International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:9, No:8, 2015

Customer Segmentation Model in E-commerce Using Clustering Techniques and LRFM Model: The Case of Online Stores in Morocco

Rachid Ait daoud, Abdellah Amine, Belaid Bouikhalene, Rachid Lbibb

***Abstract*—**Given the increase in the number of e-commerce sites, the number of competitors has become very important. This means that companies have to take appropriate decisions in order to meet the expectations of their customers and satisfy their needs. In this paper, we present a case study of applying LRFM (length, recency, frequency and monetary) model and clustering techniques in the sector of electronic commerce with a view to evaluating customers' values of the Moroccan e-commerce websites and then developing effective marketing strategies. To achieve these objectives, we adopt LRFM model by applying a two-stage clustering method. In the first stage, the self-organizing maps method is used to determine the best number of clusters and the initial centroid. In the second stage, k-means method is applied to segment 730 customers into nine clusters according to their L, R, F and M values. The results show that the cluster 6 is the most important cluster because the average values of L, R, F and M are higher than the overall average value. In addition, this study has considered another variable that describes the mode of payment used by customers to improve and strengthen clusters' analysis.

The clusters' analysis demonstrates that the payment method is one of the key indicators of a new index which allows to assess the level of customers' confidence in the company's Website.

***Keywords*—**Customer value, LRFM model, Cluster analysis, Self-Organizing Maps method (SOM), K-means algorithm, loyalty.

I. Introduction

ACCORDING to the figures released by Interbank Electronic banking Centre (IEBC), the e-commerce sector in Morocco has experienced a +10,3 % increase in the first quarter of 2015 in the number of transactions and +10,5 % in the amount of spent relative to the same period in 2014 [1].

The birth of the company Morocco Telecommerce in 2001, the first leading operator of electronic commerce in Morocco helped to crystallize the ambitions of many entrepreneurs who have found an effective way to secure their transactions through the electronic payment security system. This was setup by the company in question, which is certified and recognized by the Interbank Electronic Banking Centre (IEBC), Moroccan banks, and by international organizations such as Visa and MasterCard.

577 contracts were signed late in 2014 with certifying bodies such as Morocco Telecommerce (MTC) and the Interbank Electronic Banking Centre (IEBC). There were 140 at the end of 2010. In the end, it is 577 e-commerce sites that are currently active and referenced by Morocco Telecommerce (MTC). The trend is not ready to calm down. The year 2015 should also know its own share of novelties, since the IEBC expects to reach 700 affiliated online merchants by the end of 2015. No doubt, this continuous increase of new sites allows for the improvement and activation of more online sales [1].

With this increase in the number of merchant sites affiliated with the IEBC, they have realized 527000payment online-operations with credit cards, both local and foreign, for a total of MAD 285,3 million in the first quarter of 2015. The IEBC added that the activity by the Moroccan card rose from 487 000 transactions in the first quarter of 2014 to MAD 506 million during the same period in 2015 (+5,8%) and from 242,5 million to MAD 249,8 million (+3%) in terms of their amount [1]. According to these figures, the Moroccan internet users become more familiar with the online payment. In order to survive and cope with the competition related to the explosive growth of electronic commerce, the companies must develop innovative marketing activities to identify different customers and try its best and utmost to preserve them or, at least, care for the most loyal ones and get their satisfaction, because the identification of such segments can be the basis for an effective strategy to target and predict potential customers [2].

Market segmentation is the process of identifying key groups within the general market that share specific characteristics and consuming habits and it provides the management to customize the products or services to fulfill their needs [3]. Nowadays, RFM model, which was developed by Hughes (1994), is one of the most common methods for segmenting and identifying customer values in companies. This method depends on Recency, Frequency and Monetary measures which are considered as three important variables used to extract the behavioral characteristics of customers, and that influence their future purchasing possibilities [4]. By adopting RFM model, marketing managers can effectively target valuable customers and then develop marketing strategies for them based on their values [5].

Recent studies find that the addition of supplementary variables to the classical model RFM can improve its

Rachid Ait Daoud and Rachid Lbibb are with the Department of Physics, Sultan Moulay Slimane University, PB 523, Beni Mellal, Morocco (e-mail: daoud.rachid@ gmail.com, rachid.lbibb@gmail.com).

Abdellah Amine is with the Department of Applied Mathematics, Sultan Moulay Slimane University, PB 523, Beni Mellal, Morocco (e-mail: amine71de@yahoo.de).

Belaid Bouikhalene is with Department of Mathematics and Informatics, Sultan Moulay Slimane University, PB 523, Beni Mellal, Morocco(e-mail: bbouikhalene@yahoo.fr).

predictability when predicting customer behaviors [6]. For example, the model RFM was extended by [7] by adding an additional variable customer relation length (L) to it, to become LRFM model, because by adopting RFM model, the companies cannot effectively distinguish between the short-life and long-life customers [8]. (L) measures the time period between the first visit and the last visit of a particular customer. In this paper, we use the results of this method (LRFM) as inputs for clustering algorithms to determine the customer's loyalty for an online selling company in Morocco.

A real case study for an online selling company in Morocco is employed by combining LRFM model and data mining techniques (cluster analysis) to achieve better market segmentation and improve customer satisfaction. Data mining techniques such as Self Organizing Maps and K-means are used in this study to group all customers into clusters. The characteristics of each cluster are examined in order to determine and retain profitable and loyal customers. As mentioned before, the customers are segmented into similar clusters according to their LRFM values.

The customer purchases database consists of 730 customers who purchased directly from the company website from November 2013 to January 2015. The profile for each customer includes the customer identifier, gender, birth date, city, shopping frequency, date of first transaction, date of last purchase, the total expense, and mode of payment.

The remainder of the paper is as follows: Section II provides the literature review on RFM, LRFM models and data mining techniques. Section III reports the methodology used to conduct this study. Section IV presents the empirical results. Finally, conclusions, managerial implications, limitations and further research are depicted.

