Applying Data Mining Techniques and Extended RFM Model in Customer Loyalty Measurement

Panwad Bunnak, Sotarat Thammaboosadee, and Supaporn Kiattisin
Technology of Information System Management Division, Faculty of Engineering, Mahidol University, Nakhon Pathom, Thailand
Email: ohana_mylife@hotmail.com, {sotarat.tha, supaporn.kit }@mahidol.ac.th

Abstract—**This paper proposes a loyalty measurement model of individual customer for the benefit in creating of marketing campaign and activities as well as the suitable products and services for customers and establishment of good customer relationship. This study adapts the concept of RFM (Recency- Frequency-Monetary) model and applies to database of customer purchases and the customer type. The business type of selected organization is commercial business. To apply the RFM concept to find customer loyalty according to type of customer, the customer loyalty is partitioned into 5 classes using k-means clustering algorithm and is heuristically assigned customer types: Platinum, Gold, and Silver. Type of customers is then brought into consideration the extending of the RFM Model with customer analytics to make it even better customer classification performance. Finally, the classification system generates decision rules to find out the loyalty of new future customers using C4.5 decision tree algorithm.**

Index Terms—**customer loyalty, RFM model, k-means, decision tree, CRM**

I. INTRODUCTION

In business today, it is very important to be able to satisfy customer's needs and wants because the current customer determines the direction by choosing, in a self-guided manner. If an organization is unable to satisfy the requirements of the customers, the customers will switch product or service provider immediately which makes losing of the opportunity and competition. It also impacts organization income because in the current business climate, many companies compete for same customers who have more right to choose than the organization. In order to the company is able to keep an existing customer base, they need to understand the behaviour of customers. The company must define a clear customer segmentation to establish relationships with customers. In addition, customer segmentation can be used to determine appropriate marketing strategy and apply these to the target groups. The goal is to keep an existing customer because keep an existing customer; it cost less than finding new customers. In addition, using of existing database is benefit

the company in the way of organization can determine the customer's loyalty.

In this paper, we have extended the concept of RFM model [1] applied to database of customer purchases and the type of customer of commercial business organization. The methodology is applying RFM Model to find customer loyalty according type of customer. Then the data will be partitioned for customer loyalty by clustering method and be assigned for customer types with 3 levels: Platinum, Gold and Silver. Consequently, the customer type will be brought into co-consideration with the RFM principles to make it even practical customers and their loyalty segmentation. Finally, the decision rules are generated to find out the loyalty of new customers in the future using decision tree algorithm.

The rest of this paper is organized as follows. In section 2, we describe an overview of the related theories and researches. Section 3 presents the proposed procedure and steps of the research process. Section 4 describes the analytically experimental results and concluding the paper in Section 5.

II. RELATED WORKS

A. The CRISP-DM Process

The CRISP-DM [2] consists on a cycle that comprises six stages as follow:

Step1: (Business Understanding) Understanding and identifying business problem or business opportunity.

Step2: (Data Understanding) Collecting relevant data into analyze the data mining techniques.

Step3: (Data Preparation) usually takes a longer time. The model derived from the data mining results accurate depends on the quality of data.

Step4: (Modeling) process modeling with data mining techniques.

Step5: (Evaluation) evaluates or measure the performance of the models analyzed in the previous step.

Step6: (Deployment) conduct the result or knowledge obtained from the analysis of data with data mining techniques.

B. RFM Model Definition

RFM [1] is a method used for analyzing customer value. It is commonly used in database marketing and

direct marketing and has received particular attention in retail and professional services industries. The definition of the RFM model is described as follows:

Recency (R) is the interval between the time, which the latest consuming behavior happens, and present.

Frequency (F) is the number of transactions in a particular time period.

Monetary (M) refers amount of money in a particular time period.

According to the research proposed by Birant [3], this research provides an overview RFM that can be applied to field of Data Mining. It adopted the clustering task using K-Means algorithm and classification rules using C4.5 Decision Tree algorithm [4] and classifying the customer segment using RFM.

C. K-Means Algorithm

K-means clustering [5] is one of the simplest unsupervised learning algorithms, descriptive modeling. The algorithm is composed of the following steps:

Step1: Place K points into space and points represent initial group of centroids.

Step2: Assign each object to the group that has the closest centroid.

Step3: When all objects have been assigned to the group, recalculate the positions of the K centroids.

Step4: Repeat second step and third step until the centroids no longer move.

