modeling partial customer churn: On the value of first product-category purchase sequences, *Expert Systems with Applications*, no. 39, pp. 11250-11256, 2012.
- [8] O.Ali,U.Ariturk,"Dynamic churn prediction frame work with more effective use of rare event data: The case of private banking", *Expert systems with Applications*, No 41, pp 78897903, 2014.
- [9] E. Castro, M. Tsuzuki, "Churn Prediction in Online Games Using Players' Login Records: A Frequency Analysis Approach", *IEEE Vol 7*, 2015.
- [10] W. Buckinx, & D. Van den Poel," Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting." *European Journal of Operational Research*, 164(1), 252268, 2005.
- [11] S.L. Chan, W.H. Ip, V. Cho, "A model for predicting customer value from perspectives of product attractiveness and marketing strategy", *Expert Systems with Applications*, 37, 12071215, 2010.
- [12] A.T. Jahromi, Unpublished Thesis:" Predicting Customer churn in Telecommunications Service Providers", Lulea University of Technology, 2009
- [13] R. Kimball, M. Ross, "The Data Warehouse Toolkit", Third edition, Wiley, 2013
- [14] S. Sridharan, B. Purcell How Analytics Drives Customer Life-cycle Management. *The Customer Analytics Playbook*. 2015
- [15] B. Huang, M. T. Kechadi, B. Buckely, "Customer churn prediction in telecommunications", *Expert System with applications*, No 39, pp1414-1425, 2012.
- [16] B. Baesens, G. Verstraeten and D. Van den Poel, Bayesian network classifiers for identifying the slope of the, *European Journal of Operational Research*, vol. II, no. 156, pp. 508-523, 2003.
- [17] C. Bishop, *Pattern Recognition and Machine Learning*, Singapore: Springer, 2006.
- [18] J. Han, M. Kamber, J. Pei, "Data Mining Concepts and Techniques": The Morgan Kaufman Publishers: 3rd Edition 2012.
- [20] J.M.Tomczak, M. Zieba, "Classification Restricted Boltzmann Machine for comprehensible credit scoring model", *Expert Systems with Applications*, Np 42, pp 1789 - 1796, 2015.

Authors

Prof.AlexieiDingliB.Sc IT (Hons)(Melit.),Ph.D. (Sheffield),M.B.A (Grenoble)

Dr AlexieiDingli is a Senior Lecturer of Artificial Intelligence within the Faculty of ICT at the University of Malta. He is also one of the founder members of the ACM student chapter in Malta, founder member of the Web Science Research, founder member of the International Game Developers Association (IGDA) Malta and of the Gaming group at the same University. He pursued his Ph.D. on the Semantic Web at the University of Sheffield in the UK under the supervision of Professor Yorick Wilks. While there, he worked on various large projects but his major contribution can be attributed to the Advanced Knowledge Technologies project, one of the largest Interdisciplinary Research Collaborations (IRC) funded by the Engineering and Physics Research Council (EPSRC). For this project he created two systems which were rated World Class by a panel of international experts whose chair was Professor James Handler (one of the creators of the Semantic Web). These systems were later used as a core component of the application that won the first Semantic Web challenge (2003). His recent work in Mobile Technology and Smart Cities (2011) was also awarded a first price by the European Space Agency. He has published several posters, papers, book chapters and a book in the area. For four years, he also worked as a Senior Manager in a large government corporation where he got insight into the needs, potential and deficiencies of digital natives. During this time, he also pursued an MBA with the Grenoble Business School in France specialising on Technology Management



Dr. Vincent Marmara B.Sc.(Hons) (Melit), M.Sc (Sheff) Ph.D (Stir)

Vincent Marmara is a Statistician and Researcher by profession. He obtained his first degree in Statistics, Operational Research and Mathematics at the University of Malta. He later advanced his studies by obtaining Masters of Science in Statistics at Sheffield University, UK. Furthermore, he has obtained his PhD in Mathematics (Statistics) from the University of Stirling, Scotland. He was entrusted with numerous research projects both at a national and international level. He led research groups and analyzed data to a high-level scientific extent. Vincent has over 10 years of experience in the Remote Gaming Industry as a Business Intelligence Analyst and Consultant. He occupied several key important roles such as; member of the faculty board of science (University of Malta), President of the Science Student Society, Financial Officer of the National Youth Council, Deputy CEO and Chief Regulatory Officer within the Malta Lotteries and gaming authority. Vincent is a lecturer at the department of Management (FEMA) at the University of Malta and a fellow of the Royal Statistical Society (UK). His main research interests are research and management in the gaming industry, epidemiology and health research, sampling surveys, regression models and Bayesian analysis.



Nicole Sant Fournier B. Sc.(Hons) (LSE)

Nicole Sant Fournier is a Business Intelligence consultant. Her first degree was in Information Systems and Management at London School of Economics. She now has started her Masters in Artificial Intelligence with University of Malta. Throughout her career, she has had the opportunity to understand and meet key people within industries including the Financial, Retail and Distribution Sector as a Business Analyst and Business Intelligence Consultant. Having analysed and implemented solutions for various industries Nicole has an understanding of the business problems companies encounter. Therefore, her thesis focuses on solving the problem of churn predicting using data mining techniques.