II. Literature Review of Segmentation Methods and Clustering

A. RFM and LRFM Models

Recency, frequency and monetary (RFM) is an effective method of segmenting and it is likewise a behavioral analysis that can be employed for market segmentation [9], [10]. Reference [9] describes that the main asset of the RFM method is, on the one hand, to obtain customers' behavioral analysis in order to group them into homogeneous clusters, and, on the other hand, to develop a marketing plan tailored to each specific market segment. RFM analysis improves the market segmentation by examining the when (recency), how often (frequency), and the money spent (monetary) in a particular item or service [11]. Reference [11] summarized that customers who had bought most recently, most frequently, and had spent the most money would be much more likely to react to the future promotions.

The advantage of RFM model resides in its relevance as long as it operates on several variables which are all observable and objective. They are all available at the order's past for each customer. These variables are classified according to three independent criteria, namely recency, frequency and monetary [12]. Recency is the time interval between the last purchase and a present time reference; a lower value corresponds to a higher probability that a customer will make a repeat purchase. Frequency is the number of transactions that a customer has made in a particular time period and monetary means the amount of money spent in this specified time period [13].

The traditional approach to adopt RFM model is to sort the customers' data via each variable of RFM and then divide them into five equal quintiles [2], [14]. The process of segmentation begins with sorting all customers based on recency, then frequency and monetary. For recency, the customer database is sorted in an ascending order (most recent purchasers at the top). Customers are then sorted for frequency and monetary in a descending order (most frequently and had spent the most money were at the top).The customers are then split into quintiles (five equal groups), and given the top 20%segmentis assigned as a value of 5, the next 20% segment is coded as a value of 4, and so on. Therefore, all customers are represented by one of 125 RFM cells, namely, 555, 554, 553, . . . , 111[15], [16].

Customers who have the most score are profitable. In this study, we adopt another approach proposed by [17], it consists of using the original data rather than the coded number. The definitions are as follows: Recency is the time interval between the first day of study period and the last purchase; frequency is the number of transactions that a customer has made in a particular time period and monetary means the amount of money spent in this specified time period.

Some researchers try to develop new RFM models by adding some additional parameters to it so as to examine whether they achieve good results than the basic RFM model or not [18]-[20]. For example, [19] selected targets for direct marketing from a database by extending RFM model to RFMTC, by adding two parameters, namely time since the first purchase (T) and churn probability (C). Another version was proposed by [21] Timely RFM (TRFM) model consists of adding one additional parameter, the period of product activity to determine the relationship of product properties and purchase periodicity i.e. to analyze different product demands at different moments. Chang and Tsay propose the LRFM model, by taking the customer relation length into account, in order to resolve RFM model problem related to the difficulty of distinguishing between customers, who have long-term or short-term relationships with the company [7]. In addition, [22] suggests that the customer's loyalty and profitability depend on the relationship between a company and its customers. In this regard, in order to identify most loyal customers, it is necessary to consider the customer's relation length (L), where L is defined as the number of time periods (such as days) from the first purchase to the last purchase in the database.

B. Cluster Analysis

Data mining techniques have been widely employed in different domains. As the transactions of an organization become much larger in size, data mining techniques, particularly the clustering technique, can be applied to divide all customers into several clusters based on some similarities in these customers [23]. Clustering techniques are used to identify a set of groups that both minimize within-group variation and maximize between-group variation according to a distance or dissimilarity function [24].

The SOM (Self-Organizing Map) is an unsupervised neural network methodology, which needs only the input is used to clustering for problem solving [25] and market screening [26]. The network is formed by an unsupervised competitive learning algorithm, which can detect for itself (which means that no human intervention is needed during the learning process) patterns, strong features, and correlation in the large input data and code them in the output [27]. The patterns of SOM in a high-dimensional input space are originally very complicated. When projected on a graphical map display, its structure, after clustering, turns out to be not only understandable but more transparent as well [28].

K-means clustering is the most common algorithm used to cluster n vectors based on attributes into k partitions, where k < n, depending to some measures [29]. The name comes from the fact that k clusters are identified and the center of a cluster is the mean of all vectors within this cluster. The algorithm starts with choosing k random initial centroids, then assigns vectors to the nearest centroid using Euclidean distance and recalculates the new centroids as means of the assigned data vectors. This process is repeated many times until vectors no longer altered clusters between iterations [30]. The K-means method is arguably a non-hierarchical method. However, SOM has a few disadvantages. For example, with the result generated by SOM technique, it is difficult to detect clustering boundaries, a fact which limits their application to automatic knowledge [25]. Furthermore, in the k-means technique, the number of clusters and the initial starting point are randomly selected, which means that the algorithm has to turn several times to identify strong forms, because the final result depends on the initial starting points (different initial k objects may produce different clustering results). Due to the weakness of SOM and k-means method, the integration of these methods becomes desirable. Reference [31] took this view, adopting a two-staged clustering method by integrating the hierarchical method into the non-hierarchical.

Kuo et al. [32] have pointed out that it is preferable to use iterative partitioning methods instead of the hierarchical methods if the initial centroid and number of clusters are provided. If the information is provided, the iterative method consistently finds better clusters and higher accuracy than the hierarchical methods and yields faster results because the initialization procedure that ultimately determines the number of iteration is already executed. One example proposed by [31] is to adopt a two-staged clustering method by deploying Ward's minimum variance method to obtain the number of clusters and also to provide the starting point. Then the non-hierarchical methods, like the k-means method can use the result of the Ward's minimum variance method to find the final cluster solution. On the other hand, [32] have proposed a modified two-stage method by applying self-organizing feature maps to determine the number of clusters for K-means method. The reason is that Self-Organizing Maps can converge very fast since it is a kind of learning algorithm that can continually update or reassign the observations to the closest cluster. Therefore, this study uses self-organizing feature maps to determine the number of clusters and the initial starting points that K-means method need.

