International Journal of Advanced Trends in Computer Applications
www.ijatca.com

BUSINESS SALES ANALYSIS AND CUSTOMER RETENTION

[1]Ankit Chaplot, [2]Ayush Tiwari, [3]Gaurav Bisaria, [4]Mustafa Raj
[1,2,3,4] Savitribai Phule Pune University
[1]ankitchaplot1@gmail.com, [2]ayushtiwari811@gmail.com, [3]gbisaria7@gmail.com,
[4]mustafaraj9715@gmail.com

Abstract: *This project proposes a method that predicts customer value by focusing on purchasing behaviour. The method generates a relevance model for purchase days and amount in each period between customer value and purchasing histories beforehand based on a consumer panel survey. We have adopted the random forest method to generate the prediction model. The proposed method facilitates the provisioning of smart customer management to each customer according to level such as suggesting products or services. The problem faced by the company is how to determine potential customers and apply CRM (Customer Relationship Management) in order to perform the right marketing strategy, so it can bring benefits to the company. This research aims to perform clustering and profiling customer by using the model of Recency Frequency and Monetary (RFM) to provide customer relationship management (CRM). The method used in this study consists of four steps: data mining from transaction history data of customer sales, data mining modeling using RFM and customer classification with decision tree, determination of customer loyalty level and recommendation of customer relationship management (CRM).*

Keywords: Transaction Data Mining; RFM; Clustering; Kmeans; Decision Tree; Profiling Customers; CRM.

I. Introduction

Most companies use a customer relationship management system to improve their relation attributes such as demographic data, preference, purchase history, usage history and contact history. By using and analyzing these data, it is possible to respond to customer.
Finding excellent customers, which means those who visit and purchase frequently is important for improving sales and the efficiency of various measures taken to target customers. The project defines the metric of customer level in this paper. Specifically, customer level can be used for customer selection such as optimizing the approach to customers, campaigns to train best customer candidates and measures to activate dormant customers. By holding and using the customer level in a customer relationship management system, we can build a more productive relationship with customers.

RFM is existing approach to calculate the customer level. RFM is one of several methods that can extract customer level from purchase history data. From the purchase history, this method extracts as indicators Recency (last purchase date), Frequency (purchase frequency), Monetary (purchase price). They can be used to indicate customer's level.

However, there are many customers whose purchase history data is too scant to allow customer level determination. If we can predict future investment, human resources and expertise, to the company's value chain. Enhanced performance can result from improved communication and coordination with this set of suppliers. With fewer vendors, increased with customers. Customer level from a small amount of purchase history, we can better support existing customers and acquire new customers as excellent customers. We propose that predicting customer level with the least possible delay is also important for customer relationship management. Nevertheless, no proposal has clarified the effectiveness of the amount and types of purchase data used to predict customer. Such systems record and manage customer level.

The project is organized as follows. Section I discusses related work and existing techniques for customer level determination. Section II introduces the customer relationship management system. Purchase data used to evaluate the proposed method are shown in Section III. In Section IV, it is explained how to calculate the customer level by using RFM. The proposed method to predict the customer level is explained in Section V. Section VI details our evaluation.

II. Procedures AND METHOD

(I) Section I
1. CRM

Customer relationship management's definition is also its ambition: the development and maintenance of mutually beneficial long-term relationships with strategically significant markets.

The Customer

The customer is of key importance because only relationships with customers generate revenues for a company. Establishing a good long-term relationship with customers can take the form of the provision of benefits such as special prices and preferential treatment.

2. RFM

RFM stands for Recency, Frequency, and Monetary value, each corresponding to some key customer trait. These RFM metrics are important indicators of a customer's behaviour because frequency and monetary value affects a customer's lifetime value, and recency affects retention, a measure of engagement.

3. Random Forest

Prediction models are constructed by using the period of the features. It is assumed that the amount of data for each customer is different. It has been decided to use the data for a certain period as learning data.

Each customer is given one value (customer level), so predicting customer level can be seen as a classification problem. To solve this classification problem, we adopt the Random Forest for model construction.

By comparing the prediction power, the constructed model that exceeds the threshold value set by the user of our customer relationship management system is adopted as the final prediction model. As a result, we can confirm how much data is required to realize predictions whose accuracy exceeds the threshold. It is also possible to know how much accuracy and the amount of data that can be predicted even below the threshold.

(II) Section II

CUSTOMER LEVELS

Best Customers – Communications with this group should make them feel valued and appreciated. These customers likely generate a disproportionately high.

The Suppliers

Suppliers provide input, such as raw information system alignment and customer-information sharing becomes possible.

Thus focusing on keeping them happy should be a top priority. Further analyzing their individual preferences and affinities will provide additional opportunities for even more personalized messaging.

