

**CUSTOMER RELATIONSHIP MANAGEMENT STRATEGIES  
DISCOVERED BY CLUSTERING TECHNIQUE AND BUSINESS  
INTELLIGENCE**

**THANAKAL YOTSOMSAK**

**A THEMATIC PAPER SUBMITTED IN PARTIAL  
FULFILLMENT OF THE REQUIREMENTS FOR  
THE DEGREE OF MASTER OF SCIENCE  
(INFORMATION TECHNOLOGY MANAGEMENT)  
FACULTY OF GRADUATE STUDIES  
MAHIDOL UNIVERSITY  
2016**

**COPYRIGHT OF MAHIDOL UNIVERSITY**

Thematic Paper  
entitled

**CUSTOMER RELATIONSHIP MANAGEMENT STRATEGIES  
DISCOVERED BY CLUSTERING TECHNIQUE AND BUSINESS  
INTELLIGENCE**

*Thanakal Yotsomsak*

.....  
Miss. Thanakal Yotsomsak  
Candidate

*Sotarot et.*

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

*[Signature]*

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

*[Signature]*

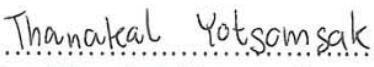
*Patcharee Lertrit*


.....  
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  
Information Technology Management  
Faculty of Engineering  
Mahidol University

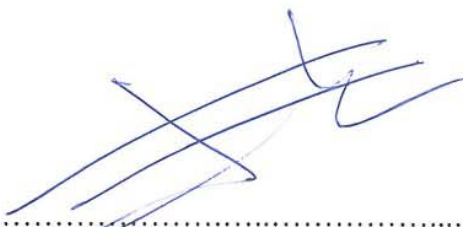
Thematic Paper  
entitled  
**CUSTOMER RELATIONSHIP MANAGEMENT STRATEGIES  
DISCOVERED BY CLUSTERING TECHNIQUE AND BUSINESS  
INTELLIGENCE**


was submitted to the Faculty of Graduate Studies, Mahidol University  
for the degree of Master of Science  
(Information Technology Management)  
on  
January 4, 2016

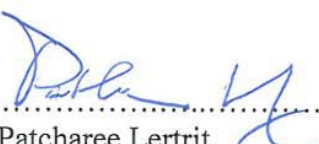
  
.....  
Miss. Thanakal Yotsomsak  
Candidate


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

  
.....  
Lect. Sotarathammaboosadee, Ph.D.  
(Information Technology)  
Member

  
.....  
Asst. Prof. Supaporn Kiattisin, Ph.D.  
(Electrical and Computer Engineering)  
Member

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

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

  
.....  
Asst. Prof. Jackrit Suthakorn, Ph.D.  
(Robotics)  
Dean  
Faculty of Engineering  
Mahidol University

## ACKNOWLEDGEMENTS

This thematic paper would not be success without advices, supports and encouragements from many people. I have to thank my research advisors Dr.Sotarat Thammaboosadee and Asst. Prof. Supaporn Kiattisin, who suggest and encourage me through the research topic and the necessary knowledge of data mining, strategic organization and tools utilization. They gave me a good advices and taught me about thinking process.

Most importantly, none of this could be happened without my family, my friends and colleagues who always stay beside and encourage me when I was under pressure and discouraged. Although they do not know deeply in this topic, they tend to support me as much as possible.

Thanakal Yotsomsak

**CUSTOMER RELATIONSHIP MANAGEMENT STRATEGIES DISCOVERED BY  
CLUSTERING TECHNIQUE AND BUSINESS INTELLIGENCE**

**THANAKAL YOTSOMSAK 5737277 EGIT/M**

**M.Sc. (INFORMATION TECHNOLOGY MANAGEMENT)**

**THEMATIC PAPER ADVISORY COMMITTEE: SOTARAT**

**THAMMABOOSADEE, Ph.D., SUPAPORN KIATTISIN, Ph.D.**

**ABSTRACT**

Customer Relationship Management (CRM) is the key management tools to gain the sustainable competitive advantage and to survive business among the intense competitive environment. Before building up CRM strategies, the organization should prioritize the customers into segment in order to manage the relationship and to provide the services properly. This research aims to develop CRM strategies in case study; Chemical Industries applying the clustering techniques to segment customer, use Business Intelligence (BI) as the visualized tools to represent the knowledge in various dimensions and finally bring out CRM strategies to deploy in business. According to the customer data used in clustering, the RFM model was applied and added to the existing attributes. The clustering methods applied in this research are K-Means and EM algorithm. The clusters are generated and defined as 4 classes, including Diamond, Platinum, Gold, and Silver, respectively. Each cluster is different in characteristic, but its interpretation still not well enough to make the decision without powerful presentation tools like BI. Finally, the knowledge acquired from dimensional aspects in BI could lead to CRM strategies. The Organization could be strengthen the good relationship with individual customer, and appropriately provides the personalized products and services based on customer segment.

**KEY WORDS: CUSTOMER RELATIONSHIP MENAGEMENT (CRM) /**

**CLUSTERING TECHNIQUES / BUSINESS INTELLIGENCE (BI) /**

**CRM STRATEGIES / RFM MODEL**

กลยุทธ์ในการบริหารจัดการความสัมพันธ์ลูกค้าจากการใช้เทคนิคการแบ่งกลุ่มและธุรกิจอัจฉริยะ  
CUSTOMER RELATIONSHIP MANAGEMENT STRATEGIES DISCOVERED BY  
CLUSTERING TECHNIQUE AND BUSINESS INTELLIGENCE

ชนกัลย์ ยศสมศักดิ์ 5737277 EGIT/M

วท.ม. (การจัดการเทคโนโลยีสารสนเทศ)

คณะกรรมการที่ปรึกษาสารนิพนธ์: โยทศรัตต ชรรณบุษดี, Ph.D., สุภาภรณ์ เกียรติสิน, Ph.D.

บทคัดย่อ

การบริหารจัดการความสัมพันธ์ลูกค้าเป็นเครื่องมือในการบริหารงานที่สำคัญที่สามารถสร้างความได้เปรียบในการแข่งขันและช่วยให้องค์กรทางธุรกิจสามารถอยู่รอดได้ท่ามกลางการแข่งขันที่รุนแรงในปัจจุบัน ก่อนที่จะกำหนดกลยุทธ์ในการบริหารจัดการความสัมพันธ์ลูกค้า องค์กรควรจะต้องจัดแบ่งลูกค้าออกเป็นกลุ่มเพื่อให้การบริหารความสัมพันธ์และการให้บริการได้อย่างเหมาะสมกับลูกค้าแต่ละกลุ่ม งานวิจัยเรื่องนี้มีวัตถุประสงค์ที่จะใช้เทคนิคการทำเหมืองข้อมูลในการแบ่งกลุ่มลูกค้า, ใช้เครื่องมือธุรกิจอัจฉริยะในการนำเสนอความรู้ที่ได้ในแง่มุมต่างๆซึ่งนำไปสู่การกำหนดกลยุทธ์ในการบริหารจัดการความสัมพันธ์ลูกค้า โดยการแบ่งกลุ่มลูกค้าได้ประยุกต์ใช้ K-Means และ RFM โมเดลเพื่อแบ่งกลุ่มลูกค้าออกเป็น 4 ระดับคือ ไดมอนด์, แพลตตินั่ม, โกลด์ และซิลเวอร์ แต่ละกลุ่มลูกค้าจะมีคุณลักษณะแตกต่างกันไป เมื่อนำมาประยุกต์ใช้ร่วมกับธุรกิจอัจฉริยะทำให้ค้นพบแง่มุมทางธุรกิจอันนำไปสู่แนวทางในการกำหนดกลยุทธ์ที่มีประสิทธิภาพในการบริหารจัดการความสัมพันธ์ลูกค้าแต่ละกลุ่ม

## CONTENTS

	<b>Page</b>
<b>ACKNOWLEDGEMENTS</b>	<b>iii</b>
<b>ABSTRACT (ENGLISH)</b>	<b>iv</b>
<b>ABSTRACT (THAI)</b>	<b>v</b>
<b>LIST OF TABLES</b>	<b>viii</b>
<b>LIST OF FIGURES</b>	<b>ix</b>
<b>CHAPTER I INTRODUCTION</b>	<b>1</b>
1.1 Background and Statement of problems	1
1.2 Objective	2
1.3 Scope of study	3
1.4 Expected Results	3
<b>CHAPTER II LITERATURE REVIEW</b>	<b>4</b>
2.1 Customer Relationship Management (CRM)	5
2.1.1 Customer value	5
2.1.2 RFM (Recency-Frequency-Monetary) model	6
2.2 Clustering Algorithms	6
2.2.1 K-Means Algorithm	6
2.2.2 EM Algorithm	7
2.3 Business Intelligence (BI)	8
2.4 Data Mining and CRM	9
2.5 BI and CRM	10
2.6 Data Mining and BI	10
2.7 Related Researches	11
<b>CHAPTER III METHODOLOGY</b>	<b>13</b>
3.1 Business understanding	14
3.2 Data Collection and Data Preprocessing	15

## **CONTENTS (cont.)**

	<b>Page</b>
3.2.1 Data Collection	15
3.2.2 Data Preprocessing	15
3.2.3 RFM Analysis	17
3.3 Customer Segmentation by Clustering Techniques	18
3.3.1 Modeling phase	18
3.3.2 Evaluation phase	18
3.4 Knowledge representation by Business Intelligence	19
3.5 CRM Strategies Development	19
3.6 Research Schedule	20
<b>CHAPTER IV RESULTS</b>	<b>21</b>
4.1 Customer Segmentation by Clustering Techniques	21
4.2 Knowledge Representation by Business Intelligence	23
4.2.1 RFM Dimension	23
4.2.2 Geographic Dimension	24
4.2.3 Incoterms Dimension	25
4.2.4 Application Dimension	26
4.2.5 Packaging Dimension	26
4.2.6 Product Categories Dimension	27
4.3 Building up CRM Strategies	28
<b>CHAPTER V CONCLUSION AND DISCUSSION</b>	<b>32</b>
5.1 Conclusion and Discussion	32
5.2 Future works	33
<b>REFERENCES</b>	<b>34</b>
<b>APPENDIX</b>	<b>36</b>
<b>BIOGRAPHY</b>	<b>52</b>

## LIST OF TABLES

<b>Table</b>	<b>Page</b>
3.1 Order quantity preprocessing	16
3.2 Payment terms preprocessing	16
3.3 Degree of product varieties preprocessing	16
3.4 Recency scaling	17
3.5 Frequency scaling	17
3.6 Monetary scaling	18
4.1 Clustering Results by K-Means	21
4.2 Clustering Results by EM	22
4.3 Customer Segment Matrix	28
4.4 CRM Strategies Matrix	30

## LIST OF FIGURES

<b>Figure</b>	<b>Page</b>
2.1 Integration between CRM, BI and Data Mining	4
2.2 K-Means Algorithm	7
2.3 Demo shows the EM algorithm used for estimation of parameters of the Gaussian mixture model	8
2.4 Data visualized by Business Intelligence	9
3.1 Research Methodology	13
3.2 Customer Segmentation classified by type of contract (AS-IS)	15
3.3 Research Schedule	20
4.1 Dimension by RFM Analysis	24
4.2 Dimension by Geographic	25
4.3 Dimension by Incoterms	25
4.4 Dimension by Application	26
4.5 Dimension by Packaging Type	27
4.6 Dimension by Product Categories	27
4.7 Customer Segmentation by K-Means Clustering	29
4.8 CRM Strategies in each Customer Segmentation	29

# CHAPTER I

## INTRODUCTION

### 1.1 Background and Statement of Problems

Optimizing customer service and customer satisfaction is in the essence of customer relationship management [1]. Among the intense competition of business at present, if any organization can precisely monitor and measure the customer service factors and customer satisfaction, it is easier to make appropriate tactic to ensure customer loyalty, customer retention, creating brand awareness, strengthening good relationship with existing customers, enhancing good reputation references, and new customer acquisition. Understanding the individual customer needs becomes the key dimension in Customer Relationship Management (CRM) on which organizations drive their sustainable competitive advantage. To effectively manage the customer relationships, the organization must set the priorities to serve the key profitable customers and to understand both buying behavior and customer satisfaction for various market segments, customer profiles, products and services [2]. With this intelligent information, organization could better set up strategies that meet customer expectations and provide the satisfied excellent services to succeed the established objectives and business needs.