In the first stage, data set is clustered via adopting the SOM. From the final output array, we can easily determine the candidate number of clusters as well as the initial centroid. In the second stage, the starting point and the derived approximation of the clusters (k) determined in the first stage are used with K-means method. Wei et al. [24] pointed out that self-organizing maps (SOM) and K-means method are commonly used for cluster analysis.

III. Research Methodology

In this section the proposed model to determine loyal and profitable customers is described.

The purpose of this case study is customer segmentation using LRFM model and clustering algorithms (SOM and K-means) to specify loyal and profitable customers for achieving maximum benefit and a win-win situation.

In order to identify most profitable customers, it is necessary to consider the ''mode of payment factor'' in the company. Fig. 1 shows the required steps for the proposed model.

A. Understanding Data

Dataset used in this case study was provided by a company selling online in Morocco and collected through its e-commerce website.

All of the transactions carried out by customers are stored in a MySQL database. From this database, we will design a data warehouse that contains a wide variety of products, descriptive information on each customer and transactional data.

The transactional data consist of 730 customers who have purchased the website of the company from November 2013 to January 2015.

Customers have four modes of payment: Cash on delivery, online credit card, bank transfer and payment in three installments.

B. Data Preparation for Segmentation

Data preprocessing is one of the most important and often time-consuming aspects of data mining project [33], [34]. In this case study, data preprocessing techniques such as data selecting, data cleaning, data integration and data transformation were used to improving the quality of data clustering.

The purchase orders included many columns such as transaction id, product id, customer id, ordering date, item price, item quantity purchased, total amount of money spent, and payment modes.

While customer table includes the following fields such as customer id, gender, marital status, birth date, email, address, city; product table included attributes such as product id, barcode, brand, category subcategory, price and quantity.

Customers must have made at least one online transaction during the study period; otherwise, they will be ignored and removed from the dataset.

According to the transformation data, and with respect to the length (L), if the first and last transaction dates are identical, the length is coded as 1. Gender equals 1 if the customer is a male and 0 if the customer is a female. The modes of payment are divided into four modes: (1) payment by cash on delivery is coded as A, (2) payment by online credit card is coded as B, (3) payment by bank transfer is coded as C and, finally (4) payment in three installments, is coded as D. In this stage, the transaction recorded for each customer must be transformed to a usable format for the LRFM model. From the Integrated dataset, the L, R, F and M variables were extracted for each customer.

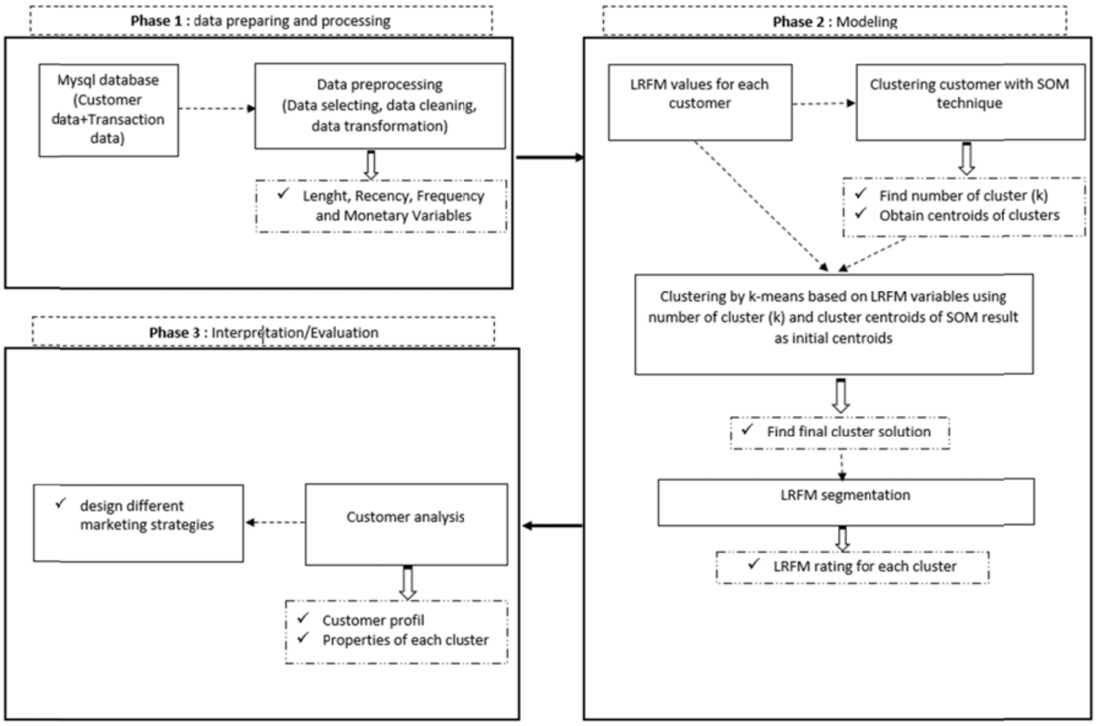

Fig. 1 Framework for customer segmentation based on LRFM model and clustering techniques

TABLE I
DATA FORM WITH SEVEN ATTRIBUTES

Attribute name	Data content
City	The city's variable of the transaction represents the city which hosted the commercial transaction
Gender	Gender of the customers
Modes of payment	Indicate the mode of payment used in each transaction
Transaction length (L)	Refers to the number of days from the first and the last purchases
Recent transaction time (R)	Refers to the number of days between the first day of study period and the last day of purchase
Frequency (F)	Refers to the total number of purchases from November 2013 to January 2015
Monetary (M)	Refers to the total amount spent by customers from November 2013 to January 2015. (Moroccan Dirhams or MAD)

TABLE II
THE DESCRIPTIONS OF LENGTH, RECENCY, FREQUENCY AND MONETARY

Variables	Max	Min	Average	Standard deviation
L	1084	1	438,86	266,18
R	438	1	296,81	113,93
F	19	1	10,16	6,17
M	13731,60	69,00	3947,18	3405,75

IV. EMPIRICAL CASE STUDY

The case studied in this research is an online selling company. This company is one of the biggest online retailers specialized in electronics, fashion, home appliances, and children's items in Morocco. It has been working and affiliated with MTC since 2008, has a contractual relation with a MTC and IEBC, and undertakes to fulfill all commitments and responsibilities stated in its general terms and conditions of sale.

