APPLYING DATA MINING TECHNIQUES AND EXTENED RFM MODEL IN CUSTOMER LOYALTY MEASUREMENT

PANWAD BUNNAK

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR
THE DEGREE OF MASTER OF SCIENCE
(INFORMATION TECHNOLOGY MANAGEMENT)
FACULTY OF GRADUATE STUDIES
MAHIDOL UNIVERSITY
2015

Thesis
entitled

APPLYING DATA MINING TECHNIQUES AND EXTENED RFM MODEL IN CUSTOMER LOYALTY MEASUREMENT

…………………………………………
Miss Panwad Bunnak
Candidate

…………………………………………
Lect. Sotarat Thammaboosadee,
Ph.D. (Information Technology)
Major advisor

…………………………………………
Asst. Prof. Supaporn Kiattisin,
Ph.D. (Electrical and Computer
Engineering)
Co-advisor

…………………………………………
Asst. Prof. Adisorn Leelasantitham,
Ph.D. (Electrical Engineering)
Co-advisor

…………………………………………
Prof. Patcharee Lertrit,
M.D., Ph.D. (Biochemistry)
Dean
Faculty of Graduate Studies
Mahidol University

…………………………………………
Asst. Prof. Supaporn Kiattisin,
Ph.D. (Electrical and Computer
Engineering)
Program Director
Master of Science Program in
Technology of Information System
Management
Faculty of Engineering
Mahidol University

Thesis
entitled

APPLYING DATA MINING TECHNIQUES AND EXTENED RFM MODEL IN CUSTOMER LOYALTY MEASUREMENT

was submitted to the Faculty of Graduate Studies, Mahidol University
for the degree of Master of Science (Information Technology Management)
on
March 28, 2015

………………………………………
Miss Panwad Bunnak
Candidate

………………………………………
Lect. Taweesak Samanchuen,
Ph.D. (Electrical Engineering)
Chair

……………………………………   ………………………………………
Lect. Sotarat Thammaboosadee,   Asst. Prof. Supaporn Kiattisin,
Ph.D. (Information Technology)   Ph.D. (Electrical and Computer
Member   Engineering)
    Member

……………………………………   ………………………………………
Asst. Prof. Adisorn Leelasantitham,   Asst. Prof. Kairoek Choeychuen,
Ph.D. (Electrical Engineering)   Ph.D. (Electrical and Computer
Member   Engineering)
    Member

……………………………………   ………………………………………
Prof. Patcharee Lertrit,   Lect. Worawit Isarangkul,
M.D., Ph.D. (Biochemistry)   M.S. (Technical Management)
Dean   Dean
Faculty of Graduate Studies   Faculty of Engineering
Mahidol University   Mahidol University

ACKNOWLEDGEMENTS

APPLYING DATA MINING TECHNIQUES AND EXTENED RFM MODEL IN CUSTOMER LOYALTY MEASUREMENT

PANWAD BUNNAK 5636644 EGIT/M

M.Sc. (INFORMATION TECHNOLOGY MANAGEMENT)

THESIS ADVISORY COMMITTEE: SOTARAT THAMMABOOSADEE, Ph.D., SUPAPORN KIATTISIN, Ph.D., ADISORN LEELASANTITHAM, Ph.D.

ABSTRACT

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

KEY WORDS: CUSTOMER LOYALTY/ RFM MODEL/ K-MEANS/ DECISION
            TREE/ CRM

40 pages

การประยุกต์ใช้เทคนิคทางด้านดาต้าไมน์นิ่งต่อยอดกับอาร์ เอฟ เอ็ม เพื่อหาความจงรักภักดีของลูกค้า

APPLYING DATA MINING TECHNIQUES AND EXTENED RFM MODEL IN CUSTOMER

LOYALTY MEASUREMENT

ปานวาด บุนนาค 5636644 EGIT/M

วท.ม. (การจัดการเทคโนโลยีการสารสนเทศ)

คณะกรรมการที่ปรึกษาวิทยานิพนธ์ : โยฑศ์รัตต ธรรมบุษดี, Ph.D., สุภาภรณ์ เกียรติสิน, Ph.D.,
อดิศร ลีลาสันติธรรม, Ph.D.

บทคัดย่อ
          งานวิจัยนี้มีจุดประสงค์เพื่อหาความจงรักภักดีของลูกค้าแต่ละรายเพื่อประโยชน์ในการ
สร้างแคมเปญทางการตลาดและกิจกรรมตลอดจนผลิตภัณฑ์และการบริการสำหรับลูกค้าที่
เหมาะสมและเป็นการสร้างความสัมพันธ์ที่ดีกับลูกค้า ซึ่งในงานวิจัยนี้ได้นำหลักการ RFM
(Recency = ระยะเวลาที่ลูกค้ามีการสั่งซื้อล่าสุดจนถึงปัจจุบัน, Frequency = จำนวนการสั่งซื้อในแต่
ละช่วงเวลา, Monetary = มูลค่าเฉลี่ยต่อการสั่งซื้อแต่ละครั้งในช่วงเวลาเดียวกัน) มาประยุกต์ใช้กับ
ฐานข้อมูลการซื้อของลูกค้าและประเภทของลูกค้า ประเภทธุรกิจที่เลือกเป็นธุรกิจเชิงพาณิชย์ โดย
นำหลักการRFM มาหาความจงรักภักดีของลูกค้าตามประเภทของลูกค้า ซึ่งความจงรักภักดีของ
ลูกค้าแบ่งออกเป็น 5 คลาส โดยใช้วิธีการจัดกลุ่มด้วยอัลกอริทึมk-means และกำหนดประเภทลูกค้า
เป็นแพลทินัม, โกลด์และซิลเวอร์ การนำประเภทลูกค้ามาต่อยอดกับหลักการRFM ทำให้การ
วิเคราะห์ประสิทธิภาพในการจำแนกความจงรักภักดีของลูกค้าได้ดียิ่งขึ้น สุดท้ายงานวิจัยนี้ได้สร้าง
กฎการตัดสินใจในการหาความจงรักภักดีของลูกค้ารายใหม่ในอนาคต โดยใช้วิธีการจำแนกข้อมูล
ด้วยอัลกอริทึมC4.5 หลังจากที่ได้ทำการสร้างกฎเพื่อหาความจงรักภักดีของลูกค้า การนำประเภท
ลูกค้ามาต่อยอดกับหลักการRFMทำให้ได้กฎที่มีรายละเอียดมากขึ้นกว่าแบบเดิม

40 หน้า

CONTENTS

CONTENTS (cont.)