- **High-spending New Customers** – It is always a good idea to carefully "incubate" all new customers, but because these new customers spent a lot on their first purchase, it's even more important. Like with the Best Customers group, it's important to make them feel valued and appreciated – and to give them terrific incentives to continue interacting with the brand.
- **Lowest-Spending Active Loyal Customers** – These repeat customers are active and loyal, but they are low spenders. Marketers should create campaigns for this group that make them feel valued, and incentivize them to increase their spend levels. As loyal customers, it often also pays to reward them with special offers if they spread the word about the brand to their friends, e.g., via social networks.
- **Churned Best Customers** –These are valuable customers who stopped transacting a long time ago. While it's often challenging to re-engage churned customers, the high value of these customers makes it worthwhile trying. Like with the Best Customers group, it's important to communicate with them on the basis of their specific preferences, as known from earlier transaction data.

(III) Section III

CRM System:-
The system is divided into three parts:
- Prediction module,
- Information Delivery module,
- User Interface module.

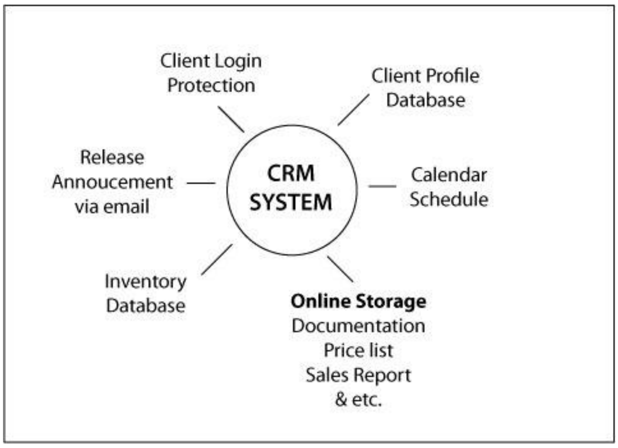

Fig. 1: CRM Architecture

The system obtains purchase data through retail POS systems. The prediction module uses purchase data to predict the customer level and products likely to be purchased. In order to deliver information that matches the condition and situation of the customer, the Information Delivery module selects, for each customer, the information to be delivered. The User Interface module generates data in a format that allows its display via the output user interface.

(IV) Section IV

1. Preparing and Pre-processing-

Data This research performed data collection of sales transaction history dataset on industries of many dataset which then were prepared and processed by determining weighting criteria previously based on variable recency, frequency and monetary.

2. Clustering

After all transaction data was transformed into numbers, then the data was able to be grouped by using K-means Method algorithm.

To be able to group the data into several clusters needed to do some steps as follows:

1. Determine the number of clusters desired. In this study the existing data will be grouped into four clusters.
2. Determine the starting centre point of each cluster. In this study the initial centre point was determined randomly and it obtained a central point of each cluster.

3. Classification with Decision Tree method

The formation of the classification model was done by implementing the data into GUI classification. Data produced a model tree and the rule of determining the characteristics of customers. The result of classification model can be seen as a tree and this classification model involved 4 attributes namely Recency, Frequency, Monetary and Cluster.

4. Profiling Customer

Best Customers – Communications with this group should make them feel valued and appreciated. These customers likely generate a disproportionately high percentage of overall revenues and thus focusing on keeping them happy should be a top priority. Further analyzing their individual preferences and affinities will provide additional opportunities for even more personalized messaging.

High-spending New Customers – It is always a good idea to carefully “incubate” all new customers, but because these new customers spent a lot on their first purchase, it’s even more important. Like with the Best Customers group, it’s important to make them feel valued and appreciated – and to give them terrific incentives to continue interacting with the brand.

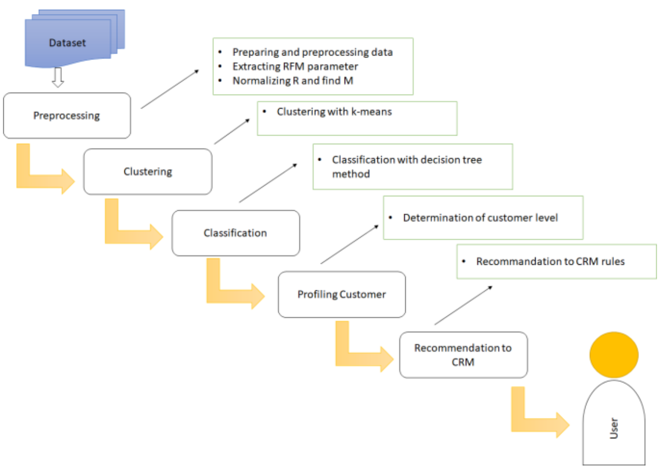

Fig. 2: Process flow diagram

Lowest-Spending Active Loyal Customers – These repeat customers are active and loyal, but they are low spenders. Marketers should create campaigns for this group that make them feel valued, and incentivize them to increase their spend levels. As loyal customers, it often also pays to reward them with special offers if they spread the word about the brand to their friends, e.g., via social networks.

Churned Best Customers – These are valuable customers who stopped transacting a long time ago. While it’s often challenging to re-engage churned customers, the high value of these customers makes it worthwhile trying. Like with the Best Customers group, it’s important to communicate with them on the basis of their specific preferences, as known from earlier transaction data.

5. Recommendation to CRM

Rule 1: Synchronization matters
Rule 2: Mobile technology is number one
Rule 3: Use predictive analytics
Rule 4: Your Company needs logical, uniform branding
Rule 5: Different messages across multiple media forms
Rule 6: The total shift to online sales
Rule 7: The personalized customer service movement
Rule 8: Customers are already making demands
Rule 9: Competition is a direct threat
Rule 10: Use SMS For customer management

(V) Section V

As the evaluation metrics, the authors adopt F-measure and Accuracy. The F-measure is a harmonic mean of precision and recall. Let i be the customer level. Let $U(i)$ be all levels (customer level) that are predicted correctly, $V(i)$ be all level predictions, and $W(i)$ be the total number of customers who have the target customer level.

Precision P(i) and Recall R(i) are computed as follows ,

P(i) = U(i) / V(i)
R(i) = U(i) / W(i)
F-measure F(i) is given by

F(i) =2 * P(i)*R(i)/(P(i)+R(i))

If O is the number of all customers in the test data , it is computed as

O = ∑I W(i)
N is the total number of customers whose values were predicated correctly. N is computed as

N = ∑I U(i)

Accuracy A is computed as

A = N/O

(VI) Section VI

To clarify which features are effective for predicting the customer level we direct your attention to shows the average prediction accuracy of all customer levels for each feature. Variations were observed in the prediction accuracy when using data as learning data for a month.

The prediction accuracy was highest when all features were used. This result shows that using all the features is effective for customer level prediction.

Second, we discuss the prediction accuracy of the customer level for each period the f-measure of each customer level and each period achieved by using all features. Among the customer levels, prediction of "Best Customer" was the most stable.

We confirmed that it can be predicted with 69% (F-measure) with the data of one month. Compared to other customer levels, "Best Customer" is the customer level for which features are easiest to extract. For data from 1M to 3M, the F-measure of "Non-Active" was the lowest. Since it was confirmed that the prediction accuracy depends on customer level, it is conceivable to decide the amount of data to be used according to the customer level being targeted. In order to manage customers, while it is important to maintain "Best Customer", it is also important to mole "Good Customer" into "Best Customer". This is because "Good Customer" can be thought of as a "Best Customer" candidate. By using this method, it is possible to predict customer quality at an early stage and use it for customer management.

III. Proposed System

The research design consists of several stages including collecting data (dataset), pre-processing data, clustering with k-means method), validity test of cluster and classification followed by customer profiling and recommendation for RFM.

The proposed research proposals have particularity mainly on the use of customer classification approach. The proposed method can be seen completely in below diagram.

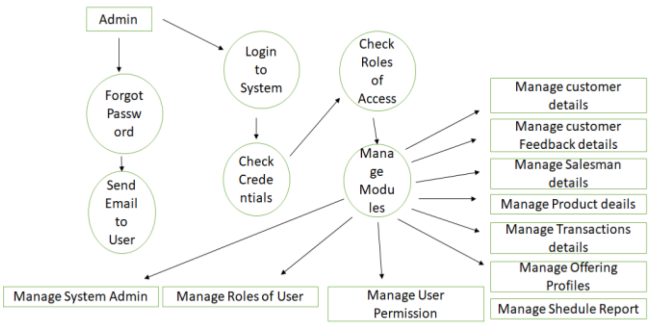

Fig. 3: System Architecture

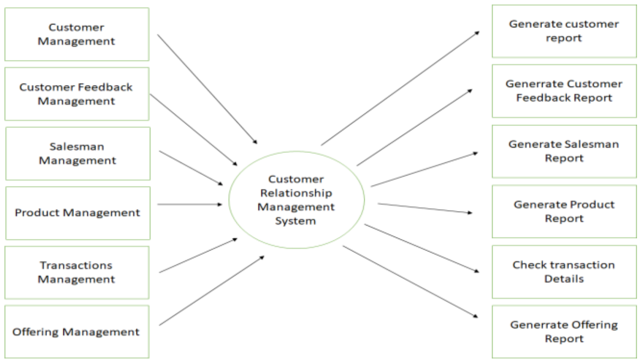

Fig. 4: Managements and reports

IV. Conclusion and Future Work

Conclusion

In this project, a customer relationship management system that uses customer level. The proposed method is able to predict customer level by focusing on customer purchase behavior.

In this it is clarified the effectiveness of the types and amounts of data for predicting customer level. Evaluations of real datasets showed the effectiveness of the proposed method. The results showed that using all features was most effective for customer level prediction. Among individual features, “Total amount for each product and each period” proved to be the most effective. The second most effective individual feature was “Total amount for each period”.

In order to manage customers, we must maintain the “Best Customer”, but also convert “Good Customer” into “Best Customer”. By using this method, it is possible to predict customer quality at an early stage and use it for customer management. The method is also useful for customer selection in campaigns and measures to activate dormant customers.

In this study, the period of the data used for learning was increased units of a month. It is also possible to employ different periods such as weekly or daily depending on the situation.

FUTURE WORK

As future works, the authors plan to extract effective Product IDs to improve the prediction accuracy of customer level. Integrating the proposed model into the author’s customer relationship management system is the next step towards demonstration experiments in the real world. The authors will clarify the ability of extracting and utilizing each predicted customer level.

V. Equations

As the evaluation metrics, the authors adopt F-measure and Accuracy. The F-measure is a harmonic mean of precision and recall. Let i be the customer level. Let $U(i)$ be all levels (customer level) that are predicted correctly, $V(i)$ be all level predictions, and $W(i)$ be the total number of customers who have the target customer level.

Precision P(i) and Recall R(i) are computed as follows ,

$$P(i) = U(i) / V(i)$$
$$R(i) = U(i) / W(i)$$

F-measure F(i) is given by

$$F(i) = 2 * P(i)*R(i)/(P(i)+R(i))$$

If O is the number of all customers in the test data , it is computed as

$$O = \sum I\ W(i)$$

N is the total number of customers whose values were predicated correctly. N is computed as

$$N = \sum I\ U(i)$$

Accuracy A is computed as

$$A = N/O$$

References

[1] I. Maryani, D. Riana, “Clustering and profiling of customers using RFM for customer relationship management recommendations,” in Proceedings of International Conference on Cyber and IT Service Management (CITSM), pp.1–6, 2017.

[2] M. Tsoy, V. Shchekoldin, “RFM-analysis as a tool for segmentation of high-tech product’s consumers,” in Proceedings of International Scientific-Technical Conference on Actual Problems of Electronics Instrument Engineering(APEIE), pp. 290–293, 2017.

[3] J. Wei, S. Lin, Y. Yang, H. Wu, “Applying data mining and RFM model to analyze customer’s values of a veterinary hospital,” in Proceedings of International Symposium of Computer, Consumer and Control (IS3C), pp.481–484, 2016.

[4] R. Daoud, A. Amine, B. Bouikhalene, “Combining RFM model and clustering techniques for customer value analysis of a company selling online,” in Proceedings of International Conference of Computer Systems and Applications (AICCSA), pp.1–6, 2015.

[5] D. Kim, J. Lee, S. Ahn, Yeongho, M, O. Kwon, “RFM analysis for detecting future core technology,” in Proceedings of the 2012 ACM Research in Applied Computation Symposium, pp.55-59, 2012.

[6] W. Zhang, L. Zhu, "Computer Simulation of Electronic Commerce Customer Churn Prediction Model Based on Web Data Mining," in Proceedings of International Conference on Smart Grid and Electrical Automation, pp.661-663, 2017.

[7] M. Tsoy, V. Shchekoldin, "RFM-analysis as a tool for segmentation of high-tech product's consumers," in Proceedings of International Scientific-Technical Conference on Actual Problems of Electronics Instrument Engineering(APEIE), pp. 290–293, 2017.

Author Profile

Ankit Chaplot received the B.E. degree in Computer Engineering from Dr. D. Y. Patil Institute Of Technology, SPPU, Pune in 2015 and 2019, respectively. During 2015-2019, he worked under the computer department of his college to study customer relationship management and data science. He is now with Larson & Toubro Infotech.

Ayush Tiwari received the B.E. degree in Computer Engineering from Dr. D. Y. Patil Institute Of Technology, SPPU, Pune in 2015 and 2019, respectively. During 2015-2019, he worked under the computer department of his college to study customer relationship management and data science.

Gaurav Bisaria received the B.E. degree in Computer Engineering from Dr. D. Y. Patil Institute Of Technology, SPPU, Pune in 2015 and 2019, respectively. During 2015-2019, he worked under the computer department of his college to study customer relationship management and data science. He is now with Tata Consultancy Services.

Mustafa Raj received the B.E. degree in Computer Engineering from Dr. D. Y. Patil Institute Of Technology, SPPU, Pune in 2015 and 2019, respectively. During 2015-2019, he worked under the computer department of his college to study customer relationship management and data science. He is now with Tata Elxsi.