The most important goal of this case study is customer segmentation using LRFM model and clustering algorithms (SOM and K-means) to specify loyal and profitable customers for achieving maximum benefits and a win-win situation.

The definition of the LRFM model used in this study is shown in Table I.

The descriptive statistics for the variables in this study (LRFM) are provided in Table II.

The maximum and minimum lengths' values in terms of days are calculated to be 1084 and 1, respectively. For recency, the larger value demonstrates that the customer has shopped recently. What is more, the maximum and minimum recency values are 438 and 1, respectively. In what concerns frequency, the maximum and minimum values are calculated to be 19 and 1, respectively. For monetary, the highest and lowest total amount spent are MAD 13731.60 and MAD 69, respectively (calculated in Moroccan Dirham).

Therefore, it is necessary to further decompose the dataset for more details. Tables III and IV present the characteristics of L, R, F and M for different genders, and different modes of payment, where the majority of customers prefer to use payment by cash on delivery to pay for their purchases. The percentage of male customers is twice as high as that of females (67,9% versus 32,1%).

TABLE III
CHARACTERISTICS OF L, R, F AND M FOR DIFFERENT GENDERS

	Gender	
	Female	Male
Size	234 [32,1%]	496 [67,9%]
Average of L	370,92	470,91
Average of R	242,53	309,31
Average of F	6,99	11,65
Average of M	2772,98	4559,20

TABLE IV
CHARACTERISTICS OF L, R, F AND M FOR DIFFERENT MODES OF PAYMENT

	Modes of payment			
	Bank transfer	Cash on delivery	Credit card	Three installments
Size	1384	2885	2520	625
Average of L	551,40	324,46	587,45	2,14
Average of R	288,61	257,08	321,01	324,93
Average of F	9,54	9,34	12,35	8,68
Average of M	3168,69	3507,10	4358,22	6639,00

According to the proposed model described in section 3, IBM SPSS Modeler 14.2 is used for self-organizing maps method and k-means method to perform cluster analysis. First step: By applying self-organizing maps (SOM) “Kohonen node” to cluster 730 customers, we found that the number nine is the best number of clustering based on the characteristics of length, recency, frequency and monetary. The result of this method is shown in Fig. 2. Later the number nine of cluster (k) generated by SOM can be used as a parameter for the second step. In this step, K-means method is applied to find the final solution. So, there are nine clusters of customers that have similar LRFM behavior. Table V is a summary of the clustering of these nine clusters, each with the corresponding number of customers, average length (L), average recency (R), average frequency (F) and average monetary (M). The last row shows the overall average for all customers. For each cluster, the values of LRFM were compared with the overall averages. If the average (L, R, F, M) value of a cluster exceeds the overall average (L, R, F, M), then an over bar appears. However, if the average (L, R, F, M) value of a cluster does not exceed the overall average (L, R F, M), an under bar appears (i.e., $\overline{R}$: Higher R value; customers have purchased recently, $\underline{R}$: Lower R value; customers have not shopped the store recently). The last column shows the LRFM pattern for each cluster. Cluster 4 includes the minimum number of customers (only 60 customers). Whereas, the number of customers in clusters 7 and 1 is very important.

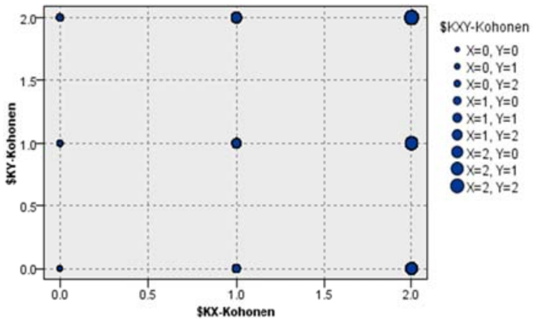

Fig. 2 Nine Clusters generated by SOM technique

Before interpreting the result for the nine clusters generated by k-means, we recall that only those customers who have made online purchases during the study period are taken into consideration.

For Cluster 3, customers have the lowest values of L, R, F and M ($\underline{L}\,\underline{R}\,\underline{F}\,\underline{M}$) compared to other clusters. In terms of length and recency, these customers have recently joined the online store, and in terms of frequency and monetary, these customers are not those who spend a lot of money, and even not those who often purchase. This cluster is called uncertain new customers. In addition to Cluster 3, customers in Cluster 9 are new customers; the only difference between the two is that customers in Cluster 9 have recently shopped at the merchant site.

Cluster 9 has a high R value ($\underline{L}\overline{R}\underline{F}\,\underline{M}$); no company wants to lose a future or potential customer. This encourages us to intelligently interact with the customers more often. Thus, we need to strive to incentivize them to buy more in order to maintain a long-term relationship with the company, and increase their frequency and monetary.

The average values of L, R, F and M for Cluster 6 are very high ($\overline{L}\,\overline{R}\,\overline{F}\,\overline{M}$) compared to the overall average values. More specifically, they have the greatest value for the money spent, the very high frequency of purchase, the highest value of recency, and, more importantly, they have established a long and close relationship with the company. Undoubtedly, these seventy-two customers are loyal and most profitable because Cluster 6 itself consists of customers who have recently made regular purchases, often purchases and spend a lot of money. These customers are probably the most valuable customers. The online store must show to the customers the interest which it carries for them, for example, rewards or incentives could be offered according to their purchases (The discount coupons, Delivery costs, the addition of a special gift in the package, etc.).

