Segmenting the Banking Market Strategy by Clustering

Varun Kumar. M	Vishnu Chaitanya. M	Madhavan. M
Assistant Professor (Junior)	Assistant Professor	Research Assistant
School of Information Technology & Engineering	School of Information Technology & Engineering	School of Information Technology & Engineering
VIT University-Vellore	VIT University-Vellore	VIT University-Vellore
Tamil Nadu-India	Tamil Nadu-India	Tamil Nadu-India

ABSTRACT
Customer segmentation is the practice of dividing a customer base into groups of individuals that are similar in specific ways relevant to marketing such as age, gender, interests, spending habits, and so on. One of the easiest definitions is "*a group of customers with shared needs*". From this definition, it's clear what we need to identify customers with shared needs. The customer segmentation consists of two phases. First phase includes K-Means clustering, where the customers are clustered according to their RFM (Recency Frequency Monetary). In the Second phase, with demographic data, each cluster is again partitioned into new clusters. Finally LTV (Life Time Value of the customers) are used to generate customer's profile.

Keywords

RFM   - Recency Frequency Monetary

SOM   - Self Organizing Map

LTV   - Life Time Value of a customer

MARC   - Mining Association Rule using Clustering

CRM   - Customer Relationship Management

OSI-ISO - Open Systems Interconnection – International

         Standards Organization

1. INTRODUCTION
Marketing has become the significant function in banking industry which makes bank to pay more attention to improve their marketing strategy. In order to attain a huge success over their market, they have to boost their strategy to identify, understand and target the profitable or valuable customers from other non-profitable customers. It is highly impossible for banks to serve customers on a one-to-one basis to find this. To overcome this problem, a two phase clustering method has been adopted where CRM plays major role in segmenting customers. Customer Segmentation is the primary stage of CRM. To enhance better customer service and to reduce operational costs we resolve on customer segmentation. In this two phase clustering method the first phase includes
k-means clustering, where RFM is used as the input variable to cluster the customers. Second, demographic data is used to form new clusters from the existing clusters that were derived from the first phase. The demographic parameters are selected by applying a variable selection procedure using Self Organizing Map (SOM) method that identifies relevant customer groups. Finally an LTV is used to generate customer profile that recognizes profitable customers.

The objectives are,
- To establish link between customer segmentation and marketing campaign and also to enhance by using RFM.
- The Profiles of customers in each group could be analyzed by marketers to make better strategies for each group.
- To establish better customer relationship management strategies.

The targeted customers can be identified by new segmentation method in the data mining domain and through some models of CRM like RFM, LTV and demographic variables. The method that is currently used is based on two-phase clustering model and this can be achieved through K-means technique. The existing customers are clustered into different groups of customers based on their transactional behaviors and characteristics. To set market strategies, the profile of customers in each group is analyzed by the marketers.

2. FEASIBILITY STUDY
The two phase clustering method has received an increasing attention in today's marketing and commercial areas. This is due to the fact that there is a fast access to the profitable customers by segmenting the customers into profitable and non-profitable. This method has overcome the disadvantages of spending more time and resources over the non-profitable segments which is unworthy. This method makes us to identify the targeted profitable customers, always retains the customer to our application for a longer period, utilizing the right amount of money and time over the profitable segments, customer satisfaction and maintaining good relationship with clients. Customer segmentation will add value by enabling it to: Target appropriate products to different consumers, Utilize marketing resources more efficiently, stay side by side of emerging market trends. Using segmentation, we can speak to the needs and interests of different groups, we can determine whether there is a product/service that fit in high opportunity segments, we can weigh whether product enhancements or new products might appeal to targeted groups. Segmentation often improves average retention and profitability of customers on a segment by segment basis. Implement individual marketing plans for each targeted segment.

3. LITERATURE SURVEY
"Clustering e- Banking Customer using Data Mining techniques and Marketing Segmentation" [1] discusses about data mining techniques like SOM, K-Means algorithm and Marketing technique-RFM analysis which are used to

segment customers into groups according to their personal profiles and e-banking usages. Later Apriori algorithm is used to find e-Banking services that are similar or applicable for segmented groups of customers.

"A general approach to the automatic content- based organization and visualization of large digital music collections" [2] discusses about Cross national analysis for online game marketing which focuses to target the loyal customers (domestic and foreign) and to concentrate their limited resources for them. The general methodology consists of two phases 1.MCFA Multi group Confirmatory Factor Analysis to test differences between national clustering factors and 2.SOM to develop actual clusters inside each nation.