LIST OF TABLES

LIST OF FIGURES

CHAPTER I
INTRODUCTION

1.1 Background and Significance of the problems

In business today, it is very important to be able to satisfy your customer's needs and wants, because current customers determine the direction by choosing, in a self-guided manner. If an organization is unable to satisfy the requirements of the customers, we will switch product or service provider immediately. Making our business lose the opportunity and lose customers in competition, it impacts organization income, because of the current business climate. Many organizations compete for same customers, who have the right to choose the organization rather than the organization chooses them. If the organization is able to keep an existing customer base, it needs to understand the behavior of customers. For example, an organization should record the transaction of the customers with regular orders, customers with a large order, and so on. The organization prioritizing this retaining group must also understand that customer types and levels of customer are classified as follows: wholesale' level of customers with Platinum, retail level of customers with Gold, and general customers level of customers with Silver. The organization must define a clear customer segmentation to establish relationships with customers. In addition, customer segmentation can be used to determine the appropriate marketing strategy and to apply the target groups. Therefore, customer segmentation is cleared. This will create the sales opportunities, and generates revenue for the organization from all customer segments.

Whenever the organization is able to understand and segment customers clearly, a strategy can be defined. A marketing campaign can be created with activity as well as the products and suitable services for customers. The goal is to keep an existing customer, because its cost is less than finding new customers. In addition, the use of existing database is benefiting the organization. Organization can determine the

customer's loyalty. Important organization should know that customer value is given by applying the principles of the RFM model.

RFM model, a method used for analyzing customer value, is described as follows: Recency (R) is the elapsed time since the customer made his most recent purchase; Frequency (F) is the total number of purchases in each period of time; and Monetary (M) is each order customer's average purchase amount. RFM model is a great first step to segment our customer base.

In this research, we have adopted the concept of RFM Model applied to the database of customer purchases and type of customer. In our case study, we select the commercial business for our model of organization. Then, the types of customers are assigned with 3 levels, given as: of Platinum, Gold, and Silver. Into consideration, the principles of the RFM Model with customer analytics makes it even better customer segmentation.

1.2 Research objectives

- To study and discover the customer segmentation by the data mining techniques with clustering method.

- To extend the principles of the RFM Model with database of customer purchases and assigns types of customers with 3 levels of Platinum, Gold, and Silver.

- To establish rules for the customer segmentation by using the data mining techniques based on classification method.

1.3 Delimitation of the research

- To determine the range of the R, F, and M by using the clustering data mining techniques for finding a range of data.

- To apply the data mining techniques to the database of customer purchases and type of customers. In our case study, we have selected the commercial business for our model of organization with the three-year-data during 2012-2014.

- To compare of the customer segmentation results between RFM (Traditional RFM) and RFM with types of customers (Extended RFM).

1.4 Expected outcomes and benefits

- To bring the customer segmentation analysis to define a strategy, and create a marketing campaign and activities as well as the products and suitable services for customers.

- To bring the customer segmentation analysis to create sales opportunities, and generate the revenue for the organization.

- To keep an existing customer, and establish the good customer relationships of extended customer Relationship Management (CRM).

- To reduce the advertising costs. Organization knows the target group for advertising.

- Organization would take an advantage over the competitors by knowledge of the target customer patterns.

This chapter introduces the thesis background, the significant problems, objectives for this research, delimitation, and expected outcomes. In the next chapter, we discuss the literature review and research.

CHAPTER II
LITERATURE REVIEW AND RESEARCH

In this research, applications of Data Mining Techniques and extended RFM Model for customer loyalty measurement are used. We select the organization of commercial business as our case study. Researchers have studied the concepts, theories, and related works, given as:

2.1 Customer Relationship Management (CRM)

2.2 RFM model definition

2.3 Data Mining Techniques

    2.3.1 Knowledge Discovery in Databases (KDD)

    2.3.2 The CRISP-DM process

    2.3.3 Data mining task

2.4 K-Means algorithm

2.5 C4.5 decision tree algorithm

2.6 Related works

2.1 Customer Relationship Management (CRM)

The CRM [1] provides a sense of customer product affinity for a service or organization. When customers bond better with an organization, we are more likely to remain loyal patron. The importance of customer base increases as well 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. Organization, ability to satisfy customers' needs and wants, is very important to understand customer behavior and analysis with our marketing strategy. Four implementation tasks for creating and managing customer relationships [2] are given as:

- Identification: Customer data such as customer name and contact information.

- Differentiation: Customer analysis and customers classification with groups based on customer value for the organization.

- Interaction: Studying the customer needs and establishing the long-term customer satisfaction.

- Customization: offering the suitable products/services for customers.

When organization able to satisfy customers' needs and wants. Important organization should know that customer value by applying the principles of the RFM model. The topic RFM model will be discussed in the next section.

2.2 RFM model definition

RFM [3] is a method used for analyzing customer value. It is commonly used in database marketing and direct marketing. The definition of the RFM model is described as follows:

- Recency (R) is the interval time between the time of the latest consuming behavior and present time. For example, a customer with most recent purchase last month will receive a higher recency score than a customer with most recent purchase in one year.

- Frequency (F) is the number of transactions in a particular time period. For example, customer with six purchases in one month will receive a higher frequency score than a customer with six purchases in one year.

- Monetary (M) refers to consumption amount of money in a particular time period. For example, customer purchase amount $100 would receive a higher monetary score than a customer purchase amount $20 in one year.

According to the research proposed by Birant [4], this research provides an overview RFM that can be applied to the field of Data Mining. Therefore, in this research, the proposed data mining consists of: the clustering task with K-Means algorithm, classification rules with C4.5 Decision Tree algorithm [5], and the customer segment classification with RFM.

2.3 Data Mining Techniques

2.3.1 Knowledge Discovery in Databases (KDD)

Knowledge Discovery in Databases is a technique to discover the patterns and hidden relationships in the massive amounts of data. From Figure 2.1, the components of subsequence process are as follows:
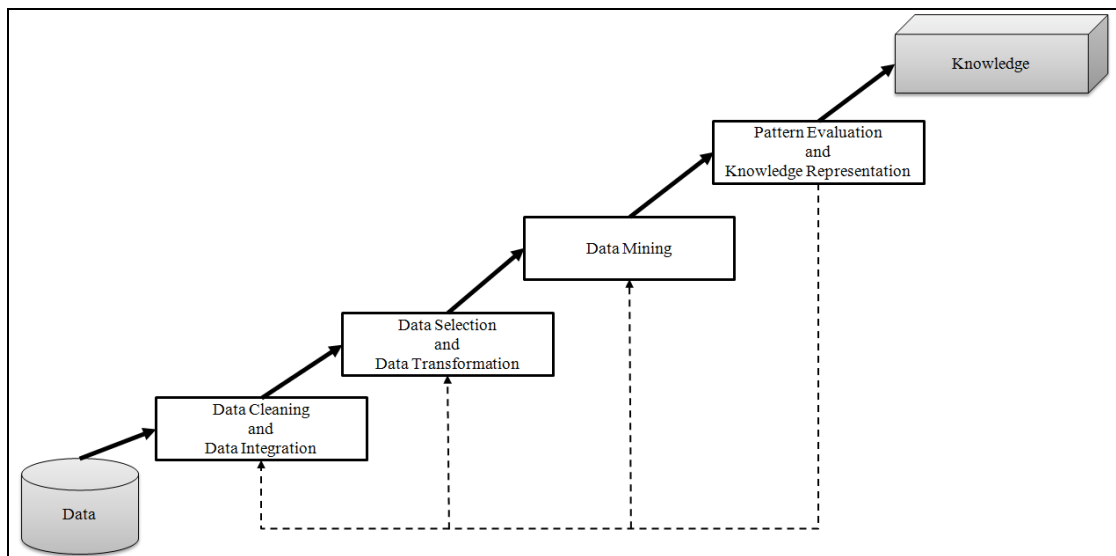
Figure 2.1 Process of Knowledge Discovery in Databases.

- Data cleaning is the procedure for data screening and filtering for the data with noise, outliers, and irrelevant data.

- Data integration is the process of recombining resources from many resources.

- Data selection is the analysis data from the recorded data.

- Data transformation is the conversion process of selected data to suit the data mining process.

- Data mining is the process for extracting patterns to take the benefit of data.

- Pattern evaluation is a process to evaluate the pattern result taken from the data mining in order to get the model representing the knowledge required.

- Knowledge representation is the process of knowledge utilization to find the solutions from the data mining.

2.3.2 The CRISP-DM process

From Figure 2.2, the CRISP-DM [6] process consists of a cycle process with six stages, given as:
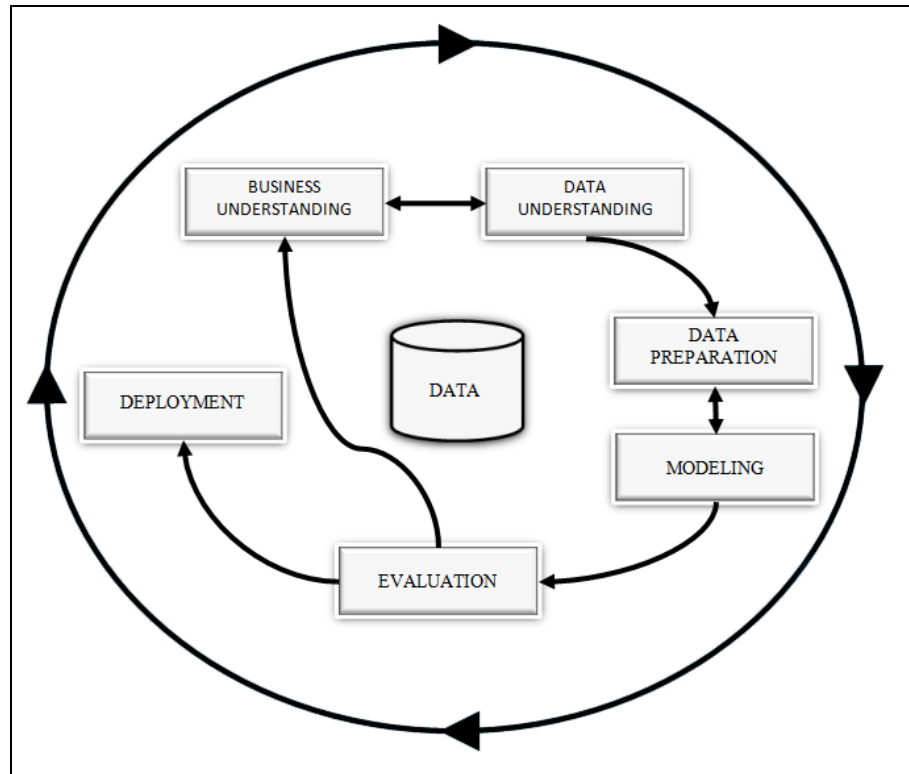
Figure 2.2 The CRISP-DM life cycle.

- The first step is to understand and identify the business problem/opportunity. Then, the knowledge is converted into a suitable form for implementation of data mining.

- The second step to collect the relevant data into the data mining techniques for analysis. Under the data gathering process, the suitable data source for the data mining analysis should be accurate, reliable, sufficient and appropriate.

- The third step, the data preparation, usually takes a longer time and higher important. The model, derived from the data mining with accurate results, depends on the quality of data. The provisions of data are such as data cleaning, data integration, data selection, and data transformation.

- The fourth step is modeling process with data mining techniques.

- The fifth step is to evaluate the performance of the analyzed models of prior step. In this process, the evaluated results of models cover the data analysis to meet the business set of the first step. Models of multiple algorithms may be evaluated for their pros and cons to find the most appropriate model for applying the business.

- The final step is to conduct the knowledge result obtained from the data analysis with data mining techniques to enhance the knowledge of data.

2.3.3 Data mining task

Data mining task [7] can be divided into two approaches including predictive modeling and descriptive modeling. The descriptions for both of models are as follows:

- Predictive modeling (or Supervised modeling): It is to predict the behavior from the occurrence in the past, and creates the prototype for the prediction of the new data or subsequent information. Relevant data mining methods are given as:

   o Classification: It is discovery of a predictive learning function, process to create rule for management data. To predict the potential information by creating rules to make decisions based on the available information, the classification is commonly represented by decision trees.

   o Regression: It is discovery of a predictive learning function, which maps a data item to a real-value prediction variable.

   o Deviation Detection: It is the discovery of a predictive learning function, the most significant changes in the data set.

- Descriptive modeling (or Unsupervised modeling): It is not intended to predict, but to describe data model for use as guidelines in making decisions. Relevant data mining methods are given as:

   o Clustering: It is common descriptive task it will be divided into groups of similar class without defining the type of information it before or do not know the number of clusters in advance.

   o Association Rule: It is common descriptive task is determining the relationship of two or more within the larger group.

2.4 K-Means algorithm

K-means clustering [8], the descriptive modeling, is one of the simplest unsupervised learning algorithms consisting of the following steps:

- First step: Place the K points into the space. This point represent the initial group of centroids.

- Second step: Assign each object to the group having the closest centroid.

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

- Final step: Repeat to the second step and third step until the centroids are no longer move.

2.5 C4.5 decision tree algorithm

C4.5 is an algorithm used to generate a decision tree developed by Ross Quinlan [9]. A decision tree is a tree like the flow chart structure. Decision tree models include such concepts as root node, internal node, leaf node, and branch. The root node represents the top-most node, that begins with creating the tree with all training data and extracts a node until leaf node. Then, it recombines and removes the branch that affects the accuracy.

2.6 Related works

We have studied the literature review and related works. The details as follows:

- Birant [4] presented the overview of RFM applied to the field of Data Mining by three approaches including: clustering, classification, and association rule in order to provide the market intelligent. Moreover, the developed RFM model was better marketing strategies.

- Cheng et al. [5] proposed the quantitative value of RFM attributes and K-means algorithm into rough set theory (RS theory). The cluster was partitioned into 3, 5, and 7 classes, and classification rule were used by three approaches including Decision tree, Neural network, and Naïve Bayes.

- Chang et al. [10] proposed the novel framework of GRFM (for group RFM) analysis to discover the customer consumption behavior. By division of the RFM different types of goods, the clustering result of PICC (Purchased Items-Constrained Clustering) contained the extra information. It could help to decision.

- Wei et al. [11] proposed the novel framework of LRFM (length, recency, frequency, and monetary) model and the extended RFM (recency, frequency, and monetary) model. By adding the variables L to the model of RFM, L is the first visit date to consider the core patients.

Based on the review of literature; in this research, we have adapted the concept of the RFM model, called Traditional RFM Model, and we have applied it to database of customer purchases and the customer type, called Extended RFM Model. By using K-Means algorithm and C4.5 Decision Tree algorithm.

In this chapter we discuss about concepts, theories and related works for this research. In the next chapter we discuss the research methodology.

CHAPTER III

METHODOLOGY

3.1 The steps of the research process

This research applies the data mining techniques to database of customer purchases for Customer Loyalty. From Figure 3.1, the processes of the research process are as follows:
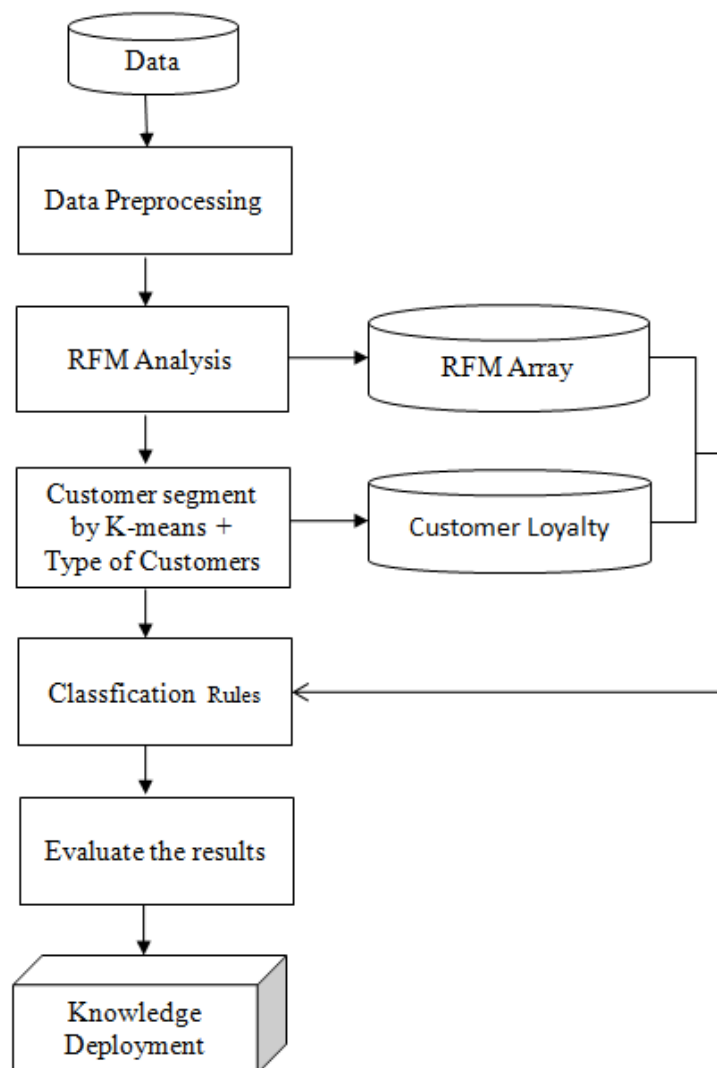
Figure 3.1 Research Methodology.

3.1.1 Data and Data Preprocessing

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

3.1.2 RFM Analysis

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

- First step: Sort the data of three R, F, M attributes by descending or ascending order.

- Second step: Partition those of 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. The level score of 5 represents to the most customer loyalty, whereas the level score of 1 represents to the least customer loyalty.

- Third step: Repeat the previous sub-processes (First step and Second step) for each R, F, and M attribute individually. 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), as shown in Figure 3.2.

R	F	M	R	F	M	R	F	M	R	F	M	R	F	M
5	5	5	4	5	5	3	5	5	2	5	5	1	5	5
5	5	4	4	5	4	3	5	4	2	5	4	1	5	4
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
5	1	1	4	1	1	3	1	1	2	1	1	1	1	1

Figure 3.2 Total RFM Array with 125 combinations.

3.1.3 Customer segment by K-means with 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 3.1, the customer type can be classified as: Silver (S), Gold (G), and Platinum (P).

Table 3.1 Three Customer Type Groups of RFM Array.

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

The procedure in this phase is that: firstly, customer loyalty is partitioned according type of customer 5 clusters by using the K-means algorithm. Then, it is 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 to the type of customer. Finally, there will be customer loyalty divided into 3 groups according to the type of customers.

3.1.4 Classification Rules

The rules of classification are discovered from RFM, customer type, and the clustering results obtained in the previous step, as exemplified in Figure 3.3. Moreover, the prediction rules of loyalty of future customers are created and discussed in the next chapter.

ID	R	F	M	Type of Customers	Loyalty of Customers
0001	4	1	1	Gold	High
0002	5	1	1	Platinum	High
0003	5	3	2	Silver	Very High
0004	5	1	1	Silver	High
0005	5	2	2	Silver	Very High
0006	3	1	1	Silver	Medium
0007	5	1	1	Platinum	High
0008	1	1	1	Silver	Very Low
0009	4	2	1	Silver	High
0010	2	1	1	Silver	Very Low
⋮	⋮	⋮	⋮	⋮	⋮

Figure 3.3 An Example of Decision table.

3.1.5 Evaluation Results

In our model, we propose K-fold cross validation technique [12], 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.

3.1.6 Knowledge Deployment

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

3.2 Schedule

Schedule of this research has start at August and uses 5 months for thesis with any topic in Figure 3.4.

No.	Research stages	August	September	October	November	December
1	Study the theory and Related works	←→				
2	Data preprocessing		←→			
3	Model Building and Evaluation			←→		
4	Knowledge Deployment			←——→		
5	Research conclusions				←——→	

Figure 3.4 Schedule of research.

In this chapter we discuss about steps of the research process. In the next chapter we discuss analysis, results and knowledge deployment for organizations.

CHAPTER IV

ANALYSIS AND RESULTS

As described in the previous chapter, we will organize the experiment results following with the step of methodology in the previous chapter, as shown in Figure 3.1.

4.1 Data and Data Preprocessing

This research uses the database for customer purchases and type of customer. In our case study, we selected the commercial business for our model of organization using the three-year-data during 2012-2014. The database contains three parts as follows:

- Customer profiles with 1,939 records. An example of customer profiles data is shown in Figure 4.1;
- Total number of customer purchases in transaction are 1,035 records:
- Customer types are classified as follows: Silver, Gold, and Platinum.

ID	Type of Customers	R	F	M
⋮	⋮	⋮	⋮	⋮
0200009	S	200	4	33,276
0400348	S	295	1	7,880
0400051	G	156	2	10,115
0400337	S	15	7	73,460
0400470	S	17	10	126,689
0701461	G	105	10	43,905
0400285	S	15	1	1,176
0500170	P	1	225	1,026,090
0400395	S	554	1	9,240
0400399	S	134	10	34,235
⋮	⋮	⋮	⋮	⋮

Figure 4.1 The Partial Dataset.

4.2 RFM Analysis

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

- First step: Sort the data of three R, F, M attributes by descending or ascending order.

- Second step: Partition those of three attributes by K-means algorithm with five clusters criteria, as shown in Table 4.1. These five parts are assigned for the score as 5, 4, 3, 2, and 1 which represent to the customer contributions for organization. The level score of 5 represents to the most customer loyalty, whereas the level score of 1 represents to the least customer loyalty.

- Third step: Repeat the previous sub-processes (First step and Second step) for each R, F, M attribute individually. There are total RFM Array of 125 (5x5x5) combinations, since each attribute in R, F, M attributes has 5 scales (5, 4, 3, 2 and 1) on output as input attributes for each customers, as shown in Figure 4.2. Customers in the 555 segment are the most valuable, whereas those in the 111 segment are the least valuable.

Table 4.1 Five scaling of RFM attributes in dataset by K-means.

Scaling	Scaling name	R – Recency (Day)	F – Frequency	M – Monetary
5 Score	Very High	under 102	over 98	over 1,240,186
4 Score	High	103-237	50-97	534,027-1,240,185
3 Score	Medium	238-396	21-49	220,565 - 534,026
2 Score	Low	370-507	6-20	61,571 - 220,565
1 Score	Very Low	over 508	under 5	under 61,570

ID	R	F	M
⋮	⋮	⋮	⋮
0200009	4	1	1
0400348	3	1	1
0400051	4	1	1
0400337	5	2	2
0400470	5	2	2
0701461	4	2	1
0400285	5	1	1
0500170	5	5	4
0400395	1	1	1
0400399	4	2	1
⋮	⋮	⋮	⋮

Figure 4.2 Quantitative value of RFM attributes for dataset.

4.3 Customer segment by K-means with 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 3.1, the customer type can be classified as: Silver (S), Gold (G), and Platinum (P). We contain the type of silver customers is with the total of 1,269 instances, as shown in Table 4.2; type of gold customers is with the total of 177 instances, as shown in Table 4.3; and type of platinum customers is with the total of 493 instances, as shown in Table 4.4. Moreover the result of traditional RFM, as shown in Table 4.5.

Table 4.2 Five cluster results by K-means with type of silver customer.

Cluster center	C1	C2	C3	C4	C5
R	4.46	4.86	3	1	2
F	1.05	2.11	1.05	1.02	1.03
M	1.01	1.97	1.05	1.02	1.08
Loyalty (output)	High	Very High	Medium	Very Low	Low
Number of instances	470	144	239	163	253

Table 4.3 Five cluster results by K-means with type of gold customer.

Cluster center	C1	C2	C3	C4	C5
R	2	3	1	4.43	4.55
F	1	1.06	1	1.08	2.27
M	1.05	1.03	1.04	1.11	2.55
Loyalty (output)	Low	Medium	Very Low	High	Very High
Number of instances	38	36	27	65	11

Table 4.4 Five cluster results by K-means with type of platinum customer.

Cluster center	C1	C2	C3	C4	C5
R	4	4.94	3	1.53	5
F	1.23	3.16	1.11	1.04	1.36
M	1.20	3.31	1.11	1.04	1.29
Loyalty (output)	Medium	Very High	Low	Very Low	High
Number of instances	84	86	70	97	156

Table 4.5 Five cluster results by K-means with traditional RFM.

Cluster center	C1	C2	C3	C4	C5
R	4.53	3.67	4.95	5.00	2.09
F	1.00	1.77	3.14	2.10	1.02
M	1.04	1.74	3.50	1.59	1.03
Loyalty (output)	Low	Medium	Very High	High	Very Low
Number of instances	669	119	101	162	888

4.4 Classification Rules

This step applies the C4.5 decision tree algorithm to create the prediction rules of loyalty of future customers. In this research, comparison of the results during the customer segmentation using RFM (Traditional RFM) is shown in Figure 4.3, and the customer segmentation using RFM with types of customers (Extended RFM) is shown in Figure 4.4. Total number of 13 rules of customer segmentation using RFM is shown in Table 4.6, and the 28 rules of customer segmentation using RFM with types of customers is shown in Table 4.7.
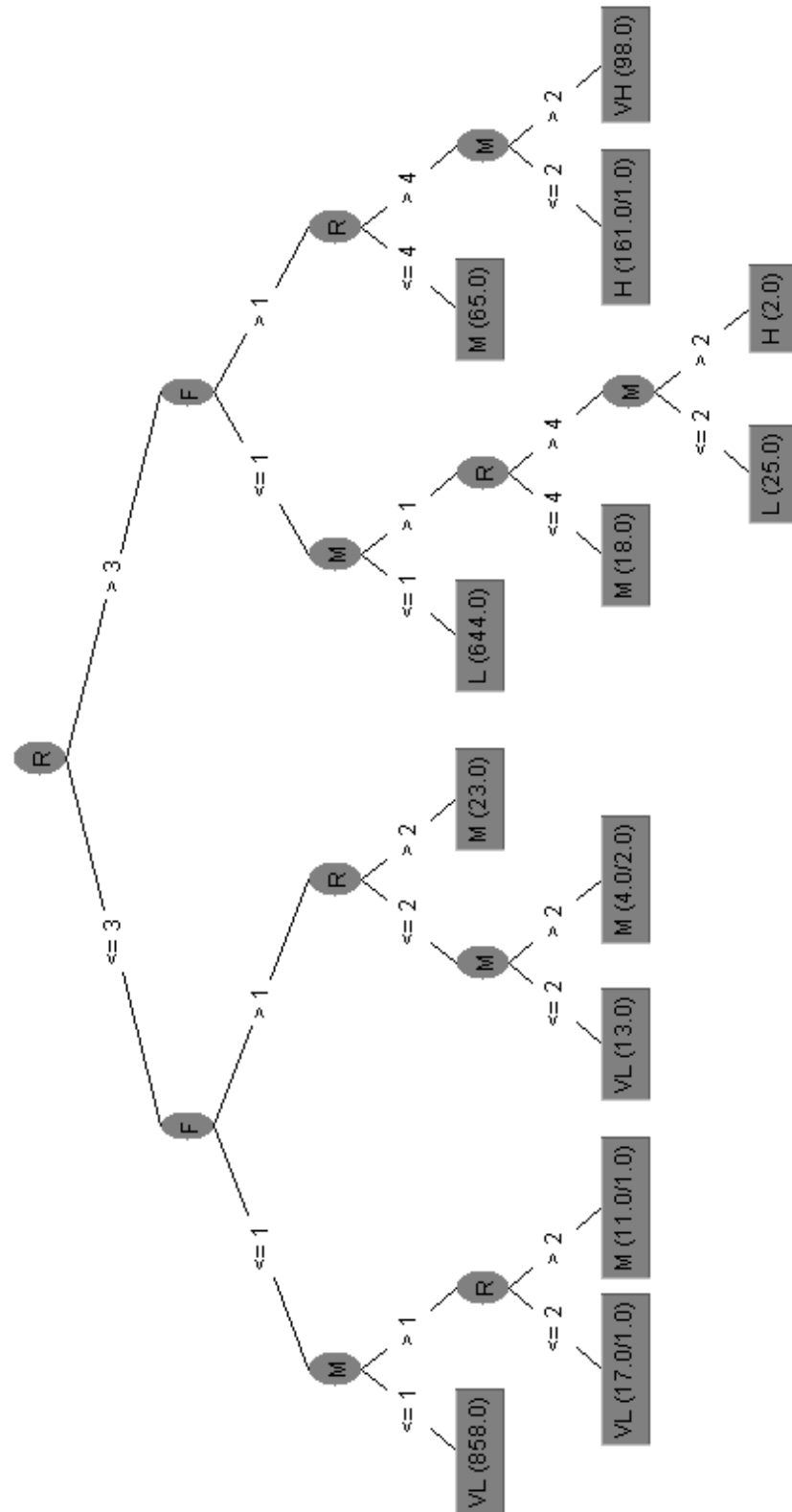
Figure 4.3 Decision tree: Traditional RFM.
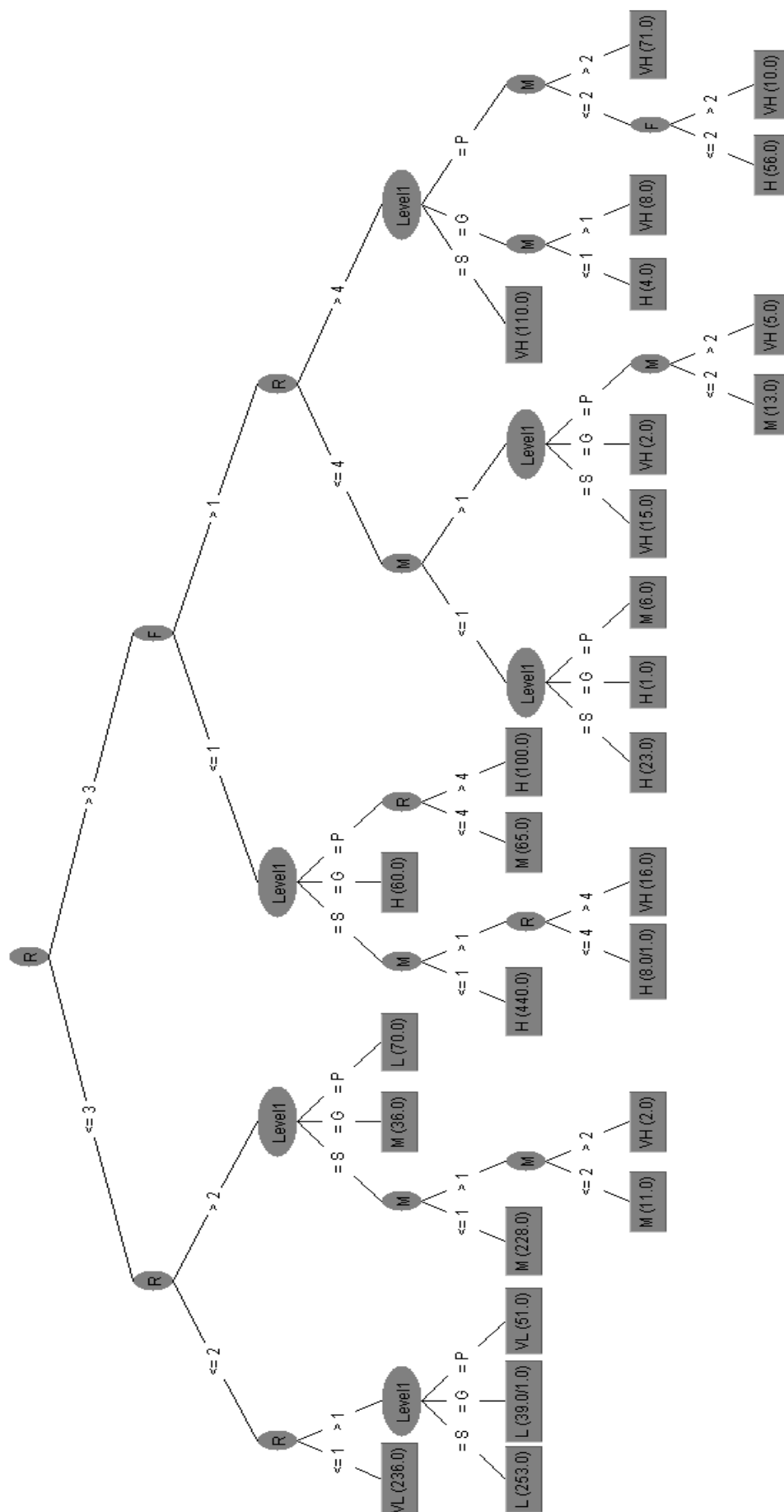
Figure 4.4 Decision tree: Extended RFM.

Table 4.6 Thirteen rules of Traditional RFM.

	Customer segment using RFM			
No.	R	F	M	Loyalty
1	≤ 2	≤ 1	> 1	VL
2	≤ 2	> 1	≤ 2	VL
3	≤ 3	≤ 1	≤ 1	VL
4	> 3	≤ 1	≤ 1	L
5	> 4	≤ 1	$= 2$	L
6	≤ 2	> 1	> 2	M
7	$= 3$	≤ 1	> 1	M
8	$= 3$	> 1	-	M
9	$= 4$	≤ 1	> 1	M
10	$= 4$	> 1	-	M
11	> 4	≤ 1	> 2	H
12	> 4	> 1	≤ 2	H
13	> 4	> 1	> 2	VH

Determined segment customer loyalty to each class, which are "Very High" (VH), "High" (H), "Medium" (M), "Low" (L), and "Very Low" (VL).

Table 4.7 Twenty-eight rules of Extended RFM.

	Customer segment using RFM + Types of customers				
No.	R	F	M	Types of customer	Loyalty
1	= 1	-	-	ALL	VL
2	= 2	-	-	S	L
3	= 3	-	≤ 1	S	M
4	= 3	-	= 2	S	M
5	= 4	≤ 1	> 1	S	H
6	= 4	> 1	≤ 1	S	H
7	> 3	≤ 1	≤ 1	S	H
8	= 3	-	> 2	S	VH
9	= 4	> 1	> 1	S	VH
10	> 4	≤ 1	> 1	S	VH
11	> 4	> 1	-	S	VH
12	= 2	-	-	G	L
13	= 3	-	-	G	M
14	= 4	> 1	≤ 1	G	H
15	> 3	≤ 1	-	G	H
16	> 4	> 1	≤ 1	G	H
17	= 4	> 1	> 1	G	VH
18	> 4	> 1	> 1	G	VH
19	= 2	-	-	P	VL
20	= 3	-	-	P	L
21	= 4	≤ 1	-	P	M
22	= 4	> 1	≤ 1	P	M
23	= 4	> 1	= 2	P	M
24	> 4	≤ 1	-	P	H
25	> 4	= 2	≤ 2	P	H
26	= 4	> 1	> 2	P	VH
27	> 4	> 1	> 2	P	VH
28	> 4	> 2	≤ 2	P	VH

Determined segment customer loyalty to each class, which are "Very High" (VH), "High" (H), "Medium" (M), "Low" (L), and "Very Low" (VL). Type of customers represents by Silver" (S), "Gold" (G), "Platinum" (P) and "All types of customers include Silver, Gold, and Platinum" (ALL).

4.5 Evaluation Results

The model is evaluated by two methods in two experimental schemes, as shown in Table 4.8.

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 of splitting evaluation method, it is to randomly split up into two sub-datasets: the 66% of dataset (1,280 instances) is used as a training set, and the other of 34% (659 instances) is 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.

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

Table 4.8 Comparisons of Accuracy Results between Traditional RFM Model and Extended RFM Model.

Methods	10-fold cross validation	Percentage of splitting
Customer segment using RFM (Traditional RFM)	99.53%	99.69%
Customer segment using RFM + types of customers (Extended RFM)	99.58%	99.84%

4.6 Knowledge Deployment

After creating rules for the group of customers segment, the customers will be able to see the differences in various groups of customers rather than the uncategorized customers data with the type of customer segmentation. It helps to reach the more targeted customers and customer loyalty according the type of customer analysis. Moreover, it can upgrade or downgrade the type of customer, it depends on the strategies for the individual organizations, as shown in Table 4.9.

Table 4.9 Example for the use of rule.

Types of customer	Loyalty	NEW Types of customer
S	VH	G (Upgrade)
P	VL	G (Downgrade)

Customer segmentation analysis, the advantage for defining a strategy, creates a marketing campaign and activities as well as the products and suitable services for customers. For example, the marketing campaigns for customer are given as: launching a campaign discount of 10 % for Silver and Gold customer, and 20% discount for Platinum customers. To increase the recency and frequency, we isolate the customer segments having the monetary score of 5. It could create the sales opportunities, and generates the revenue for the organization. In addition, it also provides an opportunity to keep an existing customer, establishes a good customer relationship, extends the CRM, and reduces the advertising costs. Moreover, the organization would takes the advantage over the competitors by knowledge of the target customer patterns.

In this chapter, we discuss the experimental results, analysis, and knowledge of deployment for organizations. In the next chapter, the conclusion and future works for research are also given.

CHAPTER V
CONCLUSION AND DISCUSSION

5.1 Conclusion and Discussion

This research proposes the extended concept of RFM model for the customer purchases database and the customer type with the historical data of 1,939 customers. By applying RFM model to find the customer loyalty according the type of customer, clustering customer loyalty is partitioned into 5 classes by k-means algorithm, and assigns the types of customers with 3 levels of Platinum, Gold, and Silver. The extension is to the bring customer type into consideration of 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 by the C4.5 decision tree algorithm. Then, we have compared the customer segmentation by traditional RFM model and the extended RFM model. Customer loyalty according to the type of customer is able to upgrade or downgrade the type of customer. For example, company can apply the rules to boost the sales, the high loyalty customer and silver customer may be offered for the promotions. If customers increase the purchase, company will upgrade the higher level type of customer to Gold level, which is more privilege than Silver level. This proposed segmentation model also makes the customers feeling good, and increases their sales for the company. Moreover, it would help to reach the more targeted customers. Finally, it could facilitate for maintaining and extending the system of customer relationship management (CRM).

5.2 Future works

Our research points to the several interesting directions for the future work as follows:

First point: Survey the customer satisfaction about assigning of types of customers with 3 levels of Platinum, Gold, and Silver. Then, evaluate the customer satisfaction for the customer level campaign. This is because the customer satisfaction depends on the responses to survey and extend the CRM.

Second point: Additional factor for creating rules of the type of customer such as: authorized capital, size of organization, etc. This makes a rule with high accurate, and increases the target of organization.

Final point: With several kinds of algorithm in classification process of data mining, we use C4.5 decision tree algorithm for making the rule. However, we can explore a new algorithm to upgrade the classification process of data mining in order to improve the solution quality.

REFERENCES

1 Thompson B. and Sims D. (2002), CRM improving demand chain intelligence for competitive advantage. Business Week, 3804, pp. 75–82.

2 Peppers D. and Rogers M. (2004), Managing Customer Relationships: A Strategic Framework. Hoboken, New Jersey: John Wiley & Sons.

3 Miglautsch, J. (2000), Thoughts on rfm scoring. Journal of Database Marketing, pp. 67–72.

4 Derya Birant (2011). Data Mining Using RFM Analysis, Knowledge-Oriented Applications in Data Mining, Prof. Kimito Funatsu (Ed.), ISBN: 978-953-307-154-1, InTech, DOI: 10.5772/13683.

5 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, pp. 4176–4184.

6 Chapman P., Clinton J., Kerber R., Khabaza T., Reinartz T., Shearer C. and Wirth R. (2000), CRISP-DM 1.0 - Step-by-step data mining guide.

7 Kantardzic, M. (2011). Data mining: concepts, models, methods, and algorithms. John Wiley & Sons.

8 MacQueen J. (1967), Some methods for classification and analysis of mulivariate observations. In: L.M. LeCam, J. Neyman (eds.): Proc. 5th Berkely Symp. Math. Statist. Probab. 1965/66. Univ. of California Press, Berkely, vol. I, pp. 281-297.

9 Quinlan J. R. (1993), C4.5 Programs for Machine Learning, Morgan Kaufmann Publishers. 302 pages.

10 Chang, Hui-Chu, and Hsiao-Ping Tsai. "Group RFM analysis as a novel framework to discover better customer consumption behavior." Expert Systems with Applications 38.12 (2011): 14499-14513.

11 Wei, J. T., Lin, S. Y., Weng, C. C., & Wu, H. H. (2012). A case study of applying LRFM model in market segmentation of a children's dental clinic. Expert Systems with Applications, 39(5), 5529-5533.

12 Kohavi R. (1995), A study of cross-validation and bootstrap for accuracy estimation and model selection. In Proceedings of International Joint Conference on AI., pp. 1137–1145.

APPENDIX

TECHNICAL PAPER OF 2015 INTERNATIONAL CONFERENCE ON INFORMATION TECHNOLOGY

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 73170, Thailand

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.

Keywords: Customer loyalty, RFM Model, k-means, decision tree, CRM.

1. 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.

2. Related works

2.1 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.

2.2 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.

2.3 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.

2.4 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.

2.5 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.

3. 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 Figure 1.
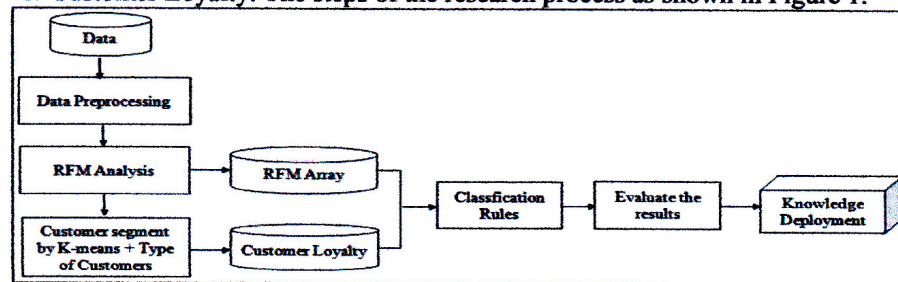

Fig.1: Overall Methodology

3.1 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.

3.2 RFM Analysis

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

Step1: Sorting the data of three R, F, M attributes by descending or ascending order.

Step2: 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.

Step3: *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).

3.3 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 1, the customer type can be classified as: Silver (S), Gold (G), and Platinum (P).

Table 1: 3 Customer type 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.

3.4 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.

3.5 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.

3.6 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.

4.   Experimental Results

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

4.1 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).

4.2 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.

4.3 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

4.4 Classification Rule

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 2.

4.5 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.

4.6 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.

Table 2: Comparison of segmentation rules between traditional RFM model and extended RFM model.

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

5. 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).

References

[1] Miglautsch, J. (2000), Thoughts on rfm scoring. Journal of Database Marketing, pp. 67–72.

[2] Chapman P., Clinton J., Kerber R., Khabaza T., Reinartz T., Shearer C. & Wirth R. (2000), CRISP-DM 1.0 - Step-by-step data mining guide.

[3] Derya Birant (2011). Data Mining Using RFM Analysis, Knowledge-Oriented Applications in Data Mining, Prof. Kimito Funatsu (Ed.), ISBN: 978-953-307-154-1, InTech, DOI: 10.5772/13683.

[4] Cheng C. H. & Chen Y. S. (2009), Classifying the segmentation of customer value via RFM model and RS theory. Expert Systems with Applications, 36, pp. 4176–4184.

[5] MacQueen J. (1967), Some methoods for classification and analysis of mulivariate observations. In: L.M. LeCam, J. Neyman (eds.): Proc. 5th Berkely Symp. Math. Statist. Probab. 1965/66. Univ. of California Press, Berkely, vol. I, pp. 281-297.

[6] Quinlan J. R. (1993), C4.5 Programs for Machine Learning, Morgan Kaufmann Publishers. 302 pages.

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

[8] Kohavi R. (1995), A study of cross-validation and bootstrap for accuracy estimation and model selection. In Proceedings of International Joint Conference on AI., pp. 1137–1145.

BIOGRAPHY

NAME	Miss Panwad Bunnak
DATE OF BIRTH	19 April 1990
PLACE OF BIRTH	Bangkok, Thailand
INSTITUTIONS ATTENDED	Srinakharinwirot University, 2008-2011
	Bachelor of Science (Computer Science)
	Mahidol University, 2013-2015
	Master of Science (Information Technology Management)
RESEARCH GRANTS	Grant to Support Graduate Students in Academic Presentations in Singapore Academic Year 2015
HOME ADDRESS	12/263 M.12 Ma Charoen Road, Nongkham, Nongkham, Bangkok, Thailand, 10160
	Tel. 089-134-5314
	E-mail : ohana_mylife@hotmail.com
PUBLICATION / PRESENTATION	Bunnak, P., Thammaboosadee, S., and Kiattisin, S. (2015), Applying Data Mining Techniques and extended RFM Model in Customer Loyalty Measurement, The Proceedings of the 2015 International Conference on Information Technology (ICIT 2015), pp. 288-295. ISSN: 1743-3517.