TABLE V
DESCRIPTIVE STATISTICS OF NINE CLUSTERS BASED ON K-MEANS METHOD

Cluster	Size	Average of L	Average of R	Average of F	Average of M	Pattern
Cluster 1	95	697.74	342.14	16.33	3 559.70	$\overline{L}\,\overline{RF}\underline{M}$
Cluster 2	83	770.63	368.29	5.81	1 743.49	$\overline{L}\,\overline{R}\underline{F}\,\underline{M}$
Cluster 3	88	143.13	84.73	2.41	684.45	$\underline{L}\,\underline{R}\,\underline{F}\,\underline{M}$
Cluster 4	60	220.00	362.67	6.80	6 284.83	$\underline{L}\overline{R}\underline{F}\overline{M}$
Cluster 5	73	211.21	182.42	13.37	4 074.19	$\underline{L}\,\underline{R}\overline{FM}$
Cluster 6	72	679.21	381.04	16.03	11 127.82	$\overline{LRFM}$
Cluster 7	104	296.78	367.78	16.76	6 340.59	$\underline{L}\overline{R}\,\overline{FM}$
Cluster 8	77	537.95	165.97	7.25	2 076.83	$\overline{L}\underline{R}\,\underline{FM}$
Cluster 9	78	355.33	334.62	4.23	924.11	$\underline{L}\overline{R}\underline{F}\,\underline{M}$
Total	730	438.86	287.90	10.16	3 986.63	

TABLE VI
MODES OF PAYMENT DISTRIBUTION AND GENDER FOR EACH CLUSTER BY K-MEANS TECHNIQUE

Variables		Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Total
Modes of payment	Cash on delivery (A)	427	125	158	178	596	67	1010	83	241	1384
	Credit card (B)	820	190	30	0	210	748	340	150	32	2885
	Bank transfer (C)	287	167	24	39	22	287	178	325	55	2520
	Three installments (D)	17	0	0	191	148	52	215	0	2	625
	Total	1551	482	212	408	976	1154	1743	558	330	7414
Gender	Female	15	24	61	14	26	12	20	24	38	234
	Male	80	59	27	46	47	60	84	53	40	496
	Total	95	83	88	60	73	72	104	77	78	730

The customers in Cluster 7 have low L value ($\underline{L}\overline{R}\,\overline{F}\,\overline{M}$). The low L value indicates that these customers have not yet established a long relationship with the online store. In addition, despite the lower value of L, it is observed that the average value of R, F and M are above the overall average, which might indicate that these customers purchase recently and frequently with a high money spent. So, they could be the customers with profit potential in the near future. The company must develop an effective marketing strategy whose aim is to encourage the customers in this cluster to migrate to the Cluster 6, by encouraging them to continue performing their online purchases, by providing a marketing program adapted to their purchases and by sending emails focused on their needs (monthly newsletter that is informing customers about the latest news of special promotions or sending emails to these customers to wish them a “happy birthday” or “happy anniversary” of the date they became customers, Request of opinion and so on).This allows to build a solid relationship between the online store and customers, and is more likely to create loyal customers.

Cluster 4 includes the minimum number of customers (only 60 customers). They have low average length and relatively low average frequency, but the average recency and monetary are very high ($\underline{L}\overline{R}\underline{F}\overline{M}$). Even though they have made very few transactions they have managed to spend a very significant amount of money.

Table VI illustrates that the majority of these customers prefer to use payment in three installments (D) as a mode of payment, which has been recently proposed by the online store, and it is reserved exclusively for the products, whose prices exceed MAD 1500. This indicates that these customers usually buy expensive items. Cluster 4 is called big spender customers potentially loyal.

The online store might concentrate its efforts in maximizing the loyalty of these big spenders. Therefore, it should place a particular emphasis on the satisfaction of these customers because the customer satisfaction contributes to increase the customer loyalty [35], e.g. by offering these customers special services such as a special discount, providing targeted and personalized promotions according to their profiles and their previous purchases. In this way, the online store keeps these customers coming back for more purchases.

Cluster 8 has higher L value but lower R, F and M values compared to the overall average L, R, F and M values ($\overline{L}\underline{R}\,\underline{F}\,\underline{M}$). The customers in Cluster 8 belong to former customers, who show little interest in items and services provided by the online store. This lack of attention is determined by the low number of transactions made by these customers, the low value of recency and their small contribution to the company. They begin to lose contact with the online store because they have not been heard of for a long time. Perhaps these customers were not satisfied with the services and products which they have received, or they have been attracted by the competitors, or they have lost confidence in the merchant site. This cluster is called the lost customers. Therefore, the company should identify the reasons and solve problems quickly to bring back those customers.

For Cluster 1 and 2, the values of L and R are above the average values. They have the characteristics of high recency, and more importantly, longer relationships with the online store. It can be said that these two clusters are loyal, but they do not have the right profile to become profitable customers, because their contribution to the company is still low even though they have the longest relationship with the online store compared with other clusters. The only difference between the customers in Cluster 1 and the customers in Cluster 2 is that the former purchased more often.

Customers in these clusters might become more important if the online store transfers them to Cluster 6, which represents the best customers of the company. The best strategy to achieve this goal would be to set up psychological pricing techniques and a special discount, which have a great impact on purchasing decisions. Again, this strategy proves efficient in that it encourages customers to purchase more frequently and spend much more money. Among these techniques, we can mention the term “FREE” that attracts the attention of any customer, even the ones who were not planning on spending anything in the first place. Another “FREE” technique is offering free delivery on all purchases of MAD150or more, proposing special promotional offers such as 1+1=3. The last technique is “The charming #9” which serves to lower items’ prices by one cent (prices ending in 9 ex. MAD 4,99) in order to boost sales.