"Intelligent Structuring and Exploration of Digital Music Collections" [3] aims at finding similar songs. For this they consider content based data like rhythm, timbre, tempo, etc. and various metadata like instrumentation, style, artist, etc. of songs to get the degrees of similarity between pieces of music. These similarities are captured to train a SOM which is projected by smoothed histograms that shows clusters with similar songs.

"Enhancing consumer behavior analysis using Data Mining Techniques" [4] presents a two stage framework of consumer behavior and key feature is a cascade involving SOM neural network to divide customers into homogeneous groups of customers and a decision tree method is simplified to identify relevant knowledge.

"Customer Behavior Analysis using CBA (Data Mining Approach)" [5] focus on customer classification and prediction of banking sector by using Naïve Bayesian classifier that accommodates the uncertainty inherent in predicting behavior of the customers. Here CRM takes customer as the center and optimizes the business process and provides way to retain profitable customers.

4. PROPOSED SYSTEM

Customer segmentation is the practice of dividing a customer base into groups of individuals that are similar in specific ways relevant to marketing such as age, gender, interests, spending habits, and so on. One of the easiest definitions is "*a group of customers with shared needs*". From this definition, it's clear what we need to do is to identify customers with shared needs. The customer segmentation consists of two phases. First phase includes K-Means clustering, where the customers are clustered according to their RFM (Recency Frequency Monetary). In the second phase, with demographic data each cluster is again partitioned into new clusters. Finally LTV (Life Time Value of the customers) are used to generate customer's profile.
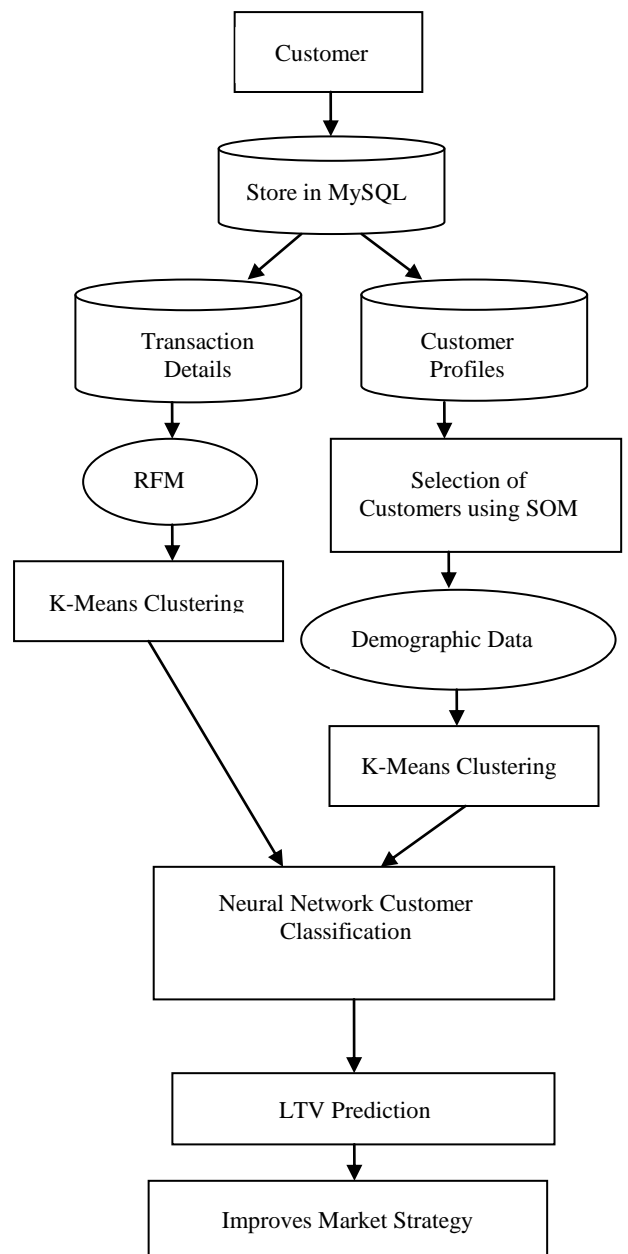
Fig 1: System Architecture

5. MODULES OF THE SYSTEM
5.1 Neural Network Customer Classification

CRM is a business strategy, the outcome of which optimizes profitability, revenue and customer satisfaction by organizing around customer segments, supporting customer satisfying behaviors and implementing customer-centric processes. CRM in Internet banking is used to improve the marketing strategy. E-banking services are offered by Fully Transactional Websites which allow the customers to operate on their accounts for transfer of funds, payment of different

bills, subscribing to other products of the bank and to transact purchase and sale of securities, etc.

5.2 K-Means Clustering Model

It generally partition n-observations into k-clusters. Clusters created based on central mean value. We consider the registered customers as n-observations using demographic data values and through RFM they are clustered (i.e.) in each cluster we consider demographic data values (student, current employee, retired people, former, senior citizen) and using RFM values (deposit, withdraw, transactions) as the central mean value, customers are grouped.

5.3 Clustering using Association Rule Mining

Here clusters consider the input variables as RFM. RFM is a method used for analyzing customer behavior and defining market segments. To create an RFM analysis, three categories for each attributes are created. Recency, Frequency and Monetary attributes can be viewed in three operations such as deposit, withdrawal and transaction.

5.4 Application Profiles for Self Organizing Map

The SOM process examines group of customer profiles and also emphasize on particular profile to perform profile integration. SOM studies user classification and customer profile and consider as data sets. These data sets are trained through unsupervised learning using existing class field thus SOM positioned arbitrarily over the data space. After performing 'n' no. of iterations nodes near trained nodes are found and thus the grid moved to the trained data. Therefore, training data will be found.

5.5 Demographic Data Collection in Life Time Value

LTV is a tactical model that is a "snapshot" of customer state at a point in time, the customer's likelihood to respond. Frequently used names for these customer states include active, lapsing, lapsed, and defected. Lifecycle is the "movie" one might put together from these snapshots of RFM states the migration from one customer state to the next are the Lifecycle trigger points.

5.6 Cluster Formation for Customer Based on Transaction

A *frequent* (used to be called large) *item set* is an item set whose support (S) is $\geq$ minSup. Apriori property (downward closure): any subsets of a frequent item set are also frequent item set.

6. ADVANTAGES OF THE PROPOSED FRAME WORK

6.1 Data Sets

Customer data is considered as data set, and assume 'n' no of customers and their transactions. During segmentation some of the factors for e-banking users are considered. They are, **RFM...**

- **R**ecency: It will let you know about the date of the recently visited user's transaction.
- **F**requency: It defines number of frequent transactions that user conduct.
- **M**onetary: The total value of financial transactions that user made within particular period.

6.2 Category of Customers

Customers are categorized into 5 groups namely,
- Students
- Current
- Employee
- Retired people
- Senior citizen.

6.3 Type of Transactions
- Withdrawal
- Deposit
- Money Transfer

6.4 Date

Customers are segmented based on the date of transactions performed (i.e. on particular date). RFM and SOM are used to perform data clustering and the information is stored into database. For clustering MARC (Mining Association Rule using Clustering) algorithm is used. The usage of this algorithm is to make a full pass over the entire database. It segments the collection of transaction so that similar transactions fall into same cluster. Then examines group of customer profiles and also emphasize on particular customer profile to perform profile integration so that trained data will be found. Then, it is given to Neural Network. It is used to repeat the transaction behavior and segment the customers of similar kind.

LTV is predicted using

$$\text{Loan to value Ratio} = \frac{\textbf{Mortgage Amount}}{\textbf{Appraised Value of the property}}$$

Managing various customer behavior clusters will be very helpful for understanding the customers. This goal is achieved by employing clustering techniques and training the samples of customer behavior variables. The proposed framework is a hybrid approach which uses clustering techniques to preprocess input samples into homogeneous clusters using K-means, and Neural network to build cluster profiles.

7. SYSTEM IMPLEMENTATION

The link between customer segmentation and marketing campaign is enhanced through using RFM. Profiles of customers in each group could be analyzed by marketers to make strategies for each group. In application of the method on our case study (in banking industry), the customers were divided into nine groups of customers according to their shared transactional behavior and characteristics. Firstly, with K-means clustering arithmetic, customers are clustered into different segments by similar survival characters (i.e. churn trend). Secondly, each cluster's Survival/hazard function is predicted by survival analyzing, then, the validity of • Profiles of customers in each group could be analyzed by marketers to make strategies for each group. Beyond simply understanding customer value in each cluster, they would gain the opportunities to establish better customer relationship management strategies, improve customer loyalty and revenue and find opportunities for up and cross selling.

7.1 Implementation Procedures

- Customer network Classification
- Clustering using association rules
- Application profiles for SOM process
- K-Means clustering model
- Demographic data collection in LTV Prediction Model
- Cluster formation for customer based on transaction

7.1.1 Customer network Classification

CRM is a business strategy, the outcomes of which optimize profitability, revenue and customer satisfaction by organizing around customer segments, supporting customer satisfying behaviors and implementing customer-centric processes. CRM is used in Internet banking to improve the marketing strategy. E-banking services are offered by Fully Transactional Websites which allow the customers to operate on their accounts for transfer of funds, payment of different bills, subscribing to other products of the bank and to transact purchase and sale of securities, etc. The above forms of Internet banking services are offered by traditional banks, as an additional method of serving the customer or by new banks, who deliver banking services primarily through Internet or other electronic delivery channels as the value added services. Some of these banks are known as 'virtual' banks or 'Internet only' banks and may not have any physical presence in a country despite offering different banking services. The term customer Network Classification means the role specification. Here user role, administrator role and the network between the customer and the administrator is specified. This allows customer to access e-banking operations by registering so they are allowed to perform deposit, withdrawal and transaction. They can also find loan amount value based on their property value. They can even change some of the details. Administrator also has their separate login so that they can view, update and also can process the customer details such as deposit, withdrawal and transaction details.

7.1.2 Clustering using association rules

RFM

      I. Recency: It let know u about the date of the recently visited user's transaction.

      II. Frequency: It defines number of frequent transactions that user conduct.

      III. Monetary: The total value of financial transactions that user made within particular period.

Type of transactions

i. Withdrawal

ii. Deposit

iii. Money Transfer.

Date

We segment customers based on date of customer transactions (i.e. from date and to date). Association rule finds co-relation between variables in larger databases. Here clusters consider the input variables as RFM. RFM is a method used for analyzing customer behavior and defining market segments. To create an RFM analysis, categories for each attribute are first created. Recency, Frequency and Monetary attributes can be viewed in three operations such as deposit, withdrawal and transaction. Administrator can view customer's RFM values by mentioning From date and To date. In the database table, the support value has been calculated (i.e.) the number of times deposit, withdrawal and transaction has been done. Here

this process is achieved only by applying one of the Data Mining Technique (i.e.) Association rule. For clustering we use MARC (Mining Association Rule using Clustering) algorithm. The usage of this algorithm is to make a full pass over the entire database. It segments the collection of transaction so that similar transactions fall into same cluster. It partition the collection of transactions so that similar transactions fall into same category, here we categorize transactions using (Recency, Frequency, Monetary) RFM then we apply association rules to classify each cluster. It has advantage of applying association rules to each cluster instead of entire data sets. It also learns association rules efficiently in single database pass.

Yields

$$\text{Support A} \longrightarrow \text{B} = \frac{\text{no of transactions A} \cap \text{no of Transactions B}}{\text{Total no of transactions}}$$

Instead of entire data set, we summarize cluster using all the frequent item sets with predetermined support and count values,

I. Based on the number of transactions the customer perform & II. By the amount value

7.1.3 Application profiles for SOM process

The SOM (Self Organizing Map) process examines group of customer profiles and also emphasize on particular profile to perform profile integration. SOM studies user classification and customer profile and consider as data sets. These data sets are trained through unsupervised learning using existing class field thus, SOM positioned arbitrarily over the data space. After, performing 'n' no. of iterations nodes near trained nodes is found and thus the grid moved to the trained data. Therefore, training data will be found.

Self-organizing maps (SOMs) are a data visualization technique which reduces the dimensions of data through the use of self-organizing neural networks. The problem is that humans cannot visualize the high dimensional data, so this technique is created to help us understand this high dimensional data. Here demographic values are taken as the training data, this categorizes customers into 5 groups i) Students ii) Current Employee iii) Retired people iv)Former v) Senior citizen.
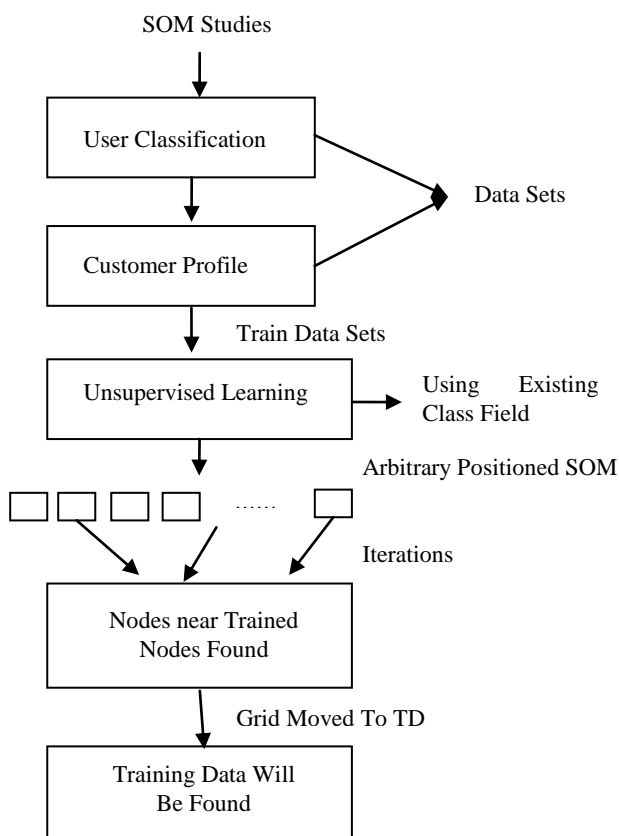
Fig 2: SOM Process

7.1.4 K-Means clustering model

K-mean algorithm creates clusters by determining a central mean for each cluster. The algorithm starts by randomly select K entities as the means of K clusters and randomly adds entities to each cluster. Then, it re-computes cluster mean and re-assigns entities to clusters to which it is most similar, based on the distance between entity and the cluster mean. Then, the mean is recomputed at each cluster, and previous entities either stay / move to a different cluster and one iteration completes. Algorithm iterates until there is no change of the means at each clusters. It generally partition n-observations into k-clusters. Clusters created based on central mean value. We consider the registered customers as n-observations using demographic data values and through RFM they are clustered (i.e.) in each cluster we consider demographic data values (student, current employee, retired people, former, senior citizen) and using RFM values (deposit, withdraw, transactions) as the central mean value, customers are grouped.
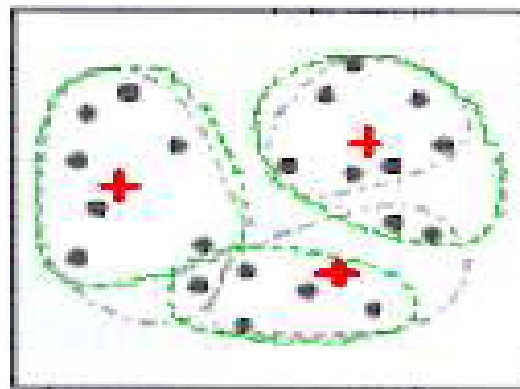

Fig 3: K- Means Cluster Formation

7.1.5 Demographic data collection in LTV Prediction Model

LTV is a tactical model that is a "snapshot" of customer state at a point in time, the customer's likelihood to respond. Frequently used names for these customer states include active, lapsing, lapsed, and defected. Lifecycle is the "movie" one might put together from these snapshots of RFM states; the migration from one customer state to the next is the Lifecycle trigger points. The "trans" is the source table for LTV prediction model. This table is subset of the churn model source table "reg". The customer joined in less than three years are filtered out from the training data set to provide a valid input into the model. Life Time Span and Life Time Value (LTV) are the two target measures to predict.
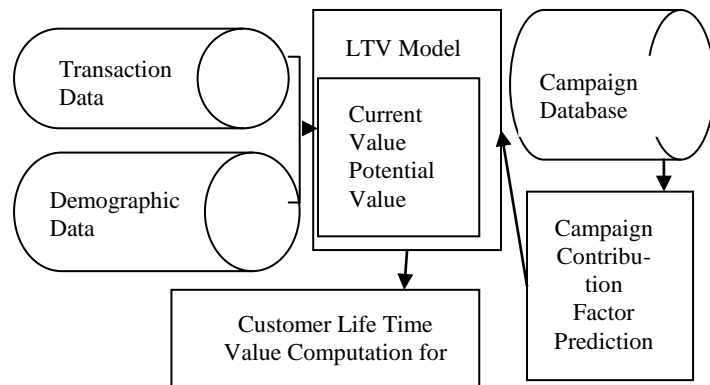

Fig 4: LTV Prediction Model

7.1.6 Cluster formation for customer based on transaction

- A *frequent* (used to be called large) *itemset* is an itemset whose support (S) is ≥ minSup.
- Apriori property (downward closure): any subsets of a frequent itemset are also frequent itemset.

Definitions:

- An *item*: an article in a basket, or an attribute-value pair
- A *transaction*: items purchased in a basket; it may have TID (transaction ID)
- A *transactional dataset*: A set of transactions
- An *itemset* is a set of items.
- An *itemset* is a set of items.
- A *k-itemset* is an itemset with k items.

- Given a dataset *D*, an itemset *X* has a (frequency) *count* in *D*
- An *association rule* is about relationships between two disjoint itemsets *X* and *Y*

$$X \Rightarrow Y$$

- It presents the pattern when *X* occurs, *Y* also occurs.

7.2 Modeling Techniques used
7.2.1 Pipes and Filters

Each component has a set of inputs and outputs. A component reads a stream of data on its input and produces a stream of data on its outputs. Input is transformed both locally and incrementally so that output begins before input is consumed (a parallel system). Components are called *filters*. Connectors serve as conduits for the information streams and are termed *pipes*.
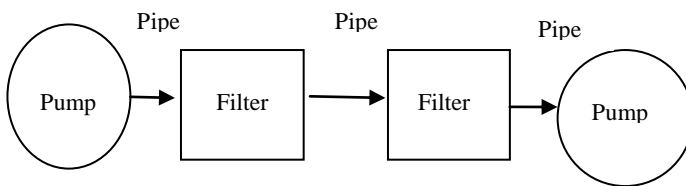
Fig 5: Pipe and Filter Architectural Style

7.2.2 Layered Systems

A layered system is organized hierarchically with each layer providing service to the layer above it and serv - ing as a client to the layer below. In some systems inner layers are hidden from all except the adjacent outer layer. Connectors are defined by the protocols that determine how layers will interact. Constraints include limiting interactions to adjacent layers. The best known example of this style appears in layered communication protocols OSI-ISO (Open Systems Interconnection – International Standards Organization) communication system. Lower levels describe hardware connections and higher levels describe application. Layered systems support designs based on increasing levels of abstraction. Complex problems may be partitioned into a series of steps. Enhancement is supported through limiting the number of other layers with which communication occurs. Disadvantages include the difficulty in structuring some systems in a layered fashion.

8. CONCLUSION

This paper focuses on clustering e-banking customer to analyze customer characteristics and behaviors with appropriated criteria: access time, transaction access and RFM Analysis, LTV, demographic variables. The benefits are valuable for the bank to improve services and the benefits includes Gain unparalleled insight into markets, Attract more

profitable customers, Maximize investments in sales and delivery channels, Secure a competitive advantage in the marketplace, Distribute branches more effectively, Align branches with defendable strategic performance, goals rooted in real-world customer behaviors and comprehensive market data Explore and quantify potential opportunities before making the critical decisions that impact the bottom line. Beyond simply understanding customer value in each cluster, the bank would gain opportunities to establish better customer relationship management strategies, improve customer loyalty and revenue and find opportunities for up and cross selling.

9. FUTURE ENHANCEMENTS

The future work will be on, the less profitable customers their accounts require highly automated interactions, while the high-yield accounts benefit from differentiated and better levels of service. Not only the marketing message will be delivered, institutions can alert the right departments and sales staff when customer communications are received, enabling appropriate and timely follow-up.

10. REFERENCES

[1] Waminee Niyagas, Anongnart Srivihok, and sukumal Kitisin "Clustering e-Banking Customer using Data Mining and Marketing Segmentation" The dataset of this study is Internet Banking customer data from one commercial bank in Thailand between January 1st and December 11th of the year 2005.

[2] Sang Chul Lee, Yung Ho Suh, Jae Kyeong Kim, Kyoung Jun Lee "A cross-national market segmentation of online game industry using SOM" in Elsevier Expert Systems with Applications 27 (2004) 559 – 570.

[3] Markus Schedl, Elias Pampalk, and Gerhard Widmer 2004 "Intelligent Structuring and Exploration of Digital Music Collection" in Austrian Research Institute for Artificial Intelligence.

[4] Nan-Chen Hsieh, Kuo-Chung Chu 2009 "Enhancing Consumer Behavior Analysis by Data Mining Techniques" in International journal of information management and sciences.

[5] K.V.Nagendra, C.Rajendra 2012 "Customer Behavior Analysis using CBA (Data Mining Approach)" in National Conference on Research trends in Computer Science and Technology.

[6] Derya Birant, Dokuz Eylul University, Turkey "Data Mining Using RFM Analysis" in an article Knowledge oriented applications in Data Mining.

[7] E.W.T. Ngai, Li Xiu, D.C.K. Chau, 2009 "Application of data mining techniques in customer relationship management: A literature review and classification" in ELSEVIER.