International Academic Institute for Science and Technology

International Academic Journal of Science and Engineering
Vol.1, No. 1, 2014, pp. 126-136.
ISSN 2454-3896

International Academic Journal of Science and Engineering

www.iaiest.com

Categories Customers Using Data Mining for Effective Communication and Increased Profitability

Mahin Tasnimi

MSc of Software Engineering, Shahre Rey Branch, Payame Noor University (PNU), Shahre Rey, Iran

Abstract

The aim of the categories of customers; identify their purchase behavior and finding similar patterns in their purchase database So that customers can buy and then the same customers for the loyalty to the organization's products. This research tries to focus on data mining models in the field of clustering, to categorize customers to improve customer relationship management and marketing strategies for each category of customers. Therefore, after reviewing the research and literature review of the research literature, RFM variables are used to category Customers and The data collected using the software SPSS Clementine was analyzed and K-means algorithm is used to cluster clients. Finally, using decision tree algorithm rules for each category of customers are extracted and the accuracy of the model was evaluated by the software. Obtained according to the rules that act on the basis of If-Then rules can be used to discover the hidden rules of payment data And to extract the pattern lies in each category. The results show that the proposed model can be used for marketing strategies.

Keywords: Customers, Data Mining, Software.

Introduction:

Competitive regional and global markets, institutions are required to create competitive advantage has and the use of data mining in customer relationship management can be a competitive advantage, he said. The market is more competitive organization's data more valuable and more valuable data mining techniques emerge. To maintain competitiveness, organizations need to develop strategies for customer focus, customer orientation and customer orientation are. All these demands of organizations in order to define the relationship with customers. Customer relationship management solution efforts to make research organizations as well as customers (ghazanfari, 2008). The basic premise is that all customers are not equal The main purpose of customer relationship management, optimization of profitable events to harmful events for groups of active customers Data mining and data analysis results to refining the right to make decisions about provides customer management strategies. Due to the high cost associated with all clients mechanism is needed to pick them up by category Customers will have the highest response rate, Customers are given the right to invest. Their customer behavior and profitability by various factors

such as competition and social factors changes. Therefore, managers should see just gain an overview of these changes and adopt an appropriate decision. Managers are effective and timely decisions about behavioral changes are not customers, was sharp decline in long-term profit and financial resources are consumed in vain. So prepare a model to identify customer behavior is essential (Koopaei and Bidgoli, 2009).

Known techniques of data mining tools to analyze their customer data. The use of statistical methods and artificial intelligence techniques, patterns in very large datasets extracted (Han and Kamber, 2006).

Data mining alone is not useful, but when it is applied in a particular cases Make sense. Data mining can roughly predict which potential customers into actual customers are As well as through the analysis of data can be models to predict the probability of customers and the possible exit attract other organizations making. To have a profitable transaction with the customer as much as possible and optimize the performance of customer relationship management system, we should have the ability to distinguish between good and bad and profitable and non-profitable customers.

We should know who they are and what differences with each other. In this study will try to Using data mining techniques to identify target customers and categorize them improve their customer relationship management strategies.

Literature Review

Customer Relation Management

Use customer profile is very important in most business activities because the first step in marketing strategy development, market segmentation and profiles of market segmentation results. In fact, the value of market segmentation, based on accurate customer profiles. Customer value analysis is a method of identifying characteristics of customer behavior (cheng & chen, 2009). Based on this analysis we can measure the level of customer loyalty. Not only maintain customer loyalty over time but creating a lasting relationship with the customer to encourage them to buy more in the future (Hosseini et al, 2010).

Customer Relationship Management refers to automate operational business processes. While analytical customer relationship management and customer behavior analysis features to support management strategies of customers. The superior and differentiated customer relationship management analytics to better allocation of resources to help the most profitable group of customers. The best tool for customer data mining tools for data analysis in analytical customer relationship management Organizations that can help to discover knowledge hidden in large amounts of data. In the CRM category is important because it provides create profiles for different customers and customer groups gives the possibility of strategic planning.

The concept of market segmentation by an American expert, Smith, Van Dell, was developed in the mid-1950s In fact, a technology to classify customers into groups with similar characteristics And tend to have the same patterns, respectively (Chang Tsai, 2011). Core market or market segmentation, marketing strategy is the core. Because marketing strategy consists of two parts: Choose a target market and develop effective marketing plan for success in the target market.

To classify customers and retail markets have different criteria and variables to be considered in the market. The ability to increase profits and return on investment point of view, the most appropriate method for the division of the market, is a method The greatest opportunities for profit and return on investment are created. Important factors in this regard include:

- Similarly, the needs of buyers in each section
- the different needs of buyers in various sectors of the market.
- The possibility of marketing activities to achieve a
- Simplicity and cost sharing different sectors of the market (Mirzazadeh et al, 2008).

There is no unique classification for the market. A marketer should try different variables that can be classified market on that basis, Identify (individually or in the aggregate) and to determine the best way to identify the market structure. Examples of variables to classify consumer market include:

1. Market Segmentation by geographic areas
2. Market segmentation based on demographic variables
3. Market Segmentation by psychological factors
4. Market Segmentation based on behavioral factors (Parsaeian, 2005).

Generally approaches in the field of market segmentation can be divided into two main groups:
The first group approach that Based on one or more known properties have been built. In this approach generally attractive market segments of the population are an option. As an example of this group can be segmented approaches, based on demographic variables mentioned.
The second group or new approaches to market segmentation query-based data operations and use of multivariate analysis. This approach has been the development of data mining and artificial intelligence techniques and the spread of information systems, developing more have found (Mirzazadeh et al, 2008).
In this research, market segmentation, market segmentation approaches second group of new approaches to be applied. For this purpose, the K-means algorithm with preset number of clusters will be used for market segmentation.
Data mining is a fundamental tool Customers are required to detect demographic characteristics (ghazanfari et al, 2008). Data mining techniques can be used to achieve a wide range of different purposes. Examples of uses include:

1. Identify profitable customers and their profiles
2. Predict customer buying behavior
3. Focus marketing efforts on potential customers who are likely to buy more.
4. Estimate the effectiveness of advertising
5. Optimize share of the customer's shopping cart
6. Buy the side and overhead to customers based on previous products
7. Predict fraud

One of the most useful areas of data mining, customer relationship management. Data mining and customer relationship management in all areas, including maintaining customer entered (Ngai, 2009).
Customer retention is a central component of customer relationship management cycle and of the utmost importance. The customer satisfaction is the difference between the current situation and the expectations he goes the condition is considered necessary to keep customers (Kracklauer & Seifert, 2004). According to research conducted by American researchers gain a new customer costs five times the cost of retaining an existing customer is This can be achieved in particular through the Service department (Cheng & Chen, 2009).
Customer loyalty programs are one of the main components of the stage which contains the efforts and support activities to maintain long-term relationship with the customer. Analysis of customer defection, actually measure the quality of service and customer satisfaction are customer loyalty programs.

Data analysis

Term data mining for the first time in 1989 at the Conference on Artificial Intelligence in the city of Detroit was introduced by Shapiro and Gyodgy Pyattsky. Knowledge discovery and data mining is an interdisciplinary field and growing in fields such as databases, statistics, machine learning and other related fields combined with the To information and valuable knowledge hidden in large volumes of data to extract (Ghazanfari et al, 2008).
Various methods for the implementation of data mining projects there. One of the methods is very strong, is the CRISP methodology. The methodology of the steps recognition system, data understanding, data preparation, modelling, evaluation and development system was formed. Each of these steps are divided into sub-sections in Figure 1, it has been shown.

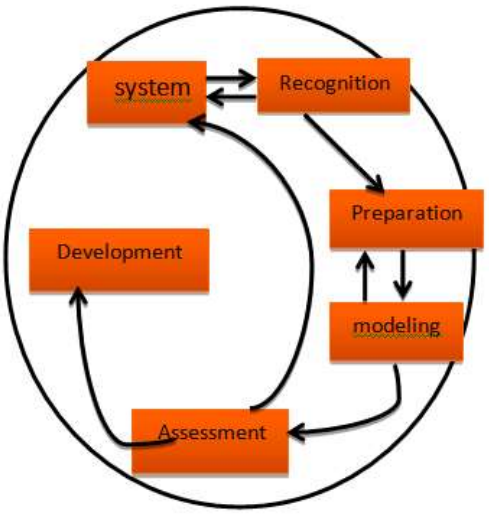


Fig 1: step methodology CRISP (spss clementine, 2010).

Data preparation for data mining Art squeeze and push out existing data and valuable data. While data mining is the art of finding meaningful patterns in the data (Ghazanfari et al, 2008).
Data preparation, 60 to 90 percent spend the time required for data exploration and data mining project to its successful 75 to 90 percent of the (Han & Kamber, 2006).
Pre-processing and data preparation stage of the project the most important and the best data mining. In pre-processing processes are as follows: Data cleansing, data integration, data conversion, data reduction. Based on the type of applications that data mining operations must be performed, various techniques used for each of these actions (Ngoi, 2009).
Data mining tasks can be divided into two main categories: Predictive and descriptive. The goal of predictive tasks that a special feature values based on the values of other properties can be predicted. Where used in the description of duties The goal-directed patterns (correlations, trends, spikes, curves and anomalies) in order to summarize the principles in data communications (MT et al, 2006).
Background research
According to research conducted in 2011-2000, 14 972 articles related to data mining techniques (DMT) is Among these articles, 216 articles from 159 journals from data mining applications and is based on surveys conducted, Growth has been driven expertise in data mining techniques While growth in the Articles related to data mining applications for solving problems or inconvenience this may have been in the organization (Liao & others, 2012).
Chen's credit rating performance in 2002 with the merger of feedback neural networks with traditional diagnostic analysis explained. Kim Yu Sun in 2004 has used neural networks to manage loans. Term data mining for the first time in 1989 at the Conference on Artificial Intelligence in the city of Detroit was introduced by Shapiro and Doctor Gyodgy Pyattsky. From that moment has written various articles in the field.

- Nan Chen Hsieh in the 2004 investigation as integrated data mining and scoring behavioral model for analyzing customer behavior Bank has done. In this study, using RFM model was used to evaluate the customers' loyalty to the bank and behavioral patterns were extracted.
- Hang And colleagues in 2005 to provide a hybrid model for validation and bank customers And classify groups of customers with high credit worthiness to identify the target group of customers with

the most profitable deals In this paper, a combination of decision trees and genetic algorithm was used.

- Hussein abdo and John Pyonton in 2008 compared to classic models (Such as logit, probit and discriminant analysis) with neural networks (artificial potential) customers have paid for accreditation.
- Ngay In 2009 in a study to review and categorize the literature in the field of data mining applications focused on customer relationship management. In this study, researchers examined the papers the period 2000 to 2006.
- Mahboobeh Khaje vand in 2011 under the title of an article estimating customer lifetime value based on RFM analysis of customer buying behavior on customer loyalty and profitability to enhance the market share concentration. In this paper, the customer lifetime value to categorize customers a cosmetic used. For this purpose, two approaches have been used: In the first approach, the methodology RFM is used for segmenting customers and the latter added an additional parameter to the method RFM Product count the items. A comparison between the two items show The code to add this parameter does not affect the category Customers Therefore, customer lifetime value based on the weighted RFM method for each section, is assigned The outcome of customer lifetime value to different segments can be used to describe the marketing and marketing strategies.

Hussein jewel respondents in 2008 to provide a model for master's thesis focused on direct marketing. This study is based on data mining techniques to identify the target market. The purpose of this research is to identify those customers who are more likely to respond to the products offered, In this study it was shown that Using a decision tree algorithms and correlation rules can be used to discover these laws to provide products related to paying customer's needs.

Mirzazadeh and colleagues in 2008 to measure customer loyalty models using self-organizing neural networks for master's thesis The goal is to provide an efficient method for classification of clients into different categories Loyalty, knowing the variables affecting customer loyalty and modeling relationships between variables is For the purpose of self-organizing neural network concepts used in modeling customer loyalty.

Research questions

- The lifetime value of customers using data mining and RFM model can be measured and how?
- How can customers based on intelligent data mining models classified?
- How can using RFM model and customer data, to improve management strategies to communicate effectively with paying customers?

Research methodology

The study of the purpose and the method of data analysis are described. As well as to assess the actual data model, customers will also be used.

Statistical Society

In this study, analysis of the raw data will be the database for fast food companies. The high volume of data that has been stored in a multi-year historical range. The population of the entire data stored by customers.

The important variables to categorize customers by reviewing the literature are extracted. In this study, two sets of data, including demographic data, financial transactions and customer data can be used to cluster clients. In this study, in order to separate customers from important behavioural factors, such as: 1-variable monetary 2-frequency, 3-current customer transactions to concisely (RFM) will be used In going after determining variables using the software SPSS-clementine customers are classified For this purpose the K-Means Clustering method will be used.

Research Findings

The methodology proposed model:

This study analyzed data fast food chain's baguette. The data from 01/01/2012 to 11/29/2012 4 baguette branch through the application speeds is collected in Excel output. The total number in 4000 is So that the customer purchase transactions during the year show. The process of analyzing data baguette with five major steps that followed each step will be explained in detail.

The first step: pre-processing data

The Bank recorded a database of customer transactions in 2012 related to the 4 branches of baguette we extracted separately in Excel output In order to convert the data into the correct format, records and fields with incorrect amounts of waste we remove. Finally, a database with approximately 2150 records and four fields are as follows: ID (ID) of each client as the client code is shared, the total number of purchases per customer in a year (Frequency), the volume of currency purchase per customer per year (Monetary) and date of last purchase per customer (Regency).

After pre-processing of data and remote data delete three main variables: the number of purchase, date of purchase and the purchase amount as the major variables are selected for this purpose, the distribution of data in these three variables is as follows.

Table 1. Statistical distribution of variables

	The number of purchases	Purchase	Last Buy
Count	2113	2113	2113
Mean	16.354	141600.700	127.158
Min	1	2900	4
Max	384	2402150	365
Range	383	2399250	361

Step 2. Determine the optimal number of clusters by the algorithm Two-step:

In order to find the optimal number of clusters Two Step algorithm with shaping distinct groups, each including at least one object are Starts The objects or groups close to one of the So that eventually an entire group be established at the highest level. Two Step Clustering is a two-step clustering method The first step with a transition of the input data is compressed them into manageable clusters. The second step uses a hierarchical clustering method, in order to integrate the evolution of this cluster of clusters larger and larger, benefit. This step does not have to pass this data. In order to find the optimal number of clusters of BSC index is used According to the table is the maximum distance of cluster 5 Which represents the number of clusters is optimal.

Step 3. Customers value segmentation algorithm using k-means:

In this step-by-RFM measures introduced by Hayes (1994) and we offered and little value RFM variables to consider as inputs. The customer value can be added by the algorithm k-means clustering. The details of this step is sub-divided into two sub-categories:

Step criteria defined variables 1-3 RFM: This section includes five sub-categories are:

• Select RFM variables with equal weights (eg 1: 1: 1).

• criteria define three variables (RFM Score) RFM: For this purpose, numbers 5, 4, 3, 2 and 1 assign to the variable. These figures reflect customer share in the profits of the company. The number 5 to those customers who have the highest share refers to the organization and the number 1 to those customers profit organization refers to the last level.

• Sort RFM variables are small to big
• RFM segmentation variables, respectively, in 5 equal parts so that all records accommodate.

Table 2. The real criterion variables R-F-M in the database baguette

Classification criteria	Benchmark Name	R-Regency	M-Monetary	F-Frequency
5	Too much	29/12 / 2012-18 / 12/2012	203300-2402150	23≤x<384
4	much	18/12 / 2012-06 / 11/2012	79000-203300	9≤x<23
3	middle	06/11 / 2012-14 / 08/2012	33000-79000	4≤x<9
2	little	14/08 / 2012-20 / 05/2012	14400-33000	2≤x<4
1	Too little	20/05 / 2012-01 / 01/2012	2900-14400	1≤x<2

• Calculation of little value based on Table 2, RFM variables as inputs for each customer

Table 3. Little value for customer's baguette RFM variables (according to Table 2).

ID	R	F	M
.....	….	…..	….
6472	5	4	4
6524	5	3	2
6599	5	2	2
6614	5	3	3
3973	4	4	4
1181	3	5	5
6327	3	1	1

Step 2-3. Clustering customer value by applying the algorithm K-means: According to quantitative value RFM variables for each client, the data is classified into 5 clusters. The K-means algorithm for clustering customer value we use. By k-means clustering results are shown in the table below.

Table 4. Clustering results with 5 clusters as output by means

Cluster centers	Cluster3	Cluster1	Cluster5	Cluster2	Cluster4
C1R	4.33	1.72	3.95	3.69	1.47
C2-AF	4.65	3.92	2.98	1.36	1.56
C3-RM	4.64	3.91	2.98	1.47	1.61
Distance to the zero point	7.68	5.80	5.78	4.20	2.68
Loyalty	Too much	much	middle	little	Too little
Number of records	507	435	331	371	469

Step 4. Creating rules:
After steps 1 and 2, three characteristics are obtained as input data. These three characteristics are distributed according to Table 4: 5. Last buy the latest transaction data per customer in 2012, buying the entire sequence of transactions per customer orders in 2012 and amount of money paid to buy the average purchase per customer in 2012. The study points out. Next, based clusters obtained in Step 2-2 and each of the three variables (RFM) can be extracted to make the decision and payment rules. For this purpose, the decision tree algorithms to analyze the categories used. 4 algorithms for analysis which is mainly an action similar to implement. They are all traits check the database has reached trait that classification and

prediction by splitting the data into subgroups does. This activity is repeated recursively until their subgroups, subgroups of another break.
C & R provides decision tree algorithms that try to predict the future and Classification observations. It is trying to reduce the impurities in each category. When a node is completely free of impurities that all elements of a subgroup it belongs to a category of the target field. If-then rules can be extracted from this algorithm to be used by RFM model.

Table 5. Describes the characteristics of RFM

features	(scaling name)
(Regency)	Very high, high, medium, low, low
(Frequency)	Very high, high, medium, low, low
(Monetary)	Very high, high, medium, low, low

Table 6. The decision based on RFM variables and the status of customer loyalty baguette

Clientele		Variables		Decision
ID	Regency	Frequency	Monetary	Loyalty
..... 5488 5674 4391 4128 3155	..... Too little Too little Too little middle middle	..... middle Too little much Too much middle	..... much middle much Too much little	..... much Too little much Too much middle .....

Based on clustering by k-means, RFM model laws that can be obtained from the output decision tree that With this legislation will be made according to how customer buying behavior, customer loyalty measure and identify the relevant category.

Conclusion:

The ultimate goal of this research is finding easy rules for decision-making in the field of customer relationship management strategies. According to Pareto principle 80% of an organization's profit comes from 20% of its clients and organizations need to identify their clients and know the importance of which share in the profit organization. The research in order to respond to your first question about the categories of customers using data mining techniques could RFM model and algorithm K-means, to categorize customers pay baguette. Also in line with the objectives of this study after identifying target customers Can be marketing strategies to promote effective communication with the customer. The following table is designed to respond to both questions listed Baguette so that the company will be marketing strategies.

Marketing strategies	Meme	Loyalty	Name Category
1. Especially given gifts on the occasion of Children's Day and Women's Day for those consumers 2. Nowruz greeting card containing a percentage discount food cards baguette in April 3. Send SMS for participating in the annual lottery baguette 4. Send email to this category of customers with content such as introduction of new food baguette, group discounts and....	Rule 1 for cluster-3 if Recency Score > 2.500 and Frequency Score > 3.500 and Monetary Score > 3.500 and Recency Score <= 3.500 and Monetary Score > 4.500 then cluster-3 Rule 2 for cluster-3 if Recency Score > 2.500 and Frequency Score > 3.500 and Monetary Score > 3.500 and Recency Score > 3.500 then cluster-3	Too much	Golden Customers
1. Send food orders with free starters (including salads, drinks, potatoes, etc.) 2.Telephone communication with those consumers one day after taking the quality and how to serve food and Feedback	Rule 1 for cluster-1 if Recency Score <= 2.500 and Frequency Score > 2.500 and Monetary Score > 2.500 then cluster-1 Rule 2 for cluster-1 if Recency Score > 2.500 and Frequency Score > 3.500 and Monetary Score > 3.500 and Recency Score <= 3.500 and Monetary Score <= 4.500 then cluster-1	much	Silver customers

Marketing strategies	Meme	Loyalty	Name Category
1. Especially given gifts on the occasion of Children's Day and Women's Day for those consumers .2. Nowruz greeting card containing a percentage discount food cards baguette in April 3. Send SMS special annual lottery baguette 4. Send e-mail to this category of customers with themes such as the introduction of new food baguette, group discounts and	Rule 1 for cluster-3 if Recency Score > 2.500 and Frequency Score > 3.500 and Monetary Score > 3.500 and Recency Score <= 3.500 and Monetary Score > 4.500 then cluster-3 Rule 2 for cluster-3 if Recency Score > 2.500 and Frequency Score > 3.500 and Monetary Score > 3.500 and Recency Score > 3.500 then cluster-3	Too much	Golden Customers
1. Send food orders with free starters (including salads, drinks, potatoes, etc.) 2.Telephone communication with those consumers one day after taking the quality and how to serve food and Feedback	Rule 1 for cluster-1 if Recency Score <= 2.500 and Frequency Score > 2.500 and Monetary Score > 2.500 then cluster-1 Rule 2 for cluster-1 if Recency Score > 2.500 and Frequency Score > 3.500 and Monetary Score > 3.500 and Recency Score <= 3.500 and Monetary Score <= 4.500 then cluster-1	much	Silver customers

Suggestions:

Practical suggestions

- Designed to evaluate the system should be taught how to work with the system to branch managers and short and long-term benefits for its members theories.
- It is also recommended to reduce the resistance; the system is designed to enjoy the support of the central office baguette.
- Recommendations for future research
- Given that this research is based on RFM variables to categorize customer's Personality traits and tastes and interests of their clients in fast food consumption is not considered it can be, if possible, to categorize customers based on features within them.
- Weight RFM variables considered in this study is same while depending on the industry and expert panel reviewed these weights can have different degrees of importance.
- In addition to the categories of customers based on RFM variables can be purchased products as well as another variable to add these variables and categorize customers based on the four variables.

References:

Chang Hui-Chu, Hsiao-Ping Tsai. (2011). "Group RFM analysis as a novel framework to discover better customer onsumption behavior". Expert Systems with Applications 38.

Cheng Ching-Hsue, You-Shyang Chen. (2009). "Classifying the segmentation of customer value via RFM model and RS theory". Expert Systems with Applications.

Ghazanfari Mehdi, Alizadeh S, T. pour Babak. (2008). "Data mining and knowledge discovery", Tehran: Publication of Science and Technology, First Edition.

Ghanbari-Baghestan, A., Esfandiari, H., Maghazei, O., & Khorasani, S. (2012). The Impact of Service Recovery on Customer Satisfaction: Case of Iran. *Archives Des Sciences*, *65*(11).

Han, J. and Kamber, M. (2006). "Data Mining : Concepts and Techniques " , Second Edition , Morgan Kaufman Publisher.

Hosseini, S. M. S., Maleki, A., & Gholamian, M. R. (2010). Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty. Expert Systems with Applications, 37(7), 5259-5264.

HosseinJavaheri, S. (2008). Response modeling in direct marketing: a data mining based approach for target selection.

Khajvand Mahboubeh, Kiyana Zolfaghar, Sarah Ashoori, Somayeh Alizadeh. (2001). "Estimating customer lifetime value based on RFM analysis of customer purchase behavior": case study. Procedia Computer Science 3.

Khorasani, S. T. (2014, August). Design-Driven Integrated-Comprehensive Model CDFS Strategic Relationships. In *ASME 2014 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference* (pp. V004T06A020-V004T06A020). American Society of Mechanical Engineers.

Koopaei, M., & Minaei-Bidgoli, B. (2009). August). A new method for ranking changes in customer's behavioral patterns in department stores. InProceedings of the 11th International Conference on Electronic Commerce (317-322). ACM.

Kracklauer, A. H., Mills, D. Q., & Seifert, D. (2004). "Customer management as the origin of collaborative customer relationship management", Collaborative Customer Relationship Management - taking CRM to the next level, 3-6.

Liao Shu-Hsien, Pei-Hui Chu, Pei-Yuan Hsiao. (2012). "Data mining techniques and applications – A decade review from 2000 to2011. Expert Systems with Applications".

M. MIRZAZADEH. (2008). "A Model to measure customer loyalty by using self-organizing neural networks". MA thesis, University of Allameh Tabatabai.

Ngai E.W.T, Li Xiu, D.C.K. Chau. (1999). "Application of data mining techniques in customer relationship management": A literature review and classification. Expert Systems with Applications.

Parsaeian Ali. (2005). "Principles of Marketing", Tehran: Tehran University Press.

Tan Pang-ning. (2006). Michael Steinbach, Vipin Kumar. "Introduction to data mining".