A number of studies and experiments have confirmed this trend, for example, the experiments of winter clearance catalog of a direct-mail women’s clothing retailer conducted by Drs. Robert Schindler and Thomas Kibarian in 1996 [36], the cheese experience was conducted in 2005 by Nicolas Gueguen and Odile Jacob [37] and the experience of pancakes with door-to-door still in 2005 was conducted by Nicolas Gueguen and Odile Jacob [37] in order to confirm the results of the previous experiments. All of these experiences confirm increased numbers of customers to 29,7% by just lowering prices by one cent (e.g., $5.99 vs. $6.00). Finally, Cluster 5 has high F and M values but low L and R values ($\underline{LR}\overline{FM}$), spending and number of transactions indicate that these customers are more frequent and spend enough money in a short period. They can be considered as profitable customers, but the very low value of R indicates that these customers have not purchased from the online store for a long period of time. They are dormant customers. Something went wrong with these customers, because they have shown a keen interest in the online store so early. Perhaps these customers will soon stop making purchases on the Website due to several reasons including: The customers no longer need products and services offered by the company, or they are dissatisfied with the poor quality of the produce. Therefore, the online store must maintain a close contact with these customers. The major marketing strategies for these customers is to come into contact with them by the creation of a customer reactivation program, e.g., providing exceptional promotional offers limited in time to establish a sense of urgency to trigger customer purchasing to restore the relationship with these customers and, therefore, to increase the retention rate.

To be responsive and in order to make the right decisions for the development of the activity of the online store, we used a tool called Microsoft Power Business Intelligence, which integrates a decision-making approach. This tool allows us to produce secured and interactive dashboards which provide marketing managers and sales directors to: consult regularly with the statistics and the reports, share the reports with the main actors (direction committee, marketing team, PDG, etc.) and announce the reports according to the period, with the intention of having a global vision and a high quality in terms of the performances of the merchant site.

A more detailed analysis regarding the length of the relation, recency, frequency, monetary, gender and the mode of payment for the nine clusters are reported in Fig. 3.

Fig. 3 is a report that contains multiple visualizations that energize clusters data. With the exception of customers in Cluster 3, the number of male customers is always larger than that of females. The majority of customers in Cluster 3, 9, 7 and 5 prefer to use payment by cash on delivery to pay for their purchases. Payment by credit card is the most popular mode used by Clusters 6, 1 and 2. The mode of payment by bank transfer is often used by customers in Cluster 8. Finally payment in three installments remains the preferred mode by the customers in Cluster 4. When examining the length (L) among different payment methods, the results have shown that the longer the relationship between the merchant site and the customers is; the more customers put their trust in the electronic system of payment proposed by the online store. This applies to Clusters 1, 2 and 6 wherein customers prefer to use the credit card such as the payment method. Moreover, clusters that have a low value of L such as Clusters 3, 5, 7 and 9 prefer to pay for their purchases using cash on delivery. Perhaps those customers that have recently joined the merchant site did not trust the electronic payment systems.

The payment method is one of the key indicators of a new index which allows assessing the level of customer confidence in the company's Website. The online store should, therefore, encourage customers to pay for their online purchases by the use of Moroccan or foreign cards in order to create a climate of trust with these customers, Because if the company manages to establish a link of trust with its customers, this will help it in promoting a sense of satisfaction and encouraging a long-term relationship [38], [39].

Fig. 4 shows how the visualizations belonging to the same report can be filtered, exploit other visualizations, and interact with them. For example, to visualize the results of Cluster 7, just click in the legend CLUSTER on Cluster 7. Therefore, the results for this Cluster are highlighted in the report and the rest of the results are dimmed. Another report illustrated in Fig. 5 provides more details about the nine clusters, including the total revenue, the relative value of each cluster in terms of total revenue and number of transactions, the value of the average basket per cluster and, finally, the number of transactions and revenue generated by payment method.

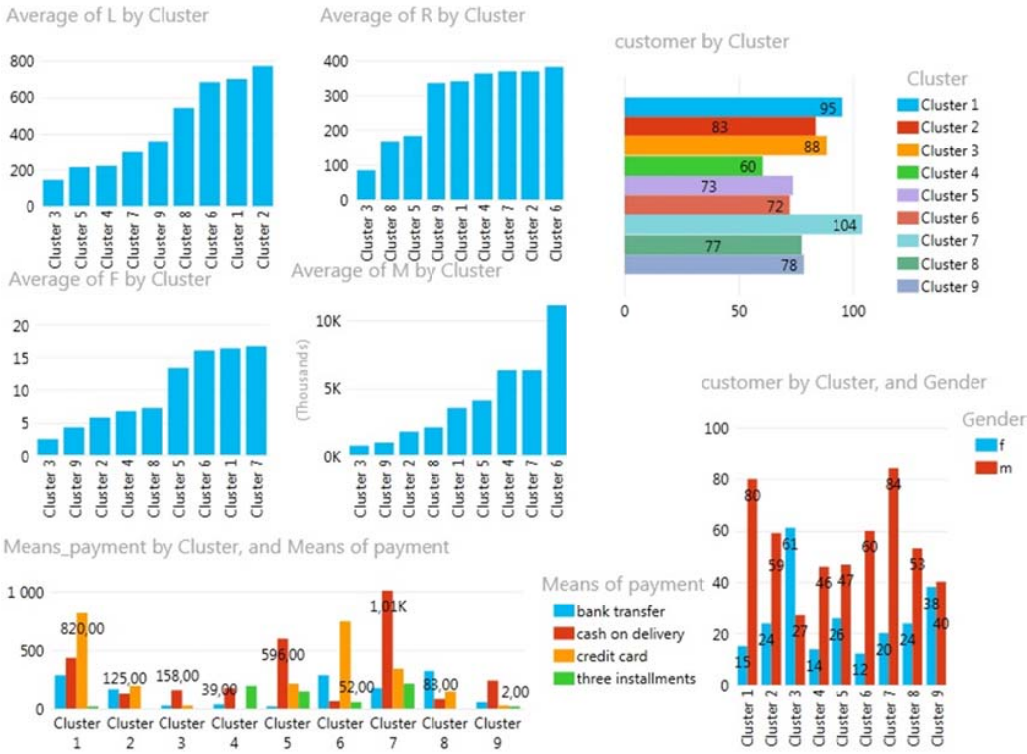

Fig. 3 More details about the nine clusters

Fig. 4 More details about the cluster 7

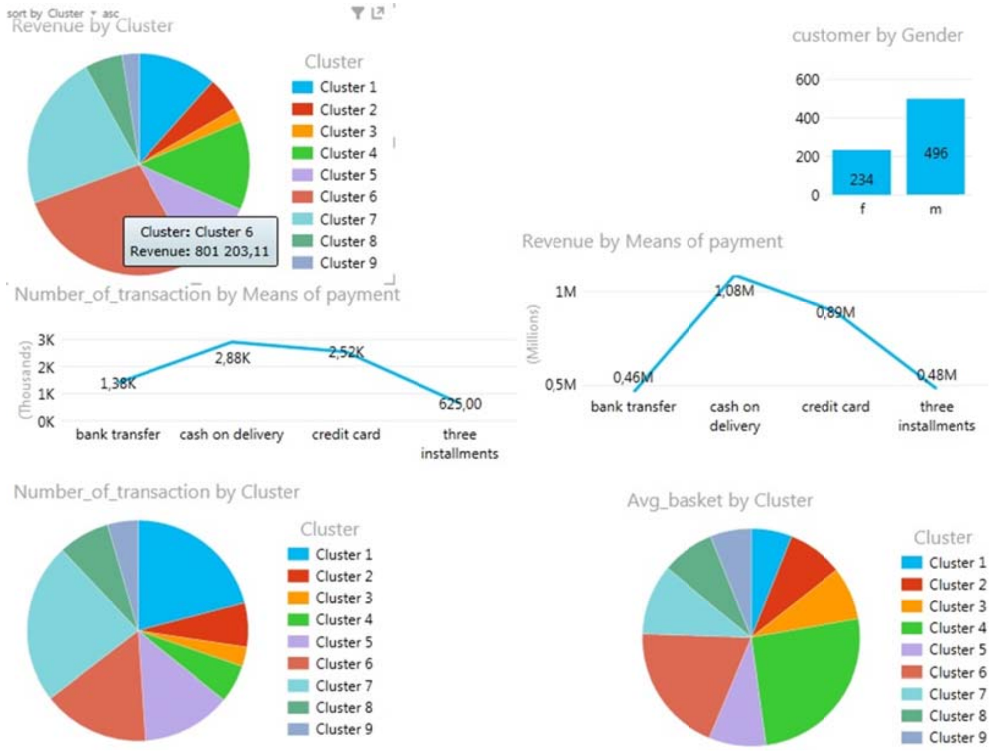

Fig. 5 An overview on the performance of each cluster

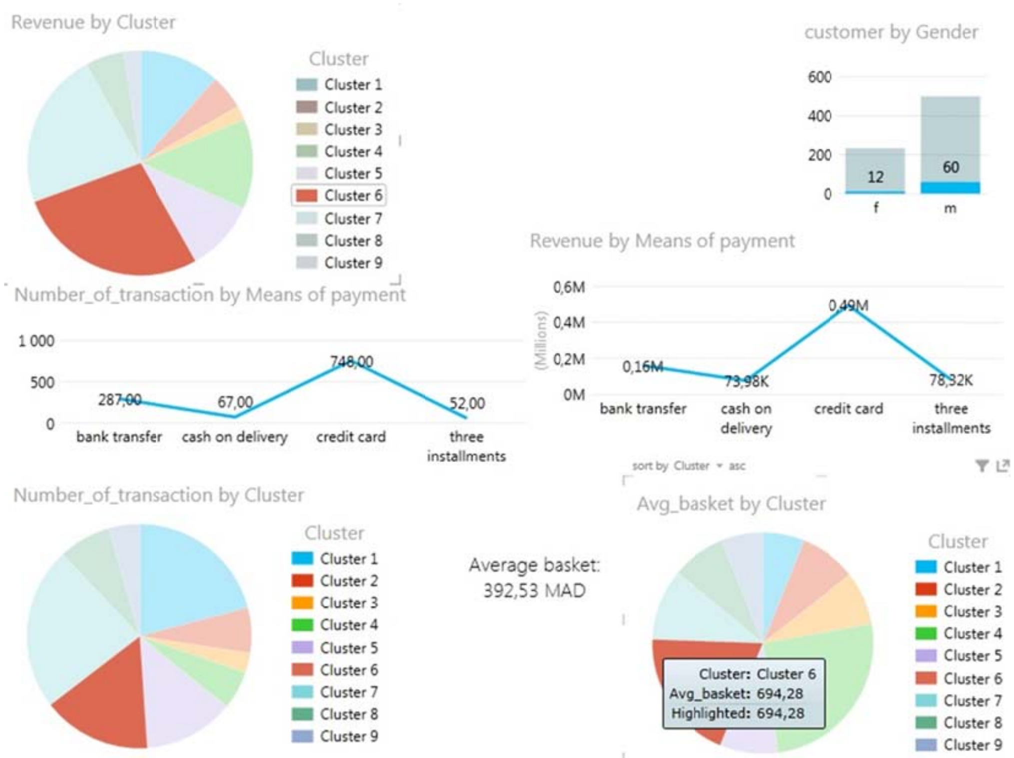

Fig. 6 An overview on the performance of the Cluster 6

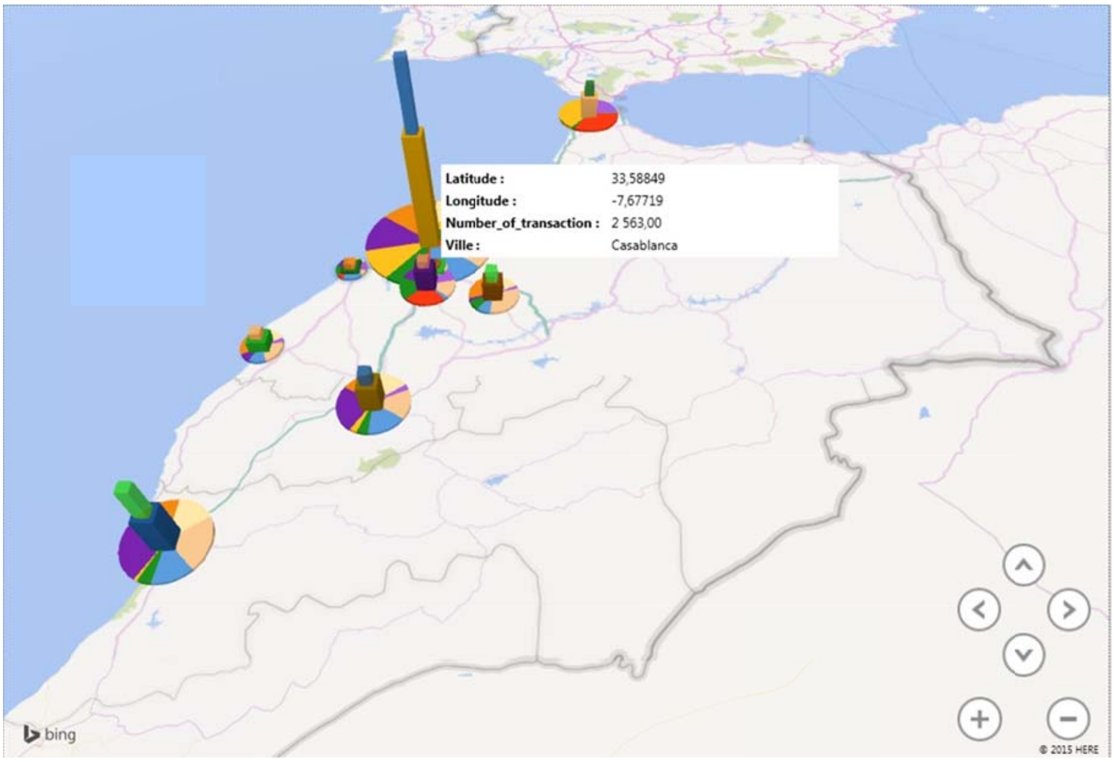

Fig. 7 Number of transactions and revenue by city

In terms of revenue, Clusters 6 and 7 represent the clusters that generate the biggest revenue 28% (MAD 801 203.11), 23% (MAD 659 420.98), respectively. In terms of the number of transactions, Clusters 1, 7, 6 and 5 represent the clusters that have recorded the highest number of transactions.

The mode of payment which is widely used by the customers remains the mode "pay on delivery". It represents 38,91% of the transactions (2885), that is to say, MAD 1,08 million. This principle consists in paying at the time of delivery.

In the second position of the modes of payment comes payment via credit cards (VISA, Master CARD and the national mark cmi) which have recorded 2520 operations. That is to say, 33.99% of the overall online purchases with an amount of money that reaches MAD 889 076,20. The third mode of payment used by customers, in terms of importance, is payment via bank transfer, which has represented 18,67% of the transactions (1384), that is to say, MAD 459 460,42. This system of payment continues to develop since the majority of Moroccan banks have set up web interfaces for customers' accounts.

As for the payment mode "pay via three installments), it comes in the fourth position with only 8,43% of the transactions (625), that is to say, MAD 478 007,97.

Payment via this mode remains the weakest since it has been recently proposed by the company. Also, this mode of payment is reserved exclusively for all the items whose price exceeds MAD 1500. This principle consists in paying the item in 3 months.

V. Conclusion

This study aims at proposing a methodology as regards the segmentation of customers in the field of e-commerce, by combining LRFM model and data mining techniques (Clustering analysis). In this study, the use of SOM method has indicated that the number 9 is more likely to be the best number of clusters. Then, the number K of K-means method is fixed at 9 in order to segment 730 customers on 9 clusters in terms of their L, R, F, and M values.

After reclassifying all the customers, we have used Power BI tool so as to produce interactive dashboards to analyse the performances of each type of cluster, and hence, assist decision-makers to take good decisions.

The results have demonstrated that group 6 is the most important group, as the average values of L, R, F, and M are superior to the overall average value. The analysis of clusters has also shown that the mode of payment is definitely a key indicator of a new index, allowing to evaluate the level of customers' trust unto the merchant site of the company. Finally, after analysing each cluster, we have provided certain strategies which might be taken into account to promote the relationship between the online store and its customers.

Acknowledgment

We would like to thank the Interbank Electronic banking Centre (IEBC) which provided us the state of e-commerce in Morocco in figures, in the first quarter of 2015, and the Moroccan online store specialized in electronics, fashion, home appliances, and children's items for providing us with the data.

REFERENCES

[1] Interbank Electronic banking Centre, Morocco, "Activité monétique 1er trimestre 2015 au Maroc", https://www.cmi.co.ma/PDF/Mon%C3%A9tique%20Marocaine%20au%2031%20Mars%202015.pdf
[2] G. C O'Connor, B O'Keefe, Viewing the web as a marketplace: the case of small companies, Decision Support Systems, vol. 21(3), 1997, pp. 171–183.
[3] H. H. Wu, S. Y. Lin, C. W. Liu, Analyzing Patients' Values by Applying Cluster Analysis and LRFM Model in a Pediatric Dental Clinic in Taiwan, Hindawi Publishing Corporation The Scientific World Journal, Vol 2014, Article ID 685495, 2014, 7 pages.
[4] C. H. Cheng, Y. S. Chen, Classifying the segmentation of customer value via RFM model and RS theory, Expert Systems with Applications, Vol.36, N.3, 2009, pp.4176–4184.
[5] S. Irvin, Using lifetime value analysis for selecting new customers, Credit World, Vol.82, No.3, 1994, pp. 37-40.
[6] J. T. Wei, S. Y. Lin, C. C. Weng, H. H. Wu, A case study of applying LRFM model in market segmentation of a children's dental clinic. Expert Systems with Applications Vol.39, No.5, 2012, pp. 5529–5533.
[7] H. H. Chang, S. F. Tsay, Integrating of SOM and K-man in data mining clustering: An empirical study of CRM and profitability evