A hybrid machine-learning approach for loyalty prediction

Harold Lee1, Dr. Ming Jiang2

1Faculty of Computer Science, University of Sunderland

bh46lu@student.sunderland.ac.uk

2Faculty of Technology, University of Sunderland

ming.jiang@sunderland.ac.uk

**Abstract:** Customer loyalty prediction is one of the most common application of machine learning in CRM, and many research studies have tried to compare effectiveness of different machine learning techniques by developing prediction models with behavioural variables of customers. In selecting features in model development, due to the simplicity and effectiveness, behavioural RFM attributes are commonly used in predicting the customer lifetime value as a measure of loyalty. However, RFM focuses on the purchase behaviours of customers only, thus overlooking the effect of some other important factors to loyalty, for example, customer satisfaction and product experience. On the other hand, prediction models in previous studies are typically designed using either supervised or unsupervised technique only. Thus the possible incremental value of hybrid model coming two learning technique is overlooked. In this paper, a two-stage model including unsupervised K-Means clustering and supervised classification model has been developed. The model is trained using not only RFM attributes capturing the effect of behavioural factors to loyalty, but also customer satisfaction and product attributes attempting to capture the effect of attitudinal factors.

***Keywords.*** *K-Means; Unsupervised learning; classification model; supervised learning; RFM; CRM*

1. INTRODUCTION
   1. Customer loyalty

Customer loyalty was a well discussed marketing concept and had received high level of attention from many businesses since 1980s. Many companies had developed customer relationship management programs with the objective to enhance customer loyalty (Pitta et al., 2006) from marketing point of view. There is also a proven relationship between customer loyalty and company performance across industries namely banking, hotel and retail (Liu and Wang, 2017; Ramanathan, Subramanian and Vijaygopal, 2017).

Customer loyalty is generally defined as the intention of repurchasing products and services (Pi & Huang 2011). Bose and Rao (2011) suggested that loyalty is the customer’s commitment to do business with a particular organization which effects in repeat purchases of goods and services of that organization consistently in future. Apart from the behavioral consequence of loyalty in a form of purchases, loyalty could also result in recommending the goods and services to friends and associates.

Besides the behavioural approach of defining loyalty, there is also the attitudinal approach. Attitudinal approach describes loyalty in forms of consumer attitudes, preferences and dispositions towards brands which explain the motives leading to the repurchase behaviours (Antonios, 2011; Shih-i, 2011).

* 1. Lifetime value as a measure of customer loyalty

Because of the difficulty of collecting large-scale attitudinal data, most empirical research papers focus on the behavioral approach which defines loyalty as a propensity to repurchase a brand (Trinh, Anesbury and Driesener, 2017; Liu et al., 2019). Since loyalty is reflected in a form repurchase or a propensity to buy a brand in behavioural approach, loyalty could then be measured quantitatively with customer lifetime value (CLV), which is calculated as a sum of the present value of all profits generated from a customer’s future purchases over a full life of relationship with a company (Gupta, Lehmann and Stuart, 2004; Rust, Lemon and Zeithaml, 2004). Recent empirical studies supported the relationship between loyalty and CLV. Studies by Zhang, Dixit and Friedmann (2010) and Chen et al (2012) supported the hypothesis that customer loyalty was positively correlated with customer lifetime value by proving customer revenue and customer retention were driven by customer loyalty.

* 1. **Estimating customer lifetime value using RFM**

To measure CLV, we can estimate the revenue generated from future purchases by using prediction techniques. The most common one according to Gupta et al. (2006) is RFM model. RFM model is an effective tool for determining high potential customers by exploring the quantitative characteristics of customers. RFM model was widely used by marketers for customer segmentation, customer loyalty and response prediction (Song et al., 2017; Cheng & Chen, 2009; Coussement & De Bock, 2013). The three components of RFM are called recency, frequency and monetary (Yeh, Yang and Ting, 2009) Recency represents the time interval between now and last transaction time of a customer; Frequency represents the number of transactions a customer made in a period of time; Monetary indicates the total amount of spending of a customer made in a period of time.

* 1. Other factors influencing customer loyalty

RFM models seem to be too simplistic and fail to capture the effect of attitudinal and other industry specific factors to customer loyalty. One factor that is believed to be directly or indirectly related to customer loyalty is customer satisfaction. Customer satisfaction refers to customer feelings or customer reaction to the state of fulfillment of their expectation and needs through the services or products (Hallowell, 1996; Oliver, 2006). There are many empirical studies have proven the positive relationship between satisfaction and loyalty in banking and even B2B industries (Hallowell, 1996; Chandrashekaran et al.,2007).

Some research studies used category or product purchasing data in customer segmentation model to capture the industry specific characteristics. For example, in a recent study, Brito et al. (2015) had segmented the shirt customers for an online fashion retailer based on style, colors and fabric. Heilman and Bowman (2002) had shown in the study of baby product industry that segmentation could be developed by considering purchase behaviour in multiple categories together.

1. RESEARCH METHOD

This section aims to introduce the proposed research method for predicting the loyalty of customers. The research method first involves the application of K-means clustering, then followed by a classification model built from different learning techniques.

* 1. K-Means Clustering

Clustering is the automatic process of grouping data of similar attributes into same group, while dissimilar data into other groups. K-means clustering is a very well-known algorithms that has been used extensively in various business applications, one of which is customer segmentation. It works by automatically partitioning a dataset into k groups with the rule of nearest means, which assigns data sample to the cluster with the closest centroid “Mean” (Zalaghi and Varzi, 2014).

To select the right number of clusters (K value), CH Index is used to compare the validity of different K values. The method is invented by Calinski & Harabasz in 1974 based on the relationship between “the sum of distances within cluster” and “sum of distances between clusters”. The CH Index formula is displayed in Figure 1, where *N* is the number of observations, *K* is the number of clusters, and *BCSM* and *WCSM* are the between- and within-cluster sums of squares, respectively. According to CH Index, the correct number of clusters is determined by the K value that can maximize CH Index.

A close up of a logo

Description automatically generated

**Figure 1. CH Index’s Formula**

* 1. **Classification models for prediction**

Classification method is deployed in the second stage to build the prediction model using the K-means clustering results and other customer features. Contrast to the unsupervised nature of clustering, classification is a supervised learning technique to identify which “class” an observation should belong to, based on a training set of data containing observations with known “class” membership (Alpaydin, 2020). The learning techniques used for classification are described in the following section.

**Decision Tree:** Decision tree learning is among the most popular learning technique for classification model. Decision tree learning method uses a tree-based model that starts from splitting data samples based on values of attributes (represented in the branches) until the “decision” about all data samples’ target values are decided (Wu et al, 2008). In classification problem, the target values are set of discrete class labels and the decision tree is called classification tree”

**Logistic Regression:** Regression is a statistical analysis concerned with describing the relationship between a dependent variable and one or more explanatory variables. There are two key categories of regression models, one is linear regression model and one is logistic regression model. The key distinguishing factor of two models is the target (dependent) variable. When the target variable is a continuous set of values, linear regression should be employed, while logistic regression should be used if target variable is a set of categorical values.

**Random Forest:** The random forest is an ensemble tree method, an aggregation scheme of many individual decision trees. Due to its popularity and its good empirical performance, Breiman’s random forest is one of the most used algorithms, and is often directly called “random forest” (Scornet, 2017). In the random forest model, the training data are sampled randomly to create multiple randomized trees, so that different classification results would be obtained from the multiple randomized trees (Livingston, 2005). The final classification result will be the one with most votes (weighted or unweighted) from randomized decision trees. Figure 2 illustrates the scheme of an un-weighted random forest algorithm.

A picture containing object, clock

Description automatically generated

**Figure 2. Scheme of un-weighted random forest algorithm**

**Boosted model:** Boosted model is also an ensemble tree method but it is different to random forest in the way trees are built and combining results. In boosted model (Schapire, 2003), the model works in a forward stage-wise manner that each tree is built once at a time, with the next tree is built with an emphasis of mis-labelled samples based on the previous tree . The mislabelled samples from previous tree are called weak learner, and by adding weighting to weak learners in next tree, it enhances the training of the model towards those previously mis-modelled.

2.3 Methodology of related research studies as comparison

At my best knowledge, the scope and methodology design of this project has a very comprehensive approach and can generate research contributions to luxury industry, which is a relatively under-researched industry. Table 1 summaries the methodology applied in recent research studies as comparison.



**Table 1. Summary of methodologies used in related studies**

1. DATA
   1. Dataset

The data we use are obtained from a global luxury fashion retailer from Europe. The dataset integrates the enterprise customer transaction data and satisfaction survey data obtained across 4 years from January 2016 to December 2019. There are 32123 rows of data capturing transaction data of customers who made first transaction between January to March 2016 and are from Greater China region including China, Hong Kong and Taiwan. The enterprise transaction customer database stores the historical purchase records of all customers which including customer details, purchased stores, purchased products, transaction date, sales amount etc. While the customer satisfaction survey database contains the responses of survey conducted by the company regarding their satisfaction level towards the shopping experience.

* 1. Feature selection and engineering

The 4-year period of dataset is split into two periods, of which the first period contains the transaction data of 1st year and 2nd year; and the second period contains those of 3rd year and 4th year. The purpose is to use the RFM variables of first 2 years’ period to predict the “customer lifetime value” in form of the next 2 years’ total spending value. The following feature engineering is performed to convert variables to the forms that can be exploited by clustering and classification models.

**CLV Classifier:** The outcome of classification is to identify customers of top 10% CLV in the “future” period. A new classifier “Second 2 Year-Top 10% spend” is derived based on total spend value in the second 2-years’ period.

**RFM attributes:** The last purchase date is transformed to the number of months from the last purchase as “Recency”; the number of transactions made under the same customer ID is aggregated as “Frequency”; the total spend under same customer ID is regarded as “Monetary”. Values are normalized so that it can be processed accurately in K-means clustering

**Product variables:** For each customer ID, product category bought most by value is regarded as “Top spend category”, while product category bought in the very first transaction is “First purchase category”. Figure 3 summarizes the variables trained in the model after feature engineering, and their corresponding data timeframe, while Table 2 describes the data type and meaning of each variable.

A screenshot of a cell phone

Description automatically generated

**Figure 3. Independent variables and classifiers for modelling**

|  |  |  |
| --- | --- | --- |
| **Variable name** | **Data type** | **Description; typical value** |
| ID | Nominal | Customer ID; unique integral number |
| Cust Market Code | Nominal | Residence location of customer; HK |
| First Transaction month | Nominal | First transaction date in YYYY-MM; 2016-03 |
| First Purchase category | Nominal | Category purchased in first transaction; MENS |
| Top spend category | Nominal | Category purchased most in value from ALL transactions |
| First 2 Year – Freq | Numeric | Number of transactions purchased in the first 2 years ie 2016 & 2017 |
| First 2 Year – Spend | Numeric | Total spent amount in GBP in the very first 2 years; 8,999 |
| First 2 Year – Recency | Nominal | Last transaction month in the first 2 years in YYYY-MM: 2017-08 |
| Second 2 Year – Freq | Numeric | Number of transactions purchased in the second 2-year period ie 2018 & 2019 |
| Second 2 Year – Spend | Numeric | Total spent amount in GBP in the second 2-year period ie 2018 & 2019; 8,999 |
| Second 2 Year – Recency | Nominal | Last transaction month in the second 2-year period ie 2018 & 2019; 2019-01 |
| Net Promoter Score | Numeric | The rating given in 10-point-scale towards overall brand recommendation |

**Table 2. Data type and description of the variables from dataset**

1. DATA ANALYSIS
   1. Data analysis process

The application of clustering and classification modelling are performed in following flow (Figure 4) using an analytics software called Alteryx Designer.

1. Clustering customers into different clusters based on RFM variables using K-Means algorithm
2. Based on the K-Means clustering result, select the one (or more) cluster with the highest RFM values
3. Split the data samples of the selected cluster into train and test dataset
4. Run the training dataset in 4 different classification method: Boosted Model, Logistic Regression, Decision Tree and Random Forest
5. Select the variable “Second 2 year - Top 10” as the target classifier, and “Net Promoter Score”, “Top spend category” and “First purchase category” as the independent variables.
6. Run the testing set of data on the 4 models
7. Select the best performing model as the final prediction model by evaluating model performance matrices

A screenshot of a cell phone

Description automatically generated

**Figure 4. Workflow of the research methodology**

* 1. Performing data analysis

### 4.**2.1 K-Means Clustering**

### The very first step of K-Means clustering is to determine the number of clusters ie. K value using Calinski & Harabasz Index (CH Index). Figure 7 shows that CH Index is maximized at 4-cluster K-Means, so in this project K value of 4 is selected.

### A screenshot of a video game Description automatically generated

**Figure 5. CH Indices of different value of K (Number of clusters)**

### 4.2.2 Standardizing RFM variables

### Once the number of clusters is determined, the clustering analysis is started on the Alteryx Designer. Since the 3 RFM variables are not measured in same unit and have very different ranges of values (Table 3), the values of RFM variables are standardized in performing the K-Means clustering analysis.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **First 2 Year - Spend** | **First 2 year – Freq** | **First 2 Year – Recency** |
| **Min** | ***1*** | **2** | **1** |
| **Max** | ***34966*** | **50** | **24** |
| **Average** | ***755.0*** | **1.8** | **20.6** |

**Table 3. Data exploration of RFM variables**

### 4.2.3 Interpretation of the K-Means clustering results

### The output of the K-Means clustering analysis is summarized in the below report, together with a tabulation of 4 clusters based on their RFM values. In this section, the cluster with high RFM values will be selected for next stage classification modelling. This is a very important step of the data analysis design as the selection will affect the subsequent prediction outcomes.

Table

Description automatically generated A screenshot of a cell phone

Description automatically generated**Figure 6. Summary report of K-Means clustering results Table 4. Average of RFM variables of 4 clusters**

From Figure 6 and Table 4, Cluster 4 represents a very exclusive group of high quality customers. They have the highest average spend in the first two years, on average they spend 4 times more than cluster 1 and cluster 3 customers do respectively. Cluster 4 customers’ average purchase frequency value at 6.74 times, 6.48 times and 6.50 times in China, Hong Kong and Taiwan respectively, compared to 2.5 to 3.4 times of other clusters. This is no surprise that the average numbers of month since last purchase is lowest for cluster 4 with an average cluster 4 customer last purchased 6-7 months ago. Accordingly, data samples of cluster 4 are selected for the next stage of classification model building.

* 1. Building classification models

The next stage of data analysis is to develop a classification model that could identify the customers from Cluster 4 who would be top 10% of customers in the “future” ie. 3rd / 4th year. Classification method is supervised learning method used to model the fit of Cluster 4 data samples to the target parameter “Second 2 year - Top 10% spend”. Based on some previous studies, product preference and brand satisfaction are perceived to have an influence to the lifetime value, so variables “Top spending category”, “First purchase category” and “Net Promoter Score” are modelled to fit the target parameter.

#### **4**.3.1 Split data samples into training and test datasets

### In our project, 568 data samples from cluster 4 are used to build the classification models, with a split of 70% / 30% training and testing data. According to Dobbin and Simon (2011), the rule of setting the split between training and testing data set is to minimize mean squared error with respect to full dataset, and their study has suggested 40-80% allocation of training data.

#### **4.3.2 Model evaluation techniques**

#### **Confusion matrix: Since the research objective of this project is to identify** customers who are top 10% spender in next 2 year, the model aims to maximize the number of positive observations and the accuracy rate of a positive class prediction i.e. “Precision” and “recall” are used.

**Lift Chart:** The lift chart can be used to compare visually the improvement different prediction models deliver against a random guess (without a model) on different proportion of data population (Jaffery and Liu, 2009).

**ROC curve:** ROC curve is regarded as one of the most important evaluation method for a binary classification model (Vuk & Curk, 2006; Ferri, Flach & Hernandez-Orallo, 2002). ROC curve is drawn by plotting the true positive rate (precision) against the false positive rate at various threshold settings. The ROC curve represents a trade off between true positive rate and false positive rate in the learning of a model. The area under the ROC curve (AUR) is used to measure of quality of a classification model where a random model has a AUR of 0.5 and a perfect model has a AUR of 1.

1. RESULT DISCUSSIONS

This chapter covers the best model selection and the assessment of model feature importance to overall loyalty prediction.

**5.1 Model performance review**

Table 5 shows the distribution of the target variable “Second 2 Year – Top 10% spend” in each cluster, where we can use it as a baseline reference for model performance assessment.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Total** | | **Top 10% Spend – Yes** | | **Top 10% Spend – No** | |
|  | **N** | ***%*** | **N** | ***%*** | **N** | ***%*** |
| **Cluster 1** | **2530** | ***100%*** | **95** | ***4%*** | **2435** | ***96%*** |
| **Cluster 2** | **1709** | ***100%*** | **230** | ***13%*** | **1479** | ***87%*** |
| **Cluster 3** | **671** | ***100%*** | **84** | ***13%*** | **587** | ***87%*** |
| **Cluster 4** | **568** | ***100%*** | **162** | ***29%*** | **406** | ***71%*** |
| **All** | **5478** | ***100%*** | **570** | ***10%*** | **4907** | ***90%*** |

**Table 5. Distribution of Y/N attributes of “Second 2 Year – Top 10% Spend” by cluster**

From table 5, 29% of the customers from cluster 4 are the top 10% spender in the second 2 years’ period, this suggests that clustering technique has produced 3 times better prediction than a “random pick” to identify top 10% spender. This 29% ratio is then used as an important referencing baseline for the model performance review in second stage.

**Confusion matrix results:** After the 4 classification models have been trained, they are then tested with the test dataset on Alteryx Designer to obtain unbiased results for model evaluation (Figure 12). Next, the confusion matrix results of each model are then used to calculate the overall model accuracy, precision and recall (table 6) to evaluate the model performance in this project.

A screenshot of a social media post

Description automatically generated

**Figure 9. Confusion matrices of 4 classification models**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Accuracy** | **Rank** | **Precision** | **Rank** | **Recall** | **Rank** |
| **Boosted Model** | **0.7570** | **3rd** | **0.6203** | **3rd** | **0.556818** | **1st** |
| **Decision Tree** | **0.7746** | **1st** | **0.6622** | **1st** | **0.556818** | **1st** |
| **Logistic Regression** | **0.7324** | **4th** | **0.6034** | **4th** | **0.5** | **3rd** |
| **Random Forest** | **0.7641** | **2nd** | **0.6567** | **2nd** | **0.397727** | **4th** |

**Table 6. Three measures calculated from confusion matrices**

Based on the confusion matrix results, the decision tree model produces the best prediction result with consistently the highest score in accuracy, precision and recall. It has an accuracy of 77.5% overall, a precision of 66.2% and recall rate of 55.7%. Random forest, a variation of decision tree technique, followed closely in terms of accuracy and precision scores at 76.4% and 65.7% respectively, however the model has much lower recall at 39.8*%* only.

**Lift curve:** Another evaluation technique is lift curve (Figure 10) which compares the lift score of the models at different proportion of data observations. The graph suggests that beyond 40% proportion of data observations, decision tree model starts to produce the most superior lift score than other models do.

**ROC curve:** ROC curves are plotted in figure 11 to select the model with highest AUC (area under the curve) where it minimizes the false positive rate at any given true positive rate. Decision tree outperforms other 3 models at true positive rate between 70% - 83% with lowest false positive rate at the same time. When a higher true positive rate is required over 83%, the boosted model then generally performs the best among 4 models.

A close up of a map

Description automatically generatedA close up of a map

Description automatically generated

**Figure 10. Lift Chart Figure 11. ROC curve**

**5.2 Illustration of the decision tree**

Based on the 3 evaluation techniques, decision tree model is selected as the best performing classification model in the measures of overall accuracy, precision, recall and lift curve, and also being one of the better performing model in AUC. In order to interpret the influence of the product preference and satisfaction towards customer loyalty, the decision tree model was illustrated in a tree as above (Figure 13). In a decision tree, there are root, branches and leave nodes. The 3 variables “First category purchase” (ACC, BAGS, MENS, WOMENS), “Top category purchase” (ACC, BAGS, MENS, WOMENS) and “Net Promoter Score” (10-point-scale rating) are the branches of the tree that extend from the root to classify data observations into each of the final leave nodes. The key decision rules of the tree are as following:

**Branch “NPS<8.5” (NO):** The tree model suggests customer satisfaction towards brand is most important to the loyalty. NPS which indicates the likelihood of brand recommendation is the first branch from the root. For customers with NPS is 8.5 or above, 61% of them are “top 10% spender”, while if NPS is below 8.5, 91% are not “top 10% spenders” eventually.

A screenshot of a cell phone

Description automatically generated

**Figure 13. Illustration of decision tree model**

**Branch “First Spend category = ACC,WOMENS” (NO):** Asthe tree extended further, “First spend category” is then used to classify customers. The branch “First spend category = ACC, WOMENS” suggests that for customers who with NPS >8.5 and purchased MENSWEARS/BAGS (Not purchased ACC, WOMENS) at the first purchase, 71% of them would eventually in top 10% spenders.

**Branch “Top spend category = BAGS, MENS, WOMENS” (NO):** “Top spend category” is used only towards very end of the tree to classify a very exclusive group of customers who have NPS score >9.5 and purchased accessories at the first purchase into “Top 10% spenders”

CONCLUSION AND FUTURE WORKS

* 1. Conclusion

This project has been demonstrated how both unsupervised technique (K-means clustering) and supervised technique (classification models) could be constructed in a two-stage model to enhance the prediction results. The values of customer-centric business intelligence have been proven to be instrumental to a fashion brand by applying RFM, customer product preference and customer satisfaction data. Besides the prediction model has the ability of identifying potential loyal customers to optimize marketing and CRM investment, the model features of the two-stage prediction model have also provided important insights of drivers towards loyalty.

* 1. Suggestions for future work

This research project have successfully developed a two-stage loyalty prediction model with accuracy rate over 77%, yet a few possible future works could further be accomplished. For example, test the prediction model in other markets given the accessibility to other market data. A different prediction models could be applied to analyse customers with 1-time purchase in first period and from cluster 2 & 3. The current methodology is designed to model only customers who have purchased 2 times or above so that the RFM model can model the “variations in frequency and recency value” of data samples. Last but not least, deep learning techniques like ANN can be deployed to enhance the prediction ability. Some research studies have used deep learning technique like neural networks to analyse RFM in various application including lifetime value model (Tkachenko, 2015) and credit scoring model in banking (Alborzi & Khanbabaei, 2016). These suggestions create an interesting research space for future.

1. REFERENCES
2. Pitta, D., Franzak, F. and Fowler, D., 2006. A strategic approach to building online customer loyalty: integrating customer profitability tiers. *Journal of consumer marketing.*
3. Liu, C. and Wang, T.Y., 2017. A study on the effect of service quality on customer loyalty and corporate performance in financial industry. *Problems and perspectives in management, (15, Iss. 2 (cont. 2)),* pp.355-363.
4. Ramanathan, U., Subramanian, N., Yu, W. and Vijaygopal, R., 2017. Impact of customer loyalty and service operations on customer behaviour and firm performance: empirical evidence from UK retail sector. *Production Planning & Control, 28(6-8),* pp.478-488.
5. Pi, W.P. and Huang, H.H., 2011. Effects of promotion on relationship quality and customer loyalty in the airline industry: The relationship marketing approach. *African Journal of Business Management, 5(11),* pp.4403-4414.
6. Bose, S. and Rao, V.G., 2011. PERCEIVED BENEFITS OF CUSTOMER LOYALTY PROGRAMS: VALIDATING THE SCALE IN THE INDIAN CONTEXT. *Management & Marketing, 6(4).*
7. Antonios, J., 2011. *Understanding the effects of customer education on customer loyalty. Business Leadership Review, 8(1),* pp.1-15.
8. Shih-I, Cheng.,2011.Comparisons of Competing Models between Attitudinal Loyalty and Behavioral Loyalty. *International Journal of Business and Social Science Vol. 2, No.10, June 2011*
9. Trinh, G.T., Anesbury, Z.W. and Driesener, C., 2017. Has behavioural loyalty to online supermarkets declined?. *Australasian Marketing Journal (AMJ), 25(4),* pp.326-333.
10. Liu, M.T., Liu, Y., Mo, Z., Zhao, Z. and Zhu, Z., 2019. How CSR influences customer behavioural loyalty in the Chinese hotel industry. *Asia Pacific Journal of Marketing and Logistics.*
11. Gupta, S., Lehmann, D.R. and Stuart, J.A., 2004. Valuing customers. *Journal of marketing research, 41(1),* pp.7-18.
12. Rust, R.T., Lemon, K.N. and Zeithaml, V.A., 2004. Return on marketing: Using customer equity to focus marketing strategy. *Journal of marketing, 68(1),* pp.109-127.
13. Zhang, J.Q., Dixit, A. and Friedmann, R., 2010. Customer loyalty and lifetime value: an empirical investigation of consumer packaged goods. *Journal of marketing theory and practice, 18(2),* pp.127-140.
14. Chen, D., Sain, S.L. and Guo, K., 2012. Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining. *Journal of Database Marketing & Customer Strategy Management, 19(3)*
15. Gupta, S., Hanssens, D., Hardie, B., Kahn, W., Kumar, V., Lin, N., Ravishanker, N. and Sriram, S., 2006. *Modeling customer lifetime value. Journal of service research, 9(2),* pp.139-155.
16. Song, M., Zhao, X., Haihong, E. and Ou, Z., 2017. Statistics-based CRM approach via time series segmenting RFM on large scale data. *Knowledge-Based Systems, 132,* pp.21-29
17. Cheng, C.H. and Chen, Y.S., 2009. Classifying the segmentation of customer value via RFM model and RS theory. *Expert systems with applications, 36(3),* pp.4176-4184.
18. Yeh, I.C., Yang, K.J. and Ting, T.M., 2009. Knowledge discovery on RFM model using Bernoulli sequence. *Expert Systems with Applications, 36(3),* pp.5866-5871.
19. Coussement, K. and De Bock, K.W., 2013. Customer churn prediction in the online gambling industry: The beneficial effect of ensemble learning. *Journal of Business Research, 66(9),* pp.1629-1636.
20. Hallowell, R., 1996. The relationships of customer satisfaction, customer loyalty, and profitability: an empirical study. *International journal of service industry management.*
21. Oliver, R.L., 2006. Customer satisfaction research. *The handbook of marketing research: Uses, misuses, and future advances, 1.*
22. Chandrashekaran, M., Rotte, K., Tax, S.S. and Grewal, R., 2007. Satisfaction strength and customer loyalty. *Journal of marketing research, 44(1),* pp.153-163.
23. Brito, P.Q., Soares, C., Almeida, S., Monte, A. and Byvoet, M., 2015. Customer segmentation in a large database of an online customized fashion business. *Robotics and Computer-Integrated Manufacturing, 36,* pp.93-100.
24. Heilman, C.M. and Bowman, D., 2002. Segmenting consumers using multiple-category purchase data. *International Journal of Research in Marketing, 19(3),* pp.225-252.
25. Zalaghi, Z. and Varzi, Y., 2014. Measuring customer loyalty using an extended RFM and clustering technique. *Management Science Letters, 4(5),* pp.905-912.
26. Alpaydin, E., 2020. *Introduction to machine learning. MIT press.*
27. Wu, X., Kumar, V., Quinlan, J.R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A., Liu, B., Philip, S.Y. and Zhou, Z.H., 2008. Top 10 algorithms in data mining. *Knowledge and information systems, 14(1),* pp.1-37.
28. Scornet, E., 2017. Tuning parameters in random forests. *ESAIM: Proceedings and Surveys, 60,* pp.144-162.
29. Livingston, F., 2005. Implementation of Breiman’s random forest machine learning algorithm. *ECE591Q Machine Learning Journal Paper,* pp.1-13
30. Schapire, R.E., 2003. The boosting approach to machine learning: An overview. *In Nonlinear estimation and classification* (pp. 149-171). *Springer, New York, NY.*
31. Machado, M.R., Karray, S. and de Sousa, I.T., 2019, August. LightGBM: An effective decision tree gradient boosting method to predict customer loyalty in the finance industry. *In 2019 14th International Conference on Computer Science & Education (ICCSE)* (pp. 1111-1116). IEEE.
32. Aleksandrova, Y., 2018. Application of Machine Learning for Churn Prediction Based on Transactional Data (RFM Analysis). *In 18 International Multidisciplinary Scientific Geoconference SGEM 2018: Conference Proceedings (Vol. 18, No. 2.1,* pp. 125-132).
33. Doğan, O., Ayçin, E. and Bulut, Z.A., 2018. Customer segmentation by using RFM model and clustering methods: a case study in retail industry. *International Journal of Contemporary Economics and Administrative Sciences, 8(1),* pp.1-19
34. Sheshasaayee, A. and Logeshwari, L., 2018, May. Implementation of Clustering Technique Based RFM Analysis for Customer Behaviour in Online Transactions. *In 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI)* (pp. 1166-1170). IEEE.
35. Le, T., Lee, M.Y., Park, J.R. and Baik, S.W., 2018. Oversampling techniques for bankruptcy prediction: novel features from a transaction dataset. *Symmetry, 10(4),* p.79.
36. Hu, W.H., Tang, S.H., Chen, Y.C., Yu, C.H. and Hsu, W.C., 2018. Promotion recommendation method and system based on random forest. *In Proceedings of the 5th Multidisciplinary International Social Networks Conference*(pp. 1-5).
37. Maryani, I. and Riana, D., 2017, August. Clustering and profiling of customers using RFM for customer relationship management recommendations. *In 2017 5th International Conference on Cyber and IT Service Management (CITSM)*(pp. 1-6). IEEE.
38. Zabkowski, T.S., 2016. RFM approach for telecom insolvency modeling. *Kybernetes.*
39. Daoud, R.A., Amine, A., Bouikhalene, B. and Lbibb, R., 2015. Combining RFM model and clustering techniques for customer value analysis of a company selling online. *In 2015 IEEE/ACS 12th International Conference of Computer Systems and Applications (AICCSA)* (pp. 1-6). IEEE.
40. You, Z., Si, Y.W., Zhang, D., Zeng, X., Leung, S.C. and Li, T., 2015. A decision-making framework for precision marketing. *Expert Systems with Applications, 42(7),* pp.3357-3367.
41. Kaur, K. and Vashisht, S., 2015. A novel approach for providing the customer churn prediction model using enhanced boosted trees technique in cloud computing*. International Journal of Computer Applications,*114(7).
42. Chiang, W.Y., 2014. Applying data mining with a new model on customer relationship management systems: a case of airline industry in Taiwan. *Transportation Letters, 6(2),* pp.89-97.
43. Coussement, K., Van den Bossche, F.A. and De Bock, K.W., 2014. Data accuracy's impact on segmentation performance: Benchmarking RFM analysis, logistic regression, and decision trees. *Journal of Business Research, 67(1),* pp.2751-2758.
44. Han, S.H., Lu, S.X. and Leung, S.C., 2012. Segmentation of telecom customers based on customer value by decision tree model. *Expert Systems with Applications, 39(4),* pp.3964-3973.
45. Hwang, H., Jung, T. and Suh, E., 2004. An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry*. Expert systems with applications, 26(2),* pp.181-188.
46. Jaffery, T. and Liu, S.X., 2009. Measuring campaign performance by using cumulative gain and lift chart. *In SAS Global Forum* (p. 196).
47. Vuk, M. and Curk, T., 2006. ROC curve, lift chart and calibration plot. *Metodoloski zvezki, 3(1),* p.89.
48. Ferri, C., Flach, P. and Hernández-Orallo, J., 2002. Learning decision trees using the area under the ROC curve. *In Icml (Vol. 2,* pp. 139-146).
49. Tkachenko, Y., 2015. Autonomous CRM control via CLV approximation with deep reinforcement learning in discrete and continuous action space.*arXiv preprint arXiv:1504.01840.*
50. Alborzi, M. and Khanbabaei, M., "Using data mining and neural networks techniques to propose a new hybrid customer behaviour analysis and credit scoring model in banking services based on a developed RFM analysis method." *International Journal of Business Information Systems 23, no. 1 (2016): 1-22.*