Customer segmentation is the fundamental technique used in marketing and CRM. Customer-purchase transaction data is deployed to track the buying behavior and to create the strategic business initiatives. Organizations can use this data to divide customers into segments based on variables, such as current customer profitability, a measure of the lifetime value of a customer, and retention probability. Traditionally, most organizations identify customer segment by using a mathematical model without considering the correlation between customer cluster and a campaign or loyalty programs [3]. As a result, the useful information is often overlooked. Data mining [4] can extract the knowledge and provide the customer insight through the analysis of data using sophisticated modeling techniques. In recent years, it has been

recognized that the clustering technique [4], one of data mining tasks, mostly used to discover a new set of categories, is well suited methodology for customer segmentation due to their relatively low computational requirements but able to mine hidden knowledge in large dataset.

Customer segmentation by clustering techniques is one of the effective methods for customer categorization. However, only straightly retrieved knowledge from clustering techniques are not well interpreted enough for decision-making supported view. Therefore, the powerful presentation tools like Business Intelligence. (BI) was used to vividly visualize the customer segment in various aspects. BI [5] is a concept of using information technology as an analytical tool to interpret sharply insight in successful strategic plans making for organization by turning the data into actionable information.

Nowadays, both clustering techniques and BI are widely used as the key technologies to analyze the customer behavior, create the valuable data for up-selling, retain the customers, and strengthen the strategy in deal with CRM. Clustering techniques will be brought out the customer segmentation which is the representatives of each similar customer groups. When we integrate each customer segment with BI tools, we will find out more intelligent customer knowledge leading to establish the strategic CRM. The imperative and appropriate CRM strategy will be developed for effectively treating individual customer segment based on customer orientation in order to secure existing profitable customers, retain and deliver each customer segment through superior services, increase satisfaction, prevent customer lost, and enhance long term relationship with them.

According to background described above, this study is to define and implement the CRM strategies for each customer segment discovered by data mining techniques and BI.

## **1.2 Objective**

The objective of this research is to develop CRM Strategies based on Customer Segmentation with Clustering Techniques represented by BI.

### **1.3 Scope of study**

Scope of this research is focused on sampling data based on historical Sales Data in Sales and Distribution Module of Enterprise Resources Planning (SAP) during year 2014 and customer profile data in the Chemical Industry. This work will emphasize on particular product that have a wide variety of grades and data comprising of product's complexity categories.

### **1.4 Expected Results**

1) To determine the customer segmentation rules by clustering techniques in order to reveal new facet of customer segmentation.

2) To visualize the determined customer segmentation derived by clustering techniques in various analytical aspects by Business Intelligence.

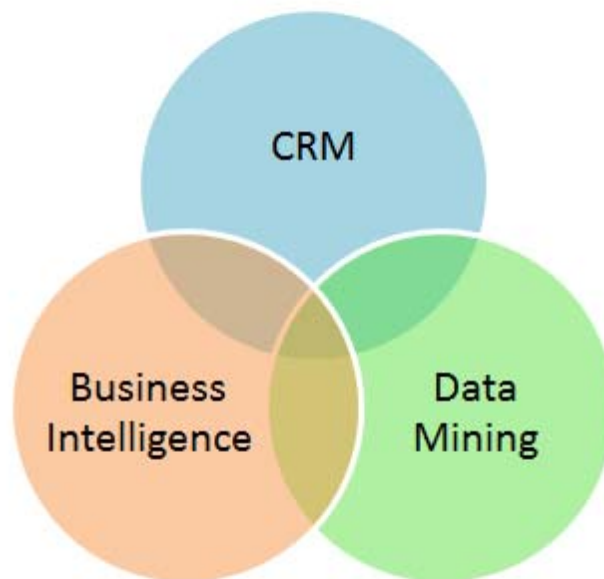
3) To align the acquired knowledge with CRM strategies in order to create the competitive advantage over competitors, expand the distribution channel from mass market to niche market, maximize the company's profit, understand the customer's buying behavior, enhance the customer's satisfaction, and strengthen the good relationship with customers in each segment.

This chapter introduces the background and statement of problems, objectives for this research, scope of studies, and expected results. In the next chapter, we discuss the literature review and research.

## CHAPTER II

### LITERATURE REVIEW

In this chapter, researcher studied about concepts, theories and applications of data mining and BI in customer relationship management (CRM) and related works. This research was mainly focused on integration between CRM, BI and data mining, as shown in Figure 2.1.



**Figure 2.1** Integration between CRM, BI, and Data mining.

Research was initiated by studying the Customer Relationship Management (CRM) regarding to concept, benefit, significance of customer value and applying RFM model in measuring customer value. Next, clustering techniques in data mining were mentioned as the methodologies in customer segmentation by focusing two algorithms given as: K-Means and Expectation Maximization (EM). Further issues were referred to Business Intelligence (BI), the intelligent tools for knowledge representation and their major roles in making decision or strategic planning. Then

integration between three core domains, including CRM, BI and data mining, were mentioned in deal with benefit and business deployment. Finally, the related researches were studied as case studies related to the above concept and theories.

## **2.1 Customer Relationship Management (CRM)**

Customer Relationship Management [6] is defined by four elements of a simple framework: Know, Target, Sell, and Services. The key concept of CRM is to understand the individual customer in order to improve the better services. CRM drives the organization to enhance customer loyalty, customer retention and organization's profitability. CRM is not only process or technology, but also the strategies that organization must implement in order to gain competitive advantage. CRM could enable the organization to:

- Identify the types of customers,
- Serve each customer as an individual and personalization,
- Provide the individual customer marketing campaigns, and
- Gain the customer insight and understand customer behaviors
- Before defining CRM strategies, organization firstly to know their

customer value which be considered as the significant process to find out customer's priority.

### **2.1.1. Customer value**

Customer value is fundamental and initial step of CRM to understand and measure the true value customers. Customer value is the difference between the benefits that organization obtained from customer and the costs incurred in attracting and serving customers. The organization could sort out the most valuable customers through RFM (Recency, Frequency, and Monetary) value. Insight through RFM analysis indexes could refine the customer value that assists the organization to acquire the potential customers and the existing customers which are the most capable and willing to buy products or services.

### **2.1.2. RFM (Recency-Frequency-Monetary) model**

The RFM analytic model [7], proposed by Hughes, was a model that categorized the significant customers by three attributes which be described as follows:

1) Recency of the last purchase (R)

R is represented the recency, which refers to how recency that customers purchase the products or services. The shorter between the latest consuming period and present is the highest score of R.

2) Frequency of the purchases (F)

F is represented frequency, which refers to the number of transactions in a particular period. The most frequency was the highest score of F.

3) Monetary value of the purchases (M)

M is represented monetary, which refers to amount of the consumption money that customers spent on each purchase in a particular period. The much purchase amount is the highest score of M.

RFM is generic model assigning the value-scores to each customer on their prior purchasing behavior basis and is usually ranged by quintile system.

## **2.2 Clustering Algorithms**

Clustering is unsupervised learning that is mostly used for non-predefined target classes by finding the similarities between data according to the characteristics found in the data and grouping the similar data objects into clusters.

### **2.2.1 K-Means Algorithm**

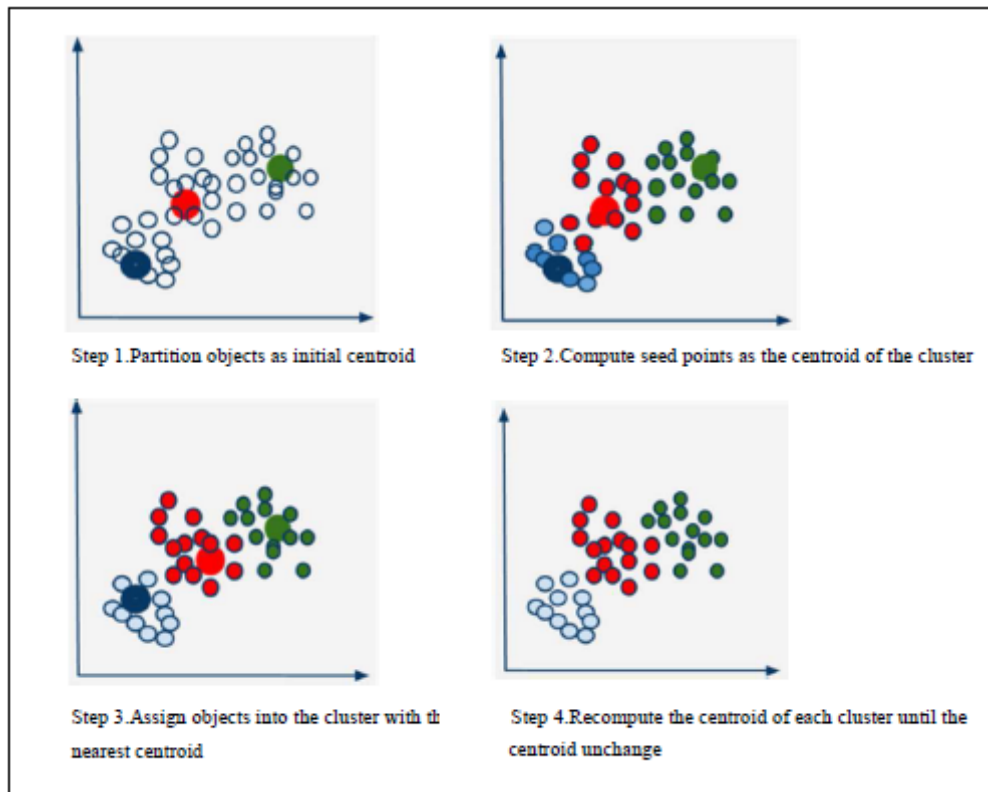
K-means [8] is one of the simplest and the most widely used of clustering algorithms. In Figure 2.2, K-means algorithm is implemented in four steps as follows:

1) Partition the objects or K as initial centroid,

2) Compute the seed points as the centroid of cluster,

3) Assign the objects into cluster with the nearest centroid,

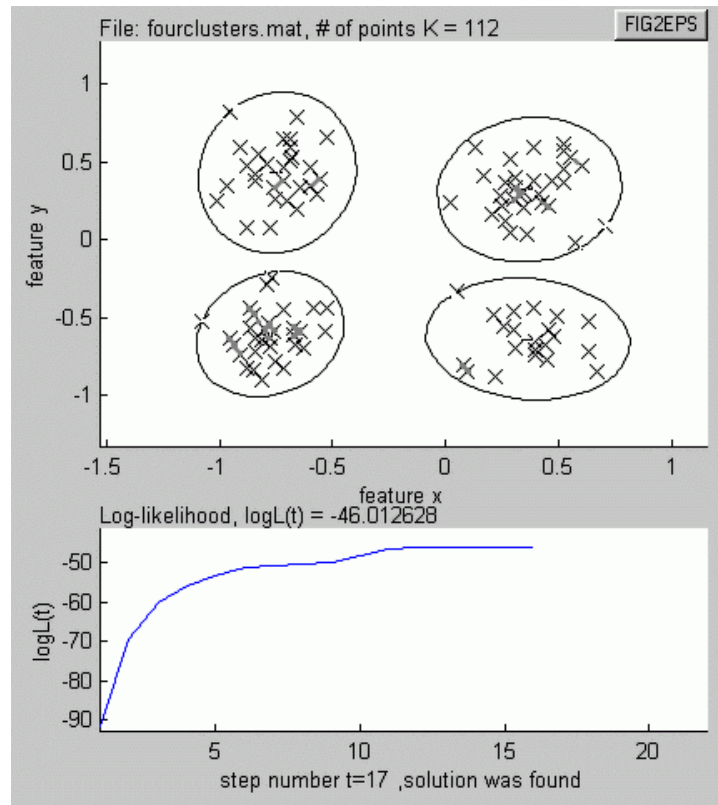
4) Recompute the centroid of each cluster until the centroid does not change.



**Figure 2.2** K-Means Algorithm [9].

### 2.2.2 EM Algorithm

Simple EM (expectation maximization) [10] assigns a probability distribution to each instance which indicates the probability of it belonging to each of the clusters. EM is a general method of finding the maximum-likelihood estimate of the parameters of an underlying distribution from a given data set when the data is incomplete or has missing values. There are two main applications of the EM algorithm. The first occurs when the data indeed has missing values, due to problems with limitations of the observation process. The second occurs when optimizing the likelihood function is analytically intractable but when the likelihood function can be simplified by assuming the existence of values for additional but missing (or hidden) parameters.



**Figure 2.3** The EM algorithm used for estimation of parameters of the Gaussian mixture model [11].

### 2.3 Business Intelligence (BI)

Business Intelligence (BI) [5] is an analytical IT application that helps organizations make sharp decisions by using technology for transforming the raw data into meaningful information and reporting. Effective BI applications [5] could shortcut the process of making business decisions due to the quality and centralized nature of data that resulted in useful and high-quality knowledge. Furthermore, BI could improve all business processes by integration of information and knowledge about customers, suppliers, competition, and overall operations. BI also improved the strategic and operational planning, control, business indicators, analysis and optimization of business operations, ensuring the successful targeting of customers, tracking the competitive activity, and predicting the future trends. An example of customer segmentation displayed in BI as shown in Figure 2.4.

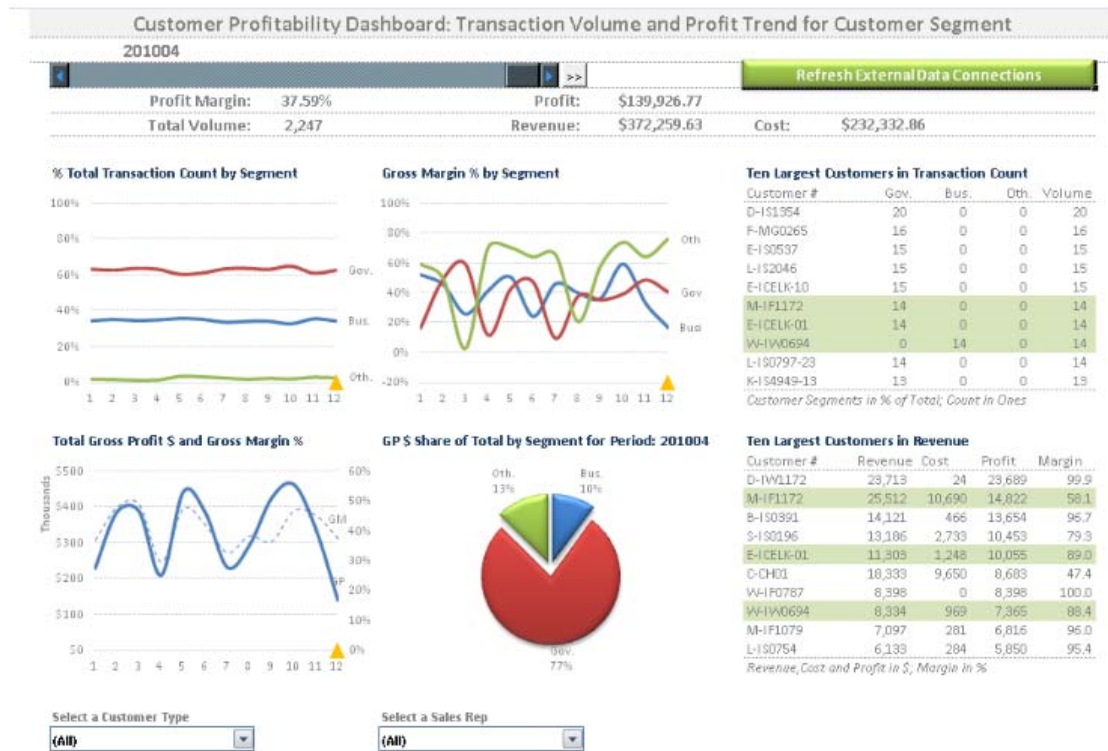


Figure 2.4 Data visualized by Business Intelligence [12].

## 2.4 Data Mining and CRM

Data mining presently takes more important role for every enterprise. Since it was regarded as a tools to reveal insight from data, thus it is regularly used by CRM .The organization rapidly and effectively extracts the complex customer data and hidden information from large databases by data mining. It could lead to identify valuable customers, gain customer insight, approach or interact with them by proper marketing tactics and enable organization to execute proactive, knowledge-driven decisions.

Data mining could provide the advantages in CRM as listed in the following [13]:

- 1) Attract potential customers, retain existing customers and increase market share,
- 2) Improve business efficiency in the least budget,

- 3) Increase customer satisfaction, customer loyalty and customer value contribution, decrease customer loss rate,
- 4) Learn customer behavior and their need to develop CRM strategy,
- 5) Evaluate the effectiveness of marketing campaign including advertisement and promotion,
- 6) Improve brand awareness,
- 7) Respond to the expectation of customer and strengthen service qualities.

## **2.5 BI and CRM**

BI was recognized as a way of delivery the right information in the right format, at the right hand and the right time. By these qualities knowledge could improve the quality of products, promotional sales, increase profits and especially enhance customer satisfaction results. Therefore, the most intriguing advantage of BI in CRM [5] was providing the organization to better understand and respond to the needs of each customer. This was to ensure each customer got precisely what they want and when they want. The connection between BI applications and CRM was frequently known as analytical CRM given the related customer data analysis for management support.

## **2.6 Data Mining and BI**

Data Mining is a powerful technology helping the organization in revealing the hidden significant information, predicting the future trends and behaviors, and making knowledge-driven decisions. Concomitantly, the most organization was realized the important role business intelligence plays in the success of their business. The organization could deploy its own data and information to gain insight, make smarter decisions and competitive advantages [14]. As business intelligence capabilities move to center stage as a top management priority, the organization meet their business intelligence needs not only in the short-term but in the long-term sustainable growth for more complex business intelligence capabilities.

## 2.7 Related Researches

Chih-Fong Tsai [15] applied two clustering techniques; K-Means and the expectation maximization algorithms (EM) in a real case study of customer segmentation for automobile dealership in Taiwan. Both clustering algorithms were adopted to generate four different customer groups: loyal, potential, VIP, and churn. This study proposed customer retention marketing strategies to increase customer satisfaction for each customer group, and improve the quality of services for effective CRM. This research was given detail of experimental results, including across-analysis results between K-Means and EM, applied customer segmentation with the real case study of enterprise in Taiwan. Nevertheless, the experimental results were presented in tabular in term of technical results which may not better visualized knowledge representation and comprehensive decision making.

Vasile [16] proposed that data mining techniques became crucial tools in CRM. Since current markets have been changed from the product-centric orientation to the customer-centric orientation, CRM strategies were developed to attract new customers, retain and restore customer who left to another competitors. Data mining methods such as descriptive (clustering) and predictive methods were utilized to achieve CRM's three major goals, given as:

- 1) Market Segmentation
- 2) Setting customer profiling
- 3) Product positioning preferences

This research proposed various data mining techniques that were used in global organizations. Website of [www.amazon.com](http://www.amazon.com) used the recommendation system to recommend the products to other customer which cause to increase company's revenue by cross-selling and up-selling. E-commerce businesses discovered knowledge by determining the target groups, approached them by the proper advertisement, and developed the effective marketing strategies. However this work has just generally mentioned to data mining techniques, but has not applied to real case study and has not given any experimental results.

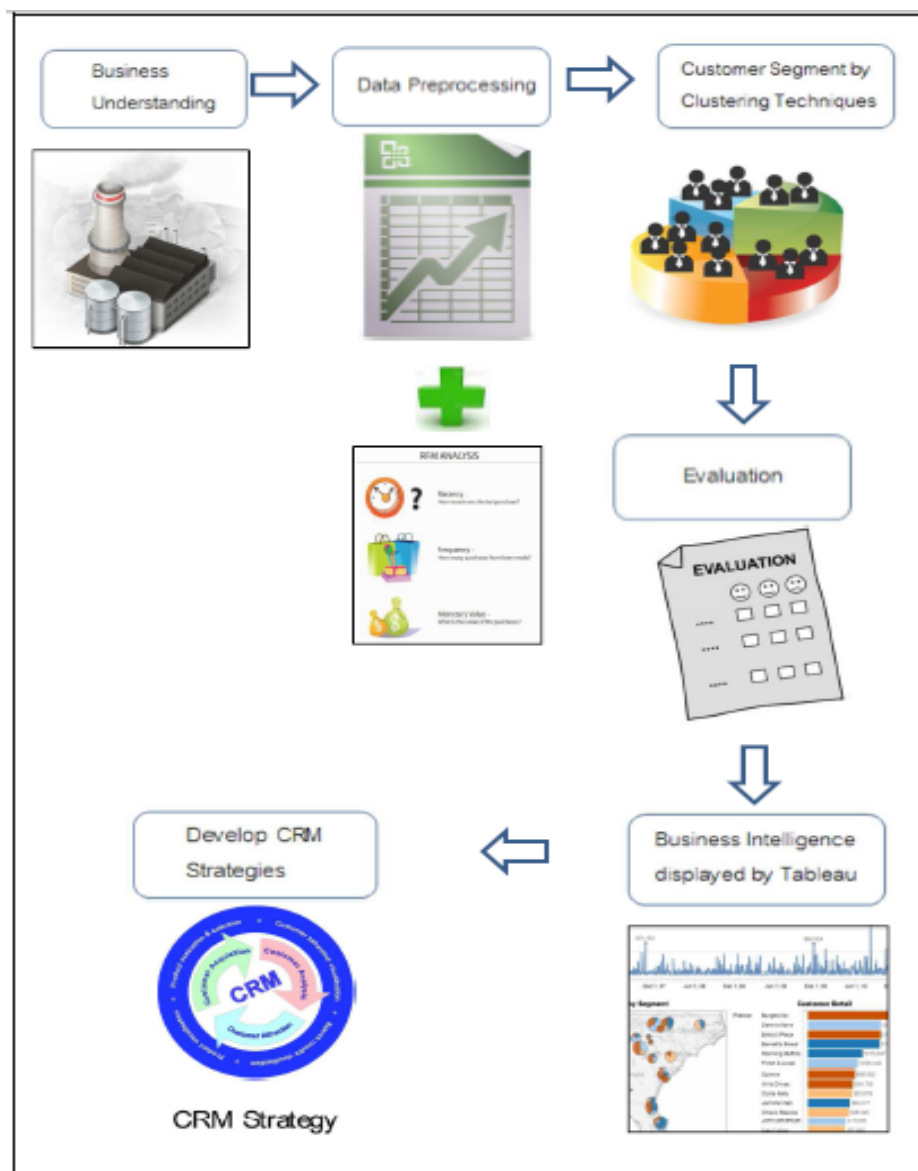
Gan Shan [17] researched on BI application in Electric Grid Industry based on data mining. The organization swiftly gained knowledge from comprehensive report and visualized dashboard from BI tools that provide the effective prompt decision process for management and optimize the enterprise resource distribution and business strategies. This works observed requirement from different business user groups and focused on a comprehensive business intelligence solution that turns the business data into information and knowledge, and help the electric grid enterprise by improving the office management, project management, technologies monitoring management, equipment inspection management, and equipment management processes. The BI platform provides the valuable support for the decision analysis. This research implemented BI system architecture for Electric grid enterprise consisting of data warehouse, online analytical processing (OLAP), and data mining technology, to support intelligent decision process of enterprises. Anyway this work did not mentioned in details of data mining tasks that adopted in their system architecture.

Based on the review of literatures, in this research, we have applied the concept of BI and clustering technique in data mining to partition the customers into segments, keenly visualize the customer segment in various business views by BI tools, and finally complies the unveiled customer data in each segment with CRM strategies.

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

## CHAPTER III METHODOLOGY

Research methodology was sequentially performed as shown in Figure 3.1.



**Figure 3.1** Research Methodology.

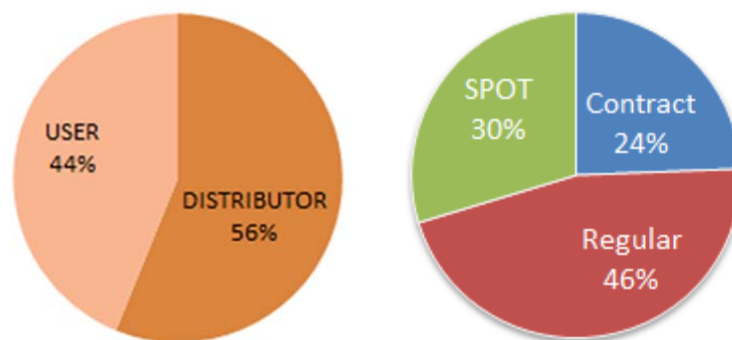
Research was initially performed by understanding business in order to clearly define the problem or requirement which can be fulfilled business need. Then, raw transactional data was accumulated and preprocessed into tabular format. RFM Analysis was applied in this stage to transform customer's data. After that, clustering techniques both K-Means algorithm and EM algorithm are selected to run model, compare the result of each method, evaluate, and choose the most appropriate one for deployment. The acquired clusters derived from clustering techniques, was represented by business intelligence tools in the next step in order to visualize data by various dimensions and providing CRM strategies. Consecutively, CRM strategies were set up to treat each customer group as per the clustering result obtained from previous process. Finally, research schedule was shown in Figure 3.3.

### **3.1 Business understanding**

At this initial stage was used for defining the objectives and requirements clearly in terms of the business or research. Then these goals were translated into the formulation of a data mining problem definition, and were prepared a preliminary strategy for achieving these objectives.

Based on business scenario of the chemical industry, the organization will unavoidably to face the intense competitiveness market by the end of year 2015 when the 10 countries that make up the Association of Southeast Asian Nations (ASEAN) are supposed to join forces as a single economic community, the Asian Economic Community (AEC). Neighboring competitors in South East Asian, such as Indonesia and Malaysia, initially playing the major roles along supply chain from upstream to downstream, will expand their market sharing. While the production cost and feedstock cost also continually increase. In fact that the organizations particularly seeking new business opportunities for survival in the market. Meanwhile organization has to secure business and existing customers as long as possible. Customer segmentation has been adopted to divide customers into homogenous and distinct groups in order to develop differentiated CRM strategies. Strategic CRM will be deployed as a key customer management tools in order to maintain, optimize company's profit, and gain competitive advantage beyond other competitors.

Referring to as-is customer groups, organization has mainly classified customers based on type of contract including spot (irregular purchasing customer), contract (committed purchase volume and fixed condition customer), regular (regular purchasing customer), and is divided by roles; user (direct end-user) and distributor (agent) as shown in Figure 3.2. By as-is categories, customer groups was roughly segmented, organization gain inadequate information to support decision making and company's strategies. That was the reason why clustering techniques should be adopted to fulfill this gap.



**Figure 3.2** Customer Segmentation classified by type of contract (AS-IS).

## 3.2 Data Collection and Data Preprocessing

### 3.2.1 Data Collection

Sampling data was derived from two main sources.

A) Historical transactional sales data in ERP covering records during year 2014.

B) Customer profile that consists of account identification number, partner function, credit status, shipping condition.

### 3.2.2 Data Preprocessing

In this stage, raw data was selected and transformed into structured tabular format. Features selection from mass transactional sales data in ERP and customer

profile by considering variables that impact on contractual agreement in terms of business which comprise of account identification number, order date, product name, packaging types, payment terms, incoterms, destination port, destination country, region, revenue, order quantities and sales channel. Each attributes was discretized and normalized by percentile range and boolean data accordingly.

**Table 3.1.** Order quantity preprocessing.

5	5th Quintile (20% of Top Volume )
4	4th Quintile
3	3rd Quintile
2	2nd Quintile
1	1st Quintile

Order Quantity was ranked in percentile and divided in quintile by purchasing volume during year 2014, every top twenty percent of each total volume was categorized as 5,4,3,2, and 1 respectively.

**Table 3.2.** Payment terms preprocessing.

T/T	NO
LC SIGHT	NO
LC TERMS	YES
T/T TERMS	YES

Payment terms was transformed into boolean data. Customer that receives credit terms was identified as yes, whereas customer who made cash payment was identified as no.

**Table 3.3.** Degree of product varieties preprocessing.

SC + MC + LC	HIGH
DUAL CHAIN	MEDIUM
SINGLE CHAIN	LOW
BY PRODUCT	EXTRA

Chemical product grade was separated by carbon chain into three categories including short chain, mid-cut and long chain. And then grouping again by degree of product varieties in purchasing history. Customer that purchase product only one chain was specified degree of product varieties purchasing as low, customer who purchase mixing two chain was specified as medium, and customer purchasing all of chains was specified high. Additionally, customer having capability to purchase special product was also specified as extra.

### 3.2.3 RFM Analysis

RFM analysis is applied by defining the scaling of R, F, and M attributes as following details.

**Table 3.4.** Recency scaling.

**Recency Scaling** (period since last purchase)

5	within 1 month (Dec)
4	within 2-3 months (Nov-Oct)
3	within 4-6 months (Sep-Jul)
2	within 7-9 months (Jun-Apr)
1	within 10-12 months (Mar-Jan)

Recency was scaled by last purchasing period of each customer during year 2014. Scale was defined as 1-5 and ranked by quarter transaction in 1-3, and the last quarter was divided into two levels as 4-5 due to the peak season of year. Customer that purchase the most recently in December was scaled as 5, purchase during October-November was scaled as 4, product purchasing during July-September was scaled as 3, purchase during April-June was scaled as 2 and purchase the least recently in January –March was scaled as 1, respectively.

**Table 3.5** Frequency scaling.

**Frequency Scaling**

5	purchase over 200 times per year
4	purchase 100-200 times per year
3	purchase 51-100 times per year
2	purchase 21-50 times per year
1	purchase below 20 times per year

Frequency was ranked by purchasing times of each customer during year 2014. Scale was defined as 1-5 and ranked by based on the minimum and maximum of purchasing times. Customer that purchase product more than 200 times per year was scaled as 5, purchasing between 100-200 times per year was scaled as 4, purchasing between 51-100 times per year was scaled as 3, purchasing between 21-50 times per year was scaled as 2 and customer that purchase product less than 20 times per year was scaled as 1, respectively.

**Table 3.6** Monetary scaling.

<b>Monetary Scaling</b>	
5	5th Quintile (20% of Top Revenue )
4	4th Quintile
3	3rd Quintile
2	2nd Quintile
1	1st Quintile

Monetary attribute was scaled in percentile and divided in quintile by revenue during year 2014. Scale was defined as 1-5, every top twenty percent of each total revenue was categorized as 5,4,3,2, and 1 respectively.

### **3.3 Customer Segmentation by Clustering Techniques**

#### **3.3.1 Modeling phase**

This phase aims to use the K-means algorithm and EM algorithm for clustering customer segment. Preprocessed dataset was run for output modeling with nominated number of cluster: 3 clusters and 4 clusters, consequently.

#### **3.3.2 Evaluation phase**

In this phase, this research compares the result of each clustering techniques performed by K-Means and EM algorithm, and selects the most appropriate method, evaluates built up model in customer aspects, and deploys in real situation.

After comparing the experimental results, K-Means algorithm was the proper method for applying in case study: Chemical Industry.

### **3.4 Knowledge representation by Business Intelligence**

At this stage , Business Intelligence will be used as the presentation layer to display visualized analytical reporting relating to customer segment that be extracted from Data Mining in various aspects of customers. Case study reporting will be controlled by parameters that has been run directly against dataset by BI including following dimensions

- 1) Profitable segment (RFM Analysis),
- 2) Regional segment,
- 3) Product Grade,
- 4) Package,
- 5) Application,
- 6) Incoterms,

The above knowledge extracted from BI discloses the various aspects of each customer segment that vitally in establishing CRM strategies.

### **3.5 CRM Strategies Development**

This is final stage when the discovering knowledge is revealed from clustering techniques and is vividly displayed by business intelligence tools like Tableau Software. Customer knowledge visualized by BI tools was presented as a dashboard in various dimensions. Each dimension could bring out the useful information, such as customer's buying behavior, new business opportunity, and significantly to identify the customer whose be the assets that company have to secure in long terms.

CRM mainly focuses on valuable customers. Uncovered knowledge and post-processed analytical data can be used to navigate sales and marketing strategy toward the existing customers. Customer relationship management does not terminate once customers having been acquired. There are involving cross-sell campaigns, up-

sell campaigns, usage stimulation campaigns, loyalty programs, and related customer centric strategies.

It is essential to understand that strategic customer relationship management consists of multiple dimensions. The organization could not only emphasize on customer relationship management in terms of technology, but discovered knowledge or analytical intelligence data is also the key elements of successful Customer relationship management. The strategic CRM was represented by matrix in order to obviously compare each customer segment with differentiated campaign.

### 3.6 Research Schedule

Research timeline has been commenced from February and consumed 7 months in details of each topic as per Figure 3.3

Sequence	Tasks	Timeline : 2015									
		FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	
1	Business Understanding and defining problem	█									
2	Studying the related researches and methodology		█								
3	Data collection and data preprocessing			█							
4	Data Mining Tasks and model evaluation				█						
5	Knowledge representation with Business Intelligence					█	█				
6	Define strategy in Customer Relationship Management						█	█			
7	Research conclusion							█	█		
8	Documentation								█	█	

**Figure 3.3** Research Schedule.

In this chapter, we discuss about sequences of research methodology. In next chapter, we discuss analysis, results and knowledge deployment for organizations.

## CHAPTER IV

### RESULTS

#### 4.1 Customer Segmentation by Clustering Techniques

This phase aims to use the K-means algorithm and EM algorithm for customer segmentation. Preprocessed dataset was run for output modeling with nominated number of clusters as Platinum, Gold, Diamond and Silver consequently. This research compares the result of each clustering techniques performed by K-Means and EM algorithm, and selects the most appropriate method to evaluate the built up model in customer aspects and deployment in real situation.

K-Means algorithm [8] is one of the simplest and the most widely used clustering algorithms by initiating the nominated cluster, finding the centroid and assigning the objects with the nearest centroid into cluster. The result of customer segmentation by K-Means is shown in Table 7.

**Table 4.1** Clustering Results by K-Means.

K-Means	Cluster 1	Cluster 2	Cluster 3	Cluster 4
	11(8%)	40(27%)	61(42%)	34(23%)
Recency	3.5455	3.6	3.7377	4.2647
Frequency	1.6364	1.675	1.1148	3.0000
Monetary	3.0000	4.175	2.5738	4.3529
Quantity	3.5455	4.075	2.541	4.3529
Chain	EXTRA	LOW	LOW	HIGH
Packaging	DRUM	TANK	BAG	BAG
Credit	NO	YES	NO	YES
Channel	0.3636	0.975	0.9508	0.4706
Incoterm	EXW	CIF	CIF	EXW
Geographic	THI	NAS	EUR	THI
Role	DIS	USER	DIS	DIS
Application	Industrial Ingr.	Personal Care	Personal Care	Personal Care

EM (expectation maximization) [10] algorithm is to generate the cluster by assigning a probability distribution to each instance which indicates the probability of it belonging to each of the clusters. The result of customer segmentation by EM is shown in Table 4.2.

**Table 4.2** Clustering Results by EM.

EM	Cluster 1	Cluster 2	Cluster 3	Cluster 4
	32(22%)	23(16%)	50(34%)	41(28%)
Recency	3.1851	4.2197	3.6759	4.1857
Frequency	1.2299	2.7943	1.0360	2.6074
Monetary	3.8709	4.0960	2.1907	4.7993
Quantity	3.7246	4.1441	2.2252	4.8247
Chain	LOW	HIGH	LOW	HIGH
Packaging	TANK	TANK	BAG	BAG
Credit	NO	YES	NO	YES
Channel	0.9999	0.0000	0.8866	1.0000
Incoterm	CIF	EXW	FOB	FOB
Geographic	NAS	THI	EUR	NAS
Role	USER	USER	DIS	DIS
Application	Personal Care	Industrial Ingr.	Personal Care	Personal Care

After comparing the experimental results as shown in Table 4.1 and Table 4.2, K-Means algorithm was the proper method for applying the case study. By comparing the size of cluster, K-Means generate the possible symmetry and differentiates characteristic of each group more obvious than EM. Cluster generated by K-Means better defines the customer group that purchase different product category, corresponded with the real situation than cluster generated by EM such as special grade product which is small portion but still have rather high purchasing frequency and impact on company’s revenue. Moreover, cluster classified by geographic is obviously represented the different market or business’s opportunity.

Cluster 1 is represented to local customers whom possibly to off-take by-products from company and make low purchasing rate, moderate in repeated purchasing and margin. This segment should be defined as Silver class.

Cluster 2 is represented to high value customers bringing the high revenue to company. Even low frequency of purchasing and moderate in repeated purchase, but consume rather high volume in particular product categories. This segment comprises of members who are the end user or manufacturer mostly located in North Asia region which is the largest and high growth consumption market. This segment should be defined as Platinum class.

Cluster 3 is represented to regular purchase customers which is the largest segment of total customers but make the lowest contribution to company. Most of members in this cluster are distributors which are sensitive in pricing and low degree of product categories purchasing. Therefore, this segment should be defined as Gold class.

Cluster 4 is represented to very high value customers who make the highest profit to company. The returned purchasing rate is also high frequency with huge volume, buying products in various categories, and having strong financial background. This segment should be defined as Diamond class.

Although each cluster was defined as representative of customer class, but there was hidden knowledge in characteristic of particular class that can better visualized by BI in the next sequence.

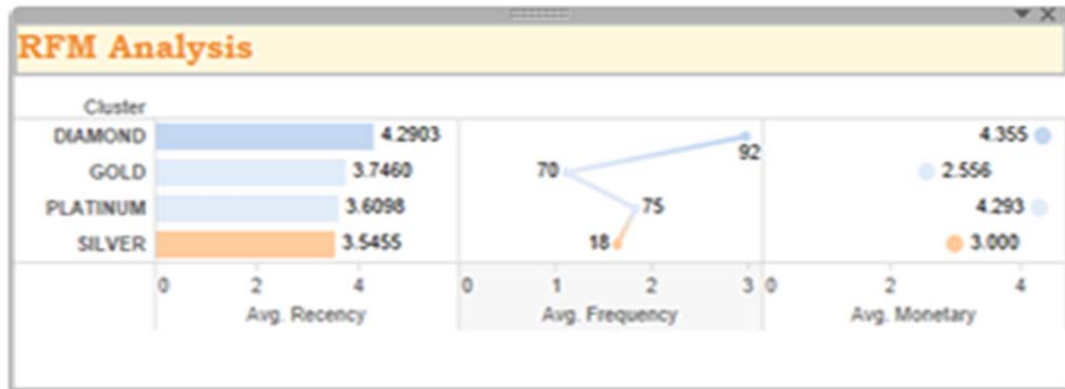
## **4.2 Knowledge Representation by Business Intelligence**

Business Intelligence (BI) represents the characteristic of each cluster in six dimensions accordingly:

### **4.2.1 RFM Dimension**

In Figure 4.1, BI is obviously shown that Diamond cluster is the most valuable customers. Customer in this segment is the asset of the company, because they have high frequency purchasing and also make the highest profit. While Platinum cluster is the second target customer that also make high contribution to company even

less frequency of purchasing than Diamond cluster. By the way, Gold cluster make regular purchasing than Platinum cluster, even though this group could not deliver high benefit to company.



**Figure 4.1** Dimension by RFM Analysis.

#### 4.2.2 Geographic Dimension

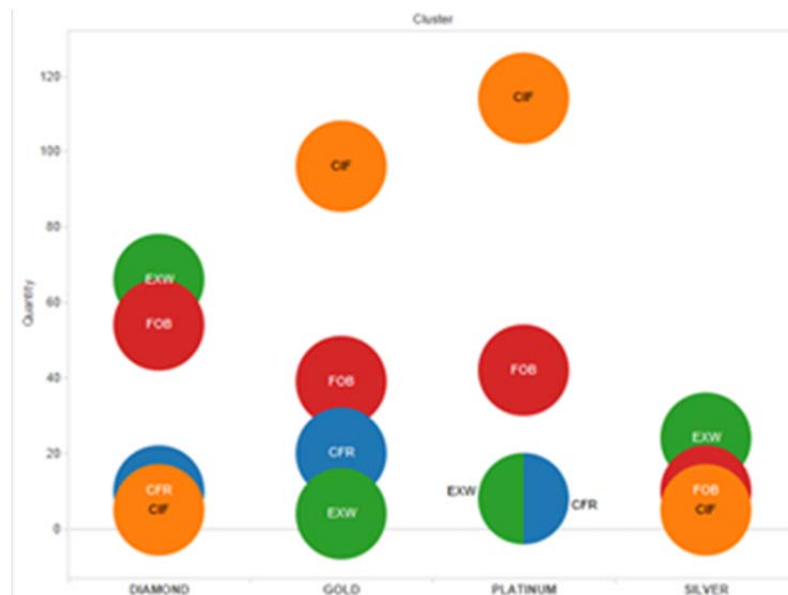
By BI visualization in Figure 4.2, high value customer in Diamond cluster that make high contribution to company are mostly located in Oceania, Africa. Gold cluster located in all regions and shipment is consigned to worldwide, but company should mainly focus on Southeast Asian area which have higher buying frequency and profit than other region. Platinum cluster is emphasized on Europe region, the largest market for this segment. North Asia is the major market of Silver cluster.



**Figure 4.2** Dimension by Geographic.

**4.2.3 Incoterms Dimension**

CIF (cost, insurance and freight) is the most frequently terms for Platinum and Gold cluster as shown in Figure 4.3. That means company may offer logistics and transportation services to both clusters. Moreover, the way company can make profit by providing all- in premium prices to customers.



**Figure 4.3** Dimension by Incoterms.

### 4.2.4 Application Dimension

As captured in Figure 4.4, Personal Care Manufacturers are the most powerful purchaser and target market that company should pay attention to strengthen relationship with. Member of both Diamond and Gold clusters are mostly in personal care industry and company could seek more opportunity in expanding market share or invest.



Figure 4.4 Dimension by Application.

### 4.2.5 Packaging Type Dimension

As shown in Figure 4.5, Platinum cluster mostly purchase product in liquid based and consigned shipment by tank. Thus packaging cost for this segment is very low and there is the room for making contribution in this area. Meanwhile Gold cluster mostly purchase products contained in bags than other packaging type. Thus company may aware and monitor concerned packaging cost for customer in this segment.



**Figure 4.5** Dimension by Packaging Type.

#### 4.2.6 Product Categories Dimension

In Figure 4.6, Gold cluster could buy some product categories, in contrast with Diamond cluster which can take all product categories. Meanwhile Silver cluster mainly purchase only special grade or by-products from company.



**Figure 4.6** Dimension by Product Categories

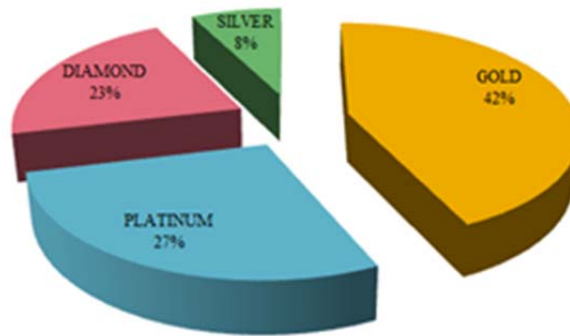
By BI visualization each dimension in BI, company could discover knowledge from customer’s buying behaviors or characteristic of each segment as shown in Table 4.3, and gain customer insight leading to develop CRM Strategies.

**Table 4.3** Customer Segment Matrix.

	<b>Diamond</b>	<b>Platinum</b>	<b>Gold</b>	<b>Silver</b>
Recency	High	Medium	Medium	Low
Frequency	High	Low	Low	Low
Monetary	High	High	Low	Medium
Quantity	High	High	Low	Low
Geographic	THI	NAS	SEA	THI
Incoterms	EXW	CIF	CIF	EXW
Application	Personal Care	Personal Care	Personal Care	Industrial Ingredient
Role	Distributor	End User	Distributor	Distributor
Packaging Type	Bag	Tank	Bag	Drum
Product Categories	High	Low	Low	Extra
	Cluster4	Cluster2	Cluster3	Cluster1

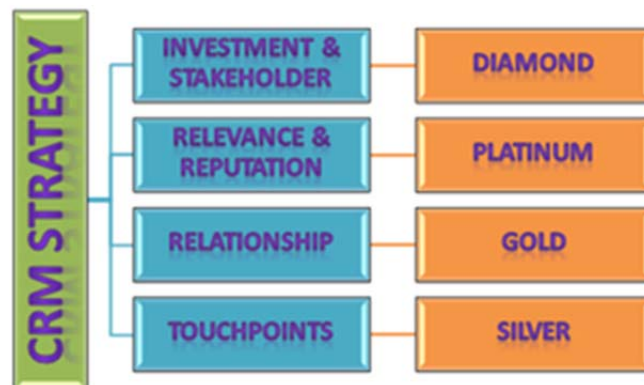
### 4.3 Building up CRM Strategies

CRM Strategies was individually set up by based on nature and different characteristic that was captured by each dimensions in BI. New customer segment is set up as; Diamond, Platinum, Gold and Silver as per break down proportion shown in Figure 4.7 that lead to building up CRM Strategies in Figure.4.8 and Table 4.4 consequently.



**Figure 4.7** Customer Segmentation by K-Means Clustering.

In Figure 4.7, new customer segment was classified by K-Means Clustering as four classes with proportion accordingly ; the largest segment is Gold (42%) , the 2<sup>nd</sup> one is Platinum (27%) , the 3<sup>rd</sup> one is Diamond (23%), and Silver (8%) is the smallest segment.



**Figure 4.8** CRM Strategies in each Customer Segmentation.

Referring to Figure 4.8, CRM Strategies was set up individually by customer segment accordingly.

- Investment and Stakeholder were used to treat customer in Diamond Class.
- Relevance and Reputation were used to treat customer in Platinum Class.
- Relationship was used to treat customer in Gold Class.
- Touchpoints was used to treat customer in Silver Class.

**Table 4.4** CRM Strategies Matrix.

	Diamond	Platinum	Gold	Silver
<b>Campaign</b>	Investment & Stakeholder	Relevance & Reputation	Relationship	Touchpoint
<b>Customer Care</b>	Superior experience	Ultimate experience	Reward	Alternative touchpoint
<b>Services Level</b>	VIP	Differentiate	Enhance	Proactive
<b>Communication</b>	Co-operative	Consultative	Commercialized	Persuasive
<b>Product</b>	Optimization	Tailor-made	Standard	Standard
<b>Pricing</b>	Joint venture	Partnership	Competitive	Attractive

Investment and Stakeholder Campaign is leaded as Diamond Strategies to use for high values or VIP customers who frequently purchase products from company with the largest volume and make the highest revenue contribution to company. All of customers in this segment are very significant and require the highest level of customer services. Company should increase level of direct attention, provide superior experience, and optimize products and services to these high value customers. Company also provides the joint venture pricing options (share in profit) with selected customer in this segment and deliver tangible award and impressive experience to these customers in order to maintain loyalty and long terms secure accounts. Meanwhile company should ensure that communication and sales channels are in co-operative rather than competitive mode.

Relevance and Reputation Campaign is leaded as Platinum Strategies to use for the new prospect customers that company could extend sales channel to expand market share and strengthen relationship with for co-creation project or becoming a long-term partnership. Company should design tailor-made products and services, co-operates in research and development for new products, and new manufacturing formula that broaden new business opportunities, and creates the values that increase mutual benefits with each other. Communication and sales channel should be consultative mode, remain consistently true to the brand, ensure core products and enhance the premium brand value. Provide the ultimate seamless purchasing experience to customers. Where appropriate, company may provide transactional consultative and partnership pricing instead of market pricing.

Relationship Campaign is leaded as Gold Strategies to use for regular purchased customer which is the largest segment in customer portal. All transaction of this group cover all regions, and products grade, but low frequency and small purchasing quantities. Customer in this segment may be sensitive with product's price than product's quality and high possibility to churn. Company should minimize cost, maintain providing standard products and offer the competitive pricing to this group. Moreover, company should extend good relationship by rewarding loyal customers in order to attract them for increasing purchased quantities, cross-sell and up-sell opportunities, and reducing churning rate. Communication and sales channel should be commercialized mode.

Touchpoint Campaign is leaded as Silver Strategies used for specific customer that occasionally purchases by-products from company. This group is very small portion compared with other segment. Although purchasing rate and consumption frequency is low, but the contribution is higher than regular customer. Company should take attention and provides the proactive strategy to deal with. Company may create alternative touchpoint scenarios for different pricing strategies or create different experiences for customers to attract them the other company's core products. Communication and sales channel should be attractive mode.

These proposed campaigns were defined as CRM Strategies to individually treat each valuable customer segment. Based on personalized characteristics of each segment, company could provide the appropriate tactics to strengthen good relationship with customers in order to create loyalty and secure them from competitors. However, this work should be fulfilled more business support if there are more input financial attributes such as pricing , transportation, delivery cost , production cost, and packaging cost. These financial view may provide more comprehensive intelligent knowledge to support decision making and contribute competitive advantage to company besides of CRM Strategies.

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

## **CHAPTER V**

### **CONCLUSION AND DISCUSSION**

#### **5.1 Conclusion and Discussion**

This research proposed CRM Strategies by applying clustering techniques and visualized BI. Raw customer data, preprocessed by RFM model approach, was transformed the daily transactional data and customer data into scale before applying clustering techniques both K-Means and EM in partitioning customer segment. The acquired result from K-Means clustering was evaluated and considered as the appropriate method for applying with case study. Because of each K-Means cluster could differentiate characteristic better than EM clustering. Based on result of clustering, customer was segmented into 4 classes; Diamond, Platinum, Gold, and Silver. When attributes in each cluster were represented and visualized by BI in RFM dimension, geographic dimension, incoterms dimension, application dimension, packaging type dimension and product categories dimension respectively, the company could acquire the knowledge to build up CRM Strategies for each customer segment.

Individual CRM Strategies were built up as four campaigns to serve each customer segment. Investment and Stakeholder Campaign was deployed as Diamond Strategies to treat VIP customer, Relevance and Reputation Campaign was deployed as Platinum Strategies to treat new prospect customers, Relationship Campaign was deployed as Gold Strategies to treat regular customers and Touchpoint Campaign was deployed as Silver Strategies to treat specific customer. Each campaign was differentiated by customer class in customer services level, the way to communicate with customer, offering products and price. Those CRM Strategies were identified and personalized by the similar characteristic of customer in each segment, and create the customer engagement to company.

Finally, it could deploy in business to strengthen and maintain good relationship with customers.

However, there are some limitation in this work relating to input data about pricing, cost comparison, and period which can be fulfilled comprehensive data for supporting CRM. Furthermore, the limited resources in the organization and the most valuable key customer segment should be the factors that be considered when CRM Strategies have been adopted to treat customers.

## **5.2 Future works**

Since this work adopted clustering techniques, BI to develop CRM Strategies and particularly applied with the enterprise in chemical industry. Therefore, the result of research may not apply with all businesses. It is subject to some specific details and nature of each type of business. Nevertheless, the acquired customer segmentation will be applied in other data mining tasks such as classification. Company could adopt the mentioned techniques in order to predict the customer's behaviors or churning rate that can be supported by CRM Strategies, Marketing campaign, predictive consumption volume, production planning and business's strategies in future.

## REFERENCES

1. M.Ghazanfari, M.Jafari and S.Rouhani, A tool to evaluate the business intelligence of enterprise systems, *Scientia Iranica, Transaction E:Industrial Engineering* 18 (6), September 2011, p1579-1590 .
2. BharatiM.Ramageri, *Data Mining Techniques and Applications*, *Indian Journal of Computer Science and Engineering*, Vol.1No.4, p301-305.
3. Konstantinos Tsiptsis and AntoniosChorianopoulos, *Data Mining Techniques in CRM*, A John Wiley and Sons, Ltd.,2009, p271-350.
4. Michael J.A.Berry and Gordon S.Linoff, *Data Mining Techniques for Marketing, Sales, and Customer Relationship Management*, Wiley Publishing,Inc., 2004, 2<sup>nd</sup> Edition, p110-113.
5. Aida Habul and AmilaPilav-Velic, *Customer Relationship Management and Business Intelligence*, [cited 2015 July 15], Available from <http://www.intechopen.com/books/advances-in-customer-relationship-management/business-intelligence-and-customer-relationship-management>.
6. Mubeena Shaik, *Data Mining Concepts with Customer Relationship Management*, *Journal of Engineering Research and Applications*, July 2014, Vol.4 Issue 7 (version 6),p98-100.
7. DeryaBriant,*Data Mining using RFM Analysis, Knowledge-Oriented in Data Mining*, In *Tech* [cited 2015 July 24], Available from <https://www.intechopen.com/books/knowledge-oriented-applications-in-data-mining/data-mining-using-rfm-analysis>.
8. LiorRokach, *Clustering Methods*, Department of Industrial Engineering. Tel-Aviv University, [cited 2015 July 20], Available from <http://www.ise.bgu.ac.il/faculty/liorr/hbchap15.pdf>
9. Kannan Chandrasekaran, *Kmeans Clustering Algorithms (Part 1)*, [cited 2015 July 20],Available from <http://humble-developer.blogspot.com/2011/01/kmeans-clustering-algorithm-part-1.html>.

10. Jeff A. Bilmes, A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models, International Computer Science Institute, Computer Science Division, Department of Electrical Engineering and Computer Science, April 1998 .
11. Vojtech Franc, Demo: Expectation-Maximization algorithm, [cited 2015 July 20], Available from [http://cmp.felk.cvut.cz/cmp/software/stprtool/examples.html#demo\\_emgmm](http://cmp.felk.cvut.cz/cmp/software/stprtool/examples.html#demo_emgmm).
12. Jon Muller, Customer Profitability Dashboard: Transaction Volume and Profit Trend for Customer Segment, [cited 2015 July 24], Available from :<http://blog.ktexcel.com/wp-content/uploads/2013/07/CustomerBySegmentProfitabilityDashboard.png#sthash.2dhWDXMM.dpbs>.
13. Samira MalekMohamadi, Application of Data Mining in Customer Relationship Management”, University of Science and Technology, 2009
14. KeyvanVahidyRodpysh, Amir Aghai and MeysamMajdi, Applying Data Mining in Customer Relationship Management, International Journal of Information Technology, Control and Automation (IJITCA) Vol.2 No.3, July 2012.
15. Chi-Fong Tsai, Ya-Han Hu and Yu-Hsin Lu, Customer segmentation issues and strategies for an automobile dealership with two clustering techniques, Expert Systems February 2015, Vol.32 no.1, p65-76.
16. Drd.Vasile, Data Mining in Customer Relationship Management, Annals of the Constantin Brancusi, University of Targu Jiu, Economy series Issue 2/2011, p161-167.
17. Gan Shan and Du Haoquan, Research on Business Intelligence Application in Electric Grid Industry based on Data Mining, 7<sup>th</sup> International Conference on Intelligent Computation Technology and Automation, 25-26 Oct 2014, p449-451.

## **APPENDIX**

# THE PROCEEDINGS OF 2<sup>nd</sup> MANAGEMENT AND INNOVATION TECHNOLOGY INTERNATIONAL CONFERENCE (MITiCON2015)

The 2nd Management Innovation Technology International Conference (MITiCON2015)

## CRM Strategies discovered by Clustering Technique and Business Intelligence; case study in Chemical Industry

Thanakal Yotsomsak and Sotarath Thammaboosadee  
Technology of Information System Management Division  
Faculty of Engineering, Mahidol University,  
Nakornpathom, Thailand  
E-mail: thanakal.y@gmail.com, sotarat.tha@mahidol.ac.th

*Abstract*— Customer Relationship Management (CRM) is the key management tools to gain sustainable competitive advantage and survive business among the intense competitive environment. Prior than planning and building up CRM strategies, the organization should prioritize customers into segment in order to properly manage relationship and provide services properly. This research aims to develop CRM strategies in case study; Chemical Industry by applying clustering techniques to segment customer, use Business Intelligence (BI) as the visualized tools to represent knowledge in various business dimensions and finally bring out CRM strategies to deploy in business. According to the customer data used in customer clustering, the RFM model was applied and added to the existing attributes. The clustering methods applied in this research are K-Means and EM algorithm. The clusters are generated and defined as 4 classes: Diamond, Platinum, Gold and Silver respectively. Each cluster is different in characteristic, but still not well interpreted enough to making decision without powerful presentation tools like BI. Finally, knowledge acquired from dimensional aspects in BI could lead to CRM strategies. The Organization could strengthen good relationship with individual customer and appropriately provide personalized products and services by based on customer segment.

*Keywords*— Customer Relationship Management, Clustering Techniques, Business Intelligence, CRM Strategies, RFM Model

## I. INTRODUCTION

Optimizing customer service and customer satisfaction is the essence of Customer Relationship Management (CRM) [1]. Among the intense competition of business at present, if any organization can precisely monitor and measure customer service factors and customer satisfaction, it would be more advantage to make appropriate tactic to ensure customer loyalty, customer retention, create brand-awareness, strengthen good relationship with existing customers, enhance good reputation references, and new customer acquisition. To effectively manage customer relationships, the organization must also set priorities to serve key profitable customers and to understand buying behavior and customer satisfaction for various market segments, customer profiles, products and services. With this intelligent information, organization could better set up strategies those meet customer expectations and provide satisfied excellent services to succeed established objectives and business needs.

Customer segmentation is the fundamental technique used in CRM. Organizations can use this data to divide customers into segments based on variables. In recent years, it has been recognized that the clustering technique [2], one of data mining tasks that mostly used to discover a new set of categories, is well suited methodology for customer segmentation due to their relatively low computational requirements but able to mine hidden knowledge in large dataset.

However only straightly retrieved knowledge of clustering techniques may not be well interpreted enough for comprehensive decision-making. Therefore, the powerful presentation tools like Business Intelligence (BI) [3] used to vividly visualize customer segment in various analytic dimensions.

Nowadays both clustering techniques and BI are widely used as the key technologies to analyze customer behavior, to create valuable data for up-selling, to retain customers and to strengthen strategy in deal with CRM. Clustering techniques will be brought out the customer segmentation which is representatives of each similar customer groups. When we integrate each customer segment with BI tools, we will find out more intelligent customer knowledge lead to establishing strategic CRM. The imperative and appropriate CRM strategy will be developed for effectively treating individual customer segment by based on customer-oriented in order to secure existing profitable customers, retain and deliver each customer segment through superior services, increase satisfaction, prevent customer lost, and enhance long term relationship with them.

According to background described above, this study is to define and implement the CRM strategies for each customer segment discovered by clustering techniques and BI. In order to paper organization, the next section we will present the related works. The section 3 is the methodology and experimental results of customer segmentation by clustering methods, dimensional visualization by BI, and the final CRM strategies.

## II. RELATED WORKS

Approaching to an application of clustering and customer segmentation, Chih-Fong Tsai [4] applied two clustering

23

techniques; K-Means and the Expectation Maximization algorithms (EM) in a real case study of customer segmentation for automobile dealership in Taiwan. Both clustering algorithms were adopted to generate four different customer groups: Loyal, Potential, VIP and Churn. This study proposed customer retention marketing strategies to increase customer satisfaction for each customer group and improve the quality of services for effective CRM. This research was given detail of experimental results including across-analysis results in between K-Means and EM, compare their results for correctness, applied customer segmentation with the real case study of enterprise in Taiwan. Nevertheless the experimental results were presented in tabular in term of technical results which may not better visualized knowledge representation and comprehensive decision making.

Vasile [5] proposed that data mining techniques became crucial tools in CRM. CRM strategies were developed to attract new customers, retain and restore customer who left to another competitors. Data mining methods such as descriptive (clustering) and predictive were utilized to achieve CRM's three major goals; Market Segmentation, Setting customer profiling and Product positioning preferences. This research proposed various data mining techniques that were used in global organizations such as to recommend products to other customer which cause to increase company's revenue by cross-selling and up-selling, to determine the target groups and approached them by the proper advertisement and developed effective marketing strategies. However this work has just generally mentioned to data mining techniques, but has not applied to real case study and has not given any experimental results.

According to an application of BI in CRM, Gan Shan [6] researched on BI application in Electric Grid Industry based on data mining. The organization swiftly gained knowledge from comprehensive report and visualized dashboard from BI tools that provide the effective and prompt decision process for management and optimize the enterprise resource distribution and business strategies. This works observed requirement from different business user groups and focused on a comprehensive business intelligence solution that turns the business data into information and knowledge, and help the electric grid enterprise by improving the office management, project management, technologies monitoring management, equipment inspection management, and equipment management processes. The BI platform provides valuable support for the decision analysis. This research implemented BI system architecture for Electric grid enterprise that consisted of data warehouse, online analytical processing (OLAP) and data mining technology to support intelligent decision process of enterprises. Anyway this work did not mentioned in details of data mining tasks that adopted in their system architecture.

Based on the review of literatures; in this research, we have applied the concept of BI and clustering technique in data mining to partition customers into segments, keenly visualize customer segment in various business views by BI tools and finally to comply the unveiled customer data in each segment with campaigns and gain customer insight and understand customer behaviors.

### III. RESEARCH METHODOLOGY AND RESULTS

This research was initially performed by understanding business in order to clearly define problem or requirement which can be fulfilled business need. Then raw transactional data was accumulated and preprocessed into tabular format. The RFM (Recency-Frequency-Monetary) Analysis [7] was applied in this stage to transform customer's data. Consequently, clustering techniques; both K-Means algorithm and EM algorithm were selected to create a model and compare the result of each method prior than evaluation and choose the most appropriate one for deployment. The acquired clusters derived from clustering techniques were represented by BI tools in the next step in order to visualize data by various dimensions and provide CRM strategies. Consecutively, CRM strategies were set up to treat each customer group as per the clustering result obtained from previous process.

*A. Data Preprocessing*

In this research, raw data in the Chemical Industry was selected as a case study and was transformed into structured tabular format. Features selection from mass transactional sales data and customer profile by considering variables that impact on contractual agreement in terms of business which comprise of account identification number, order date, product name, packaging types, payment terms, incoterms, destination country, region, revenue, order quantities and sales channel. Each attributes was discretized and normalized by percentile range and boolean data. Meanwhile RFM (Recency-Frequency-Monetary) [7] was applied as generic marketing model in order to assign value-scores and categorize customers by three attributes based on their past purchasing behavior. R was represented recency, which refers to how recently that customer purchase products or services. F was represented frequency, which refers to the number of transactions in a particular period. M was represented monetary, which refers to how much the consumption money that customers spent on each purchase in a particular period. The captured of RFM scaling and an example data is illustrated in Fig.1 and Fig.2 respectively.

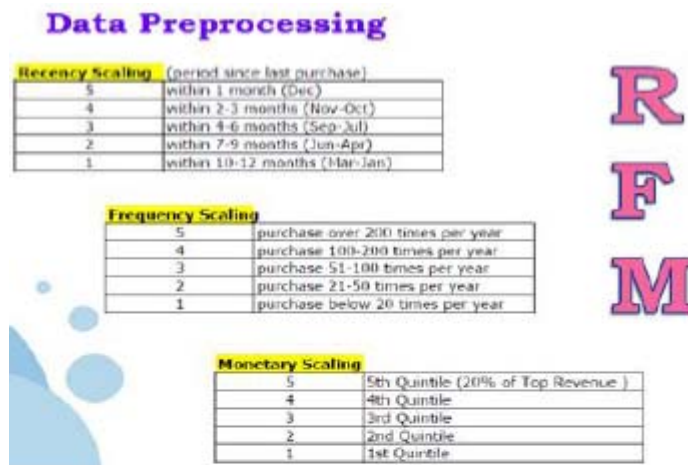


Fig. 1. Example Preprocessing Data using RFM Model.

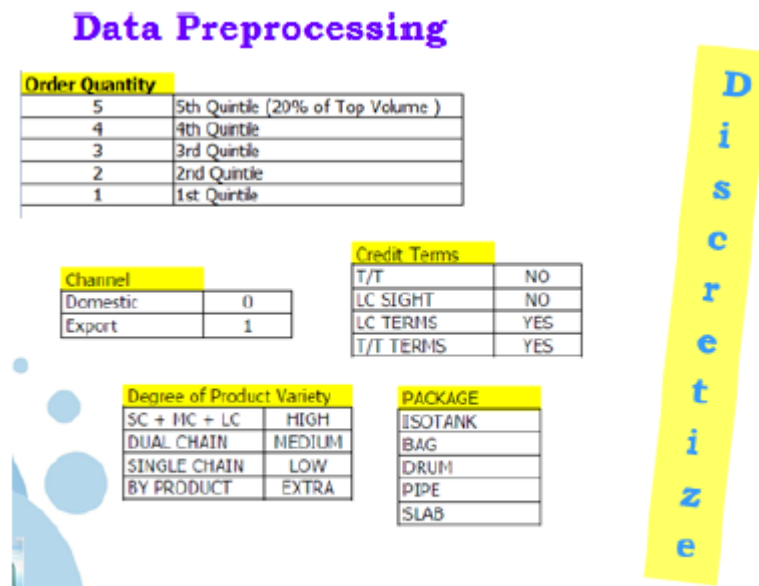


Fig. 2. Example Preprocessing Data.

#### B. Customer Segmentation by Clustering Techniques

This phase aims to use the K-means algorithm and EM algorithm for customer segmentation. Preprocessed dataset was run for output modeling with nominated number of clusters as Platinum, Gold, Diamond and Silver consequently. this research compare the result of each clustering techniques performed by K-Means and EM algorithm and select the most appropriate method , evaluate built up model in customer aspects and deployment in real situation.

K-Means algorithm [8] is one of the simplest and the most widely use of clustering algorithms by initiate nominated cluster, find centroid and assign objects with the nearest centroid into cluster. The result of customer segmentation by K-Means is shown in Table I.

TABLE I. CLUSTERING RESULTS BY K-MEANS ALGORITHM.

K-Means	Cluster 1	Cluster 2	Cluster 3	Cluster 4
	11(8%)	40(27%)	61(42%)	34(23%)
Recency	3.55	3.60	3.74	4.26
Frequency	1.64	1.68	1.11	3.00
Monetary	3.00	4.18	2.57	4.35
Quantity	3.55	4.08	2.54	4.35
Chain	EXTRA	LOW	LOW	HIGH
Packaging	DRUM	TANK	BAG	BAG
Credit	NO	YES	NO	YES
Channel	0.36	0.98	0.95	0.47
Incoterms	EXW	CIF	CIF	EXW
Geographic	THI	NAS	EUR	THI
Role	DIS	USER	DIS	DIS
Application	Industrial Ingr.	Personal Care	Personal Care	Personal Care

EM (expectation maximization) [8] algorithm is generate cluster by assigns a probability distribution to each instance which indicates the probability of it belonging to each of the clusters. The result of customer segmentation by EM is shown in Table II.

TABLE II. CLUSTERING RESULTS BY EM ALGORITHM.

EM	Cluster 1	Cluster 2	Cluster 3	Cluster 4
	32(22%)	23(16%)	50(34%)	41(28%)
Recency	3.19	4.22	3.68	4.19
Frequency	1.23	2.79	1.04	2.61
Monetary	3.87	4.10	2.19	4.80
Quantity	3.72	4.14	2.23	4.82
Chain	LOW	HIGH	LOW	HIGH
Packaging	TANK	TANK	BAG	BAG
Credit	NO	YES	NO	YES
Channel	0.10	0.00	0.89	1.00
Incoterms	CIF	EXW	FOB	FOB
Geographic	NAS	THI	EUR	NAS
Role	USER	USER	DIS	DIS
Application	Personal Care	Industrial Ingr.	Personal Care	Personal Care

After comparing the experimental results as shown in Table I and Table II, K-Means algorithm was the proper method for applying in case study. By comparing the size of cluster, K-Means generate the possible symmetry and obviously differentiate characteristic of each group than EM. Cluster generated by K-Means better define customer group that purchase different product category, corresponded with the real situation than cluster generated by EM such as special grade product which is small portion but still have rather high purchasing frequency and impact on company's revenue. Moreover cluster that classified by geographic is obviously represent different market or business's opportunity.

*Cluster 1* is represented to local customers whom possibly to off-take by-products from company and make rather low purchasing rate, moderate in repeated purchasing and margin. This segment should be defined as Silver class.

*Cluster 2* is represented to high value customers that bring high revenue to company. Even low frequency of purchasing and moderate in repeated purchase, but consume rather high volume in particular product categories. This segment comprises of members who are the end user or manufacturer mostly located in North Asia region which is the largest and high growth consumption market. This segment should be defined as Platinum class.

*Cluster 3* is represented to regular purchase customers which is the largest segment of total customers but make the lowest contribution to company. Most of members in this cluster are distributors which are sensitive in pricing and low degree of product categories purchasing. Therefore this segment should be defined as Gold class.

25

*Cluster 4* is represented to very high value customers who make the highest profit to company. Returned purchasing rate is also high frequency with huge volume, buy products in various categories and have strong financial background. This segment should be defined as Diamond class.

### *C. Knowledge Representation by Business Intelligence*

Business Intelligence (BI) represents the characteristic of each cluster in six dimensions accordingly:

#### *1) RFM Dimension*

BI as shown in Fig.3 is obviously shown that Diamond cluster is the most valuable customers. Customer in this segment is the asset of the company because they have high frequency purchasing and also make the highest profit. While Platinum cluster is the second target customer that also make high contribution to company even less frequency of purchasing than Diamond cluster. By the way, Gold cluster make regular purchasing than Platinum cluster even though this group could not deliver high benefit to company.

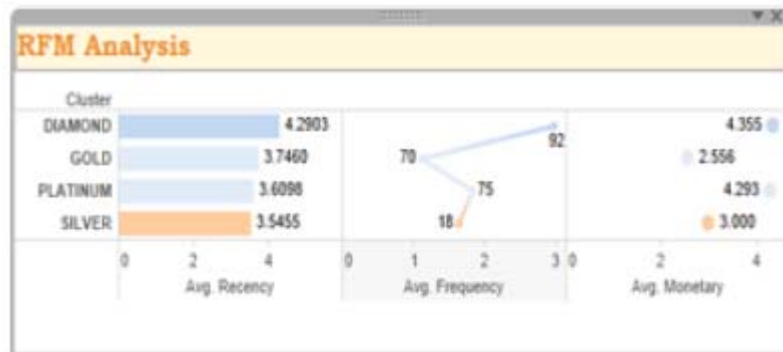


Fig. 3. Dimension by RFM Analysis.

2) *Geographic Dimension*

By BI visualization in Fig.4, high value customer in Diamond cluster that make high contribution to company are mostly located in Oceania, Africa. Gold cluster located in all regions and shipment is consigned to worldwide, but company should mainly focuses on Southeast Asian area which have higher buying frequency and profit than other region. Platinum cluster is emphasized on Europe region where is the largest market for this segment. North Asia is the major market of Silver cluster.

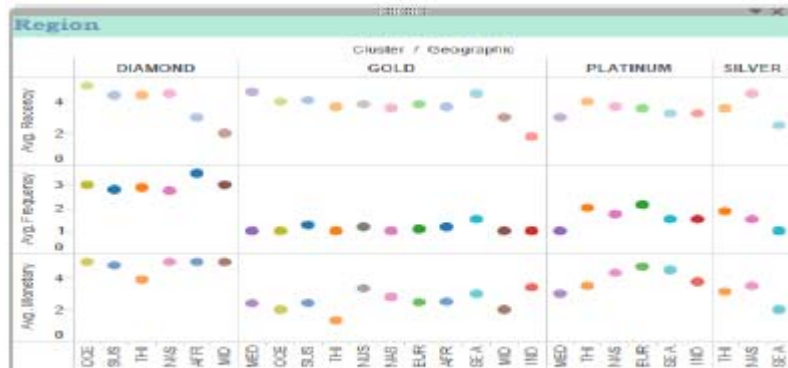


Fig. 4. Dimension by Geographic.

3) *Incoterms Dimension*

CIF (cost, insurance and freight) is the most frequently terms for Platinum and Gold cluster as shown in Fig.5. That means company may offer logistics and transportation services to both clusters. And the way company can make profit by providing all- in premium prices to customers.

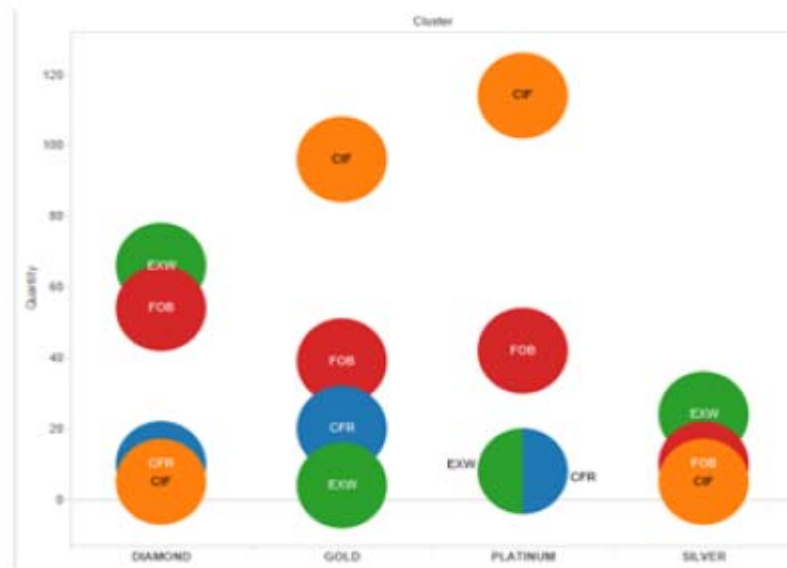


Fig. 5. Dimension by Incoterms.

4) *Application Dimension*

As captured in Fig.6, Personal Care Manufacturers are the most powerful purchaser and target market that company should pay attention to strengthen relationship with. Member of both Diamond and Gold clusters are mostly in personal care industry and company could seek more opportunity in expanding market share or invest.



Fig. 6. Dimension by Application.

5) *PackagingType Dimension*

As shown in Fig.7, Platinum cluster mostly purchase product in liquid based and consigned shipment by tank. Thus packaging cost for this segment is very low and there is the room for make contribution in this area. Meanwhile Gold cluster mostly purchase products contained in bags than other packaging type. Thus company may aware and monitor concerned packaging cost for customer in this segment.



Fig. 7. Dimension by Packaging Type.

6) *Product Categories Dimension*

As per BI shown in Fig.8, Gold cluster could buy some product categories, in contrast with Diamond cluster which can take all product categories. Meanwhile Silver cluster mainly purchase only special grade or by-products from company.

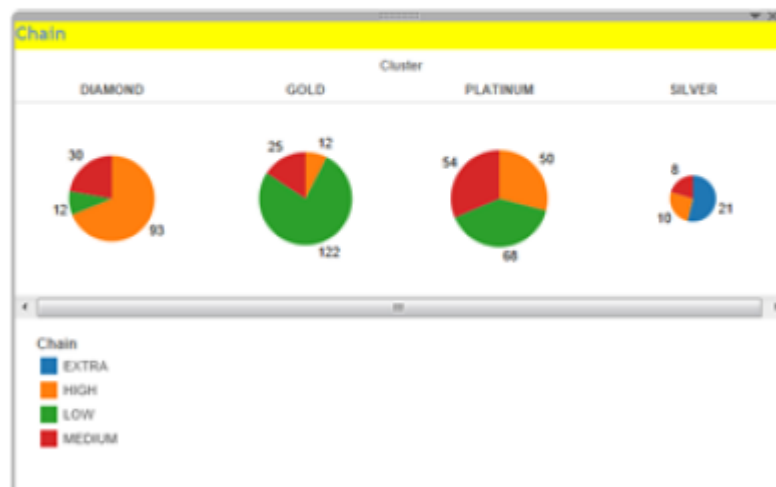


Fig. 8. Dimension by Product Categories.

By BI visualization each dimension in BI, company could discover knowledge from customer's buying behaviors or characteristic of each segment and gain customer insight that lead to develop CRM Strategies.

*D. Build up CRM Strategies*

CRM Strategies was individually set up by based on nature and different characteristic that was captured by each dimensions in BI. New customer segment is set up as; Diamond, Platinum, Gold and Silver as per break down proportion shown in Fig.9.and lead to building up CRM Strategies in Fig.10 consequently.

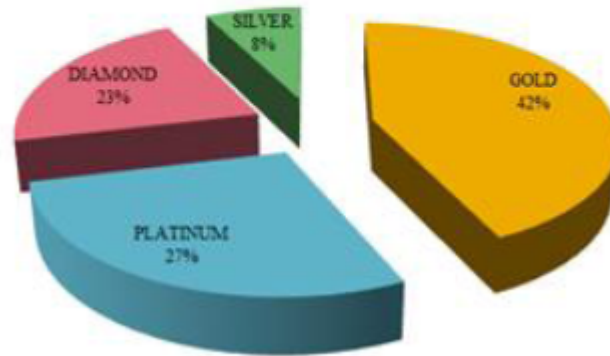


Fig. 9. Customer Segmenation by K-Means Clustering.

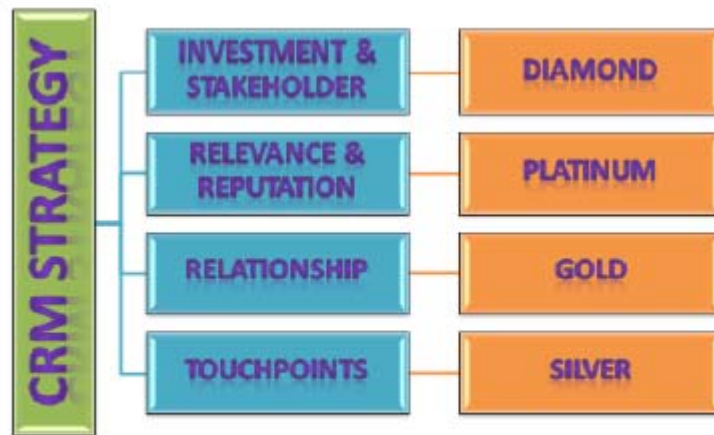


Fig. 10. CRM Strategies in each Customer Segmentation.

*Diamond Strategy:* is to define to high value or VIP customer who frequently purchase products from company with the largest volume and make the highest revenue contribution to company. All of customers in this segment are very significant and require the highest level of customer services. CRM strategies that be used for this group should be investment and stakeholders. Company should increase level of direct attention, optimize products and provide differentiated services to these high value customers. Company also provide joint venture pricing options (share in profit) with selected customer in this segment and deliver tangible award and impressive experience to these customers in order to maintain loyalty and long terms secure accounts. Meanwhile company should ensure that communication and sales channels are in co-operative rather than competitive mode.

*Platinum Strategy:* is to define to the new prospect customers that company could extend sales channel to expand market share and strengthen relationship with for co-creation project or become a long-term partnership. CRM strategies that be used for this segment should be relevance and reputation. Company should design tailored products and services, co-operate in research and development for new products, new manufacturing formula that broaden new business opportunities and create value that increase mutual benefits with each other. Communication and sales channel should remain consistently true to the brand, ensure core products and services enhance the premium brand value. Provide the ultimate seamless purchasing experience to customers. Where appropriate, company may provide transactional consultative and partnership pricing instead of market pricing.

*Gold Strategy:* is to define to regularly purchased customer which is the largest segment in customer portal. All transaction of this group cover all regions, all products grade but low frequency and small purchasing quantities. Customer in this segment may sensitive with product's price than product's quality and high possibility to churn. CRM strategies that be used should be relationship. Company should design new products and services that extend the relationship or increase purchasing, increase cross-sell and up-sell opportunities. Reward loyal customers and offer attractive pricing to reduce churn rate.

*Silver Strategy:* is to define to specific customer that occasionally purchases by-products from company. This group is very small portion compared with other segment but require attention and pro-active strategy to deal with .CRM strategies that be used for this segment should be touchpoints. Company should create alternative touchpoint scenarios for different pricing strategies or create different experiences for customers to attract them the company's core products.

#### IV. CONCLUSION

This paper proposes CRM Strategies by applying clustering techniques and visualized BI. Raw customer data was preprocessed by approach RFM model, transformed daily transactional data and customer data into scale prior than applying clustering techniques both K-Means and EM in partitioning customer segment.

The acquired result from K-Means clustering was evaluated and considered as the appropriate method for applying with case study. Due to each K-Means cluster could differentiate characteristic better than EM clustering. Based on result of clustering, customer was segmented into 4 classes; Diamond, Platinum, Gold and Silver. When attributes in each cluster were represented and visualized by BI in RFM dimension, geographic dimension, incoterms dimension, application dimension, packaging type dimension and product categories dimension respectively, company could acquire knowledge to build up CRM Strategies for each customer segment. Finally, it could deploy in business to strengthen and maintain good relationship with customers.

Since this work adopted clustering techniques, BI to develop CRM Strategies and particularly applied with the enterprise in chemical industry. Therefore the result of research may not apply with all businesses. It is subject to some specific details and nature of each type of business. Nevertheless acquired customer segmentation will be applied in other data mining tasks such as classification. Company could adopt mentioned techniques in order to predict customer's behaviors or churning rate that can be supported CRM Strategies, Marketing campaign and business's strategies in future.

#### REFERENCES

- [1] Angus Jenkinson and Michael Jacobs, *The IDM guide to CRM Mastery*, The Institute of Direct Marketing, July 2002, Session 5, pp.1-46.
- [2] Bharati M.Ramageri, "Data Mining Techniques and Applications" *Indian Journal of Computer Science and Engineering*, vol.1 No.4, pp. 301-305
- [3] Aida Habul and Amila Pilav-Velic, *Customer Relationship Management and Business Intelligence*, [ cited 15 July 2015], Available from <http://www.intechopen.com/books/advances-in-customer-relationship-management/business-intelligence-and-customer-relationship-management>.
- [4] Chi-Fong Tsai, Ya-Han Hu and Yu-Hsin Lu, "Customer segmentation issues and strategies for an automobile dealership with two clustering techniques" *Expert Systems*, vol. 32 no.1, February 2015, pp. 65-76.
- [5] Drd. Vasile, "Data Mining in Customer Relationship Management" *Annals of the Constantin Brancusi, University of Targu Jiu, Economy series*, Issue 2/2011 , pp. 161-167.
- [6] Gan Shan, Du Haoquan, "Research on Business Intelligence Application in Electric Grid Industry based on Data Mining", *7<sup>th</sup> International Conference on Intelligent Computation Technology and Automation*, October 2014, pp.449-451.
- [7] Derya Briant, "Data Mining Using RFM Analysis, Knowledge-Oriented Applications in Data Mining", Prof.Kimito Funatsu (Ed.), ISBN:978-953-307-154-1, InTech, [cited 12 August 2015], Available from: <https://www.intechopen.com/books/knowledge-oriented-applications-in-data-mining/data-mining-using-rfm-analysis>
- [8] Konstantinos Tsipstsis, Antonios Chorianopoulos, "Data Mining Techniques in CRM", A John Wiley and Sons, Ltd., Publication, 2009, pp. 189-220.

**BIOGRAPHY**

<b>NAME</b>	Miss Thanakal Yotsomsak
<b>DATE OF BIRTH</b>	10 June 1974
<b>PLACE OF BIRTH</b>	Bangkok, Thailand
<b>INSTITUTIONS ATTENDED</b>	Silpakorn University, 1993-1996 Bachelor of Arts (Thai) Mahidol University, 2014-2016 Master of Science (Information Technology Management)
<b>RESEARCH GRANTS</b>	Grant to Support Graduate Students in Academic Presentations in Thailand Academic Year 2015
<b>HOME ADDRESS</b>	36 Chomthong 3, Chomthong Road, Chomgthong, Bangkok, Thailand, 10150 Tel. 081-255-8500 E-mail : Thanakal.y@gmail.com
<b>PUBLICATION / PRESENTATION</b>	Yotsomsak, T., Thammaboosadee, S., (2015), CRM Strategies discovered by Clustering Techniques and Business Intelligence; case study in Chemical Industry, The Proceedings of 2 <sup>nd</sup> Management and Innovation Technology International Conference (MITiCON 2015), pp. 23-28.