D. C4.5 Decision Tree Algorithm

C4.5 is an algorithm used to generate a decision tree developed by Ross Quinlan [6]. A decision tree is a tree like flow chart structure. Decision tree models include such concepts as root node, internal node, leaf node and branch. The root node represents top-most node begins with creating the tree with all training data and extracts a node until leaf node. Then combines and removes branches that affect reduce the accuracy.

E. Customer Relationship Management (CRM)

The CRM [7] provides a sense of customer product affinity for a service or organization. When customers bond better with a company, they are more likely to remain loyal patron. The importance of customer base increases as the number of competitors in each business substantially increases. Competition increases as the number of customers remains the same. Businesses seeking ways to satisfy customers can lead to customer loyalty and stability. A company's ability to satisfy customers' needs and wants is very important, to understand customer behavior and analysis with our marketing strategy.

III. METHODOLOGY

In this paper, we apply data mining techniques to database of customer purchases for Customer Loyalty. The steps of the research process as shown in Fig. 1.
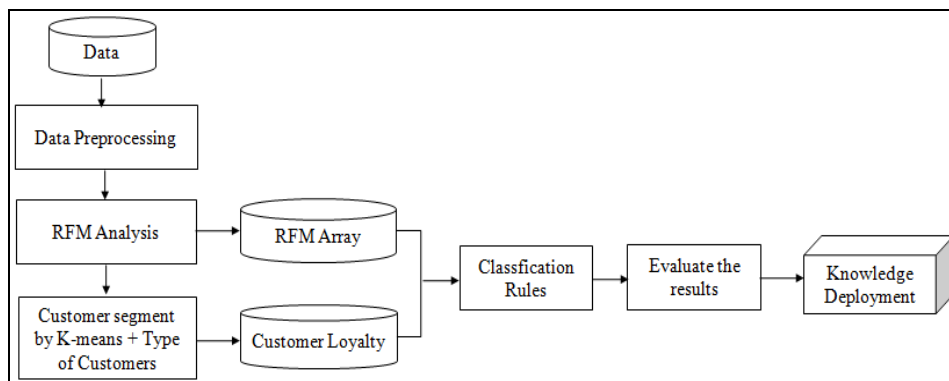

Figure 1. Overall methodology.

A. Data & Data Preprocessing

This step selects related dataset to be used in case study of data mining and then pre-processes data which is an important step. Data preprocessing eliminates irrelevant data by some methods such as data integration, data transformation, and data reduction. It helps in optimizing the effects in the precision and accuracy in the later steps.

B. RFM Analysis

In this step, RFM analysis is applied by defining the scaling of R, F, and M attributes. The procedures are as follow:

Step 1: Sorting the data of three R, F, M attributes by descending or ascending order.

Step 2: Partitioning those three attributes by K-means algorithm with five clusters criteria. These five parts are assigned for the score as 5, 4, 3, 2 and 1 which represent to the customer contributions for organization. 5 represent to most customer loyalty and 1 represent to least customer loyalty.

Step 3: Repeat First step and Second step for each R, F, M attribute individually. This, there are total RFM Array 125 (5x5x5) combinations since each attribute in R, F, M attributes has 5 scales (5, 4, 3, 2 and 1).

C. Customer Segment by K-means + Type of Customers

In this process, the RFM Array is divided into 3 groups according to the customer type before using the K-means algorithm for clustering customer segment. As summarized in Table I, the customer type can be classified as: Silver (S), Gold (G), and Platinum (P).

TABLE I.   THREE CUSTOMER TYPES GROUPS OF RFM ARRAY

Type of Customer		
Silver	Gold	Platinum
RFM Array 125 COMBINATIONS	RFM Array 125 COMBINATIONS	RFM Array 125 COMBINATIONS

The procedures in this phase are that: firstly, customer loyalty is partitioned according type of customer 5 clusters using the K-means algorithm and then determined segment customer loyalty to each class, which are "Very High" (VH), "High" (H), "Medium" (M), "Low" (L) and "Very Low" (VL). Then, the system will repeat the first step for each customer loyalty according type of customer. Finally, there will be customer loyalty divided into 3 groups according to the type of customers.

D.   Classification Rules

The rules of classification are discovered from RFM, customer type, and using the clustering results obtained in the previous step. And create prediction rules of loyalty of future customers discussed in the next section.

E.   Evaluate the Results

In our model, we propose k-fold cross validation technique [8], dataset is divided into k subsets and the method is repeated k times. In each time, one of those k subsets is used as the test set and the other k-1 subsets are put together to form a training set. On the other hands, the percentage splitting method is to split up into two sub-datasets for training set and testing set.

F.   Knowledge Deployment

This step refers to the representation and applying the obtained model to the real usage, which will be discussed in the next section.

IV.   EXPERIMENTAL RESULTS

As described in the previous section, we will organize the experiment results follows with the step of methodology in the previous section.

A.   Data & Data Preprocessing

This research used database for customer purchases for the last 3 years (A.D. 2012-2014). The database contains three parts as follows:
- Customer profiles 1,939 records.
- Transactions of customer purchases are total 11,035 records.
- Customer types are classified as follow: Silver, Gold, and Platinum

After making a selection of data, the records which include missing values and inaccurate values are removed, and eliminated the redundant attributes. Next, the data is transformed into appropriate formats. Finally, the dataset remains 1,939 instances which are characterized by the following five fields: ID, Type of Customers, Recency (R), Frequency (F), Monetary (M).

B.   RFM Analysis

This step uses data obtained in the previous step applied with the defined the scales of R, F, M attributes as described in the previous section.

C.   Customer Segment by K-means + Type of Customers

In this process, customer types are classified as follow: Silver (S), Gold (G), and Platinum (P) as shown in Table 1. They contain type of customers silver total of 1,269 instances, type of customers gold total of 177 instances, type of customers platinum total of 493 instances.

D.   Classification Rules

This step applies the C4.5 decision tree algorithm [6] to create prediction rules of loyalty of future customers. In this paper, comparison of the results during the customer segmentation using RFM and customer segmentation using RFM with types of customers. Total 13 rules of customer segmentation using RFM and 28 rules of customer segmentation using RFM with types of customers are shown in Table II and III.

E.   Evaluate the Results

The model is evaluated by two methods in two experimental.

Firstly, the 10-fold cross validation method is applied for the evaluation on a dataset (1,939 instances). The accuracy of experiment results is 99.53% for customer segmentation using RFM and accuracy of experiment results is 99.58% for customer segmentation using RFM with types of customers.

For the percentage splitting evaluation method, it is to randomly split up into two sub-datasets: the 66% of dataset (1,280 instances) are used as a training set, and the other 34% (659 instances) are used as a testing set. The accuracy of experiment results is 99.69% for customer segmentation using RFM and the accuracy of experiment results is 99.84% for customer segmentation using RFM with types of customers.

The experimental results show that, in both evaluation methods, the proposed extended model is a little superior to the traditional approach, in term of accuracy. Additionally, the extended model provides more practical customer segmentation rules which are advantage in the final phase, the knowledge deployment.

F.   Knowledge Deployment

After creating rules for the group of customers segment, the customers will be able to see the difference in various groups of customers rather than customers that are not categorized and type of customer segmentation before clustering of the customer segment, it helps to reach more targeted customers and customer loyalty according type of customer analyze it can adjusted type of customer higher or lower than original.

Customer segmentation analysis is advantage to define a strategy, create a marketing campaign and activities as well as the products and services suitable for customers such as launching a campaign discount on 10 % for Silver or Gold customer and 20% on Platinum customers. It could create sales opportunities and to generate revenue

for the company. In addition, it also provides an opportunity to keep an existing customer, establish good customer relationship, extend CRM, and reduce advertising costs. Moreover, an organization can have an advantage over the competitors by knowing the target customer patterns.

V. CONCLUSIONS

This paper proposes an extension of the concept of RFM model applied to customer purchases database and the customer type which include 1,939 historical data. By applying RFM model to find customer loyalty according type of customer, clustering customer loyalty partition into 5 classes using k-means algorithm and assigns types of customers with 3 levels of Platinum, Gold and Silver. The extension is to bring customer type into consideration the traditional RFM approach with customer analytics to make it even better and practical customer segmentation.

The output is the generated decision rules for finding out the loyalty of new customers in the future using C4.5 decision tree algorithm. Then we compared customer segmentation by traditional RFM model and the extended RFM model. Customer loyalty according type of customer can adjust type of customer to be higher or lower than the original one. For example, company can apply the rules to boost sales, customer with high loyalty and silver type customer may be offered for the promotions if customer increase purchase, company will upgrade type of customer to Gold, more privilege than Silver. This proposed segmentation model also makes customers feel good and increasing sales for the company and it helps to reach more targeted customers. Finally, it could facilitate for maintaining and extending the system of Customer relationship management (CRM).

TABLE II. SEGMENTATION RULES OF TRADITIONAL RFM MODEL

Customer segment using RFM			
R	F	M	Loyalty
≤ 2	≤ 1	> 1	VL
≤ 2	> 1	≤ 2	VL
≤ 3	≤ 1	≤ 1	VL
> 3	≤ 1	≤ 1	L
> 4	≤ 1	≤ 2	L
≤ 2	> 1	> 2	M
= 3	≤ 1	> 1	M
= 3	> 1	-	M
= 4	≤ 1	> 1	M
= 4	> 1	-	M
> 4	≤ 1	> 2	H
> 4	> 1	≤ 2	H
> 4	> 1	> 2	VH

TABLE III. COMPARISON OF SEGMENTATION RULES BETWEEN TRADITIONAL RFM MODEL AND EXTENDED RFM MODEL

Customer segment using RFM + types of customers				
R	F	M	Types of customers	Loyalty
≤ 1	-	-	ALL[a]	VL
≤ 2	-	-	S	L
≤ 3	-	≤ 1	S	M
≤ 3	-	≤ 2	S	M
= 4	≤ 1	> 1	S	H
= 4	> 1	≤ 1	S	H
> 3	≤ 1	≤ 1	S	H
≤ 3	-	> 2	S	VH
= 4	> 1	> 1	S	VH
> 4	≤ 1	> 1	S	VH
> 4	> 1	-	S	VH
≤ 2	-	-	G	L
≤ 3	-	-	G	M
= 4	> 1	≤ 1	G	H
> 3	≤ 1	-	G	H
> 4	> 1	≤ 1	G	H
= 4	> 1	> 1	G	VH
> 4	> 1	> 1	G	VH
≤ 2	-	-	P	VL
≤ 3	-	-	P	L
= 4	≤ 1	-	P	M
= 4	> 1	≤ 1	P	M
= 4	> 1	≤ 2	P	M
> 4	≤ 1	-	P	H
> 4	≤ 2	≤ 2	P	H
= 4	> 1	> 2	P	VH
> 4	> 1	> 2	P	VH
> 4	> 2	≤ 2	P	VH

REFERENCES

[1] J. Miglautsch, "Thoughts on RFM scoring," *The Journal of Database Marketing*, vol. 8, no. 1, pp. 67–72, 2000.
[2] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, and R. Wirth, CRISP-DM 1.0 - Step-by-step data mining guide, 2000.
[3] D. Birant, "Data mining using RFM analysis, knowledge-oriented applications in data mining," Prof. Kimito Funatsu (Ed.), *InTech*, 2011.
[4] C. H. Cheng and Y. S. Chen, "Classifying the segmentation of customer value via RFM model and RS theory," *Expert Systems with Applications*, vol. 36, no. 3, pp. 4176–4184, 2009.
[5] J. MacQueen, "Some methods for classification and analysis of mulivariate observations," in *Proc. The Fifth Berkeley Symposium On Mathematical Statistics And Probability*, 1967, pp. 281-297.
[6] J. R. Quinlan, *C4.5 Programs for Machine Learning*, Morgan Kaufmann Publishers, 1993.

[7] D. Peppers and M. Rogers, *Managing Customer Relationships: A Strategic Framework*, Hoboken, New Jersey: John Wiley & Sons, 2004.

[8] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection, in *Proc. International Joint Conference on AI.*, 1995, pp. 1137–1145.

Panwad Bunnak received the B.Sc. in Computer Science from Srinakharinwirot University, Bangkok, Thailand, in 2011 and received the M.Sc. degree in Information Technology Management from Mahidol University, Nakhonpathom, Thailand, in 2015. Her current research interests include data mining in customer relationship management domain.

Sotarat Thammaboosadee was born in 1982 and received his B.Eng. and M.Sc. degrees in Computer Engineering and Technology of Information System Management from Mahidol University, Thailand, in 2003 and 2005 respectively. He also received a Ph.D. in Information Technology from King Mongkut's University Technology Thonburi, Thailand, in 2013. He is now a lecturer at Technology of Information System Management Division at Faculty of Engineering, Mahidol University, Thailand. His research interests include data mining in several domains such as the legal domain, healthcare domain and management domain. He also interests the research filed of technology valuation and business process improvement in CRM section.

Supaporn Kiattisin received her B.Eng. and M.Eng. degrees in Computer Engineering and Electrical Engineering from Chiang Mai University, Thailand, in 1996 and King Mongkut's University Technology Thonburi, Thailand, in 2000 respectively. She also received a Ph.D. in Electrical and Computer Engineering from King Mongkut's University Technology Thonburi, Thailand, in 2007. She is now a vice-director at Technology of Information System Management Division and program chair of Information Technology Management program at Faculty of Engineering, Mahidol University, Thailand. Her research interests include enterprise architecture, risk management, business process improvement, customer relationship management, and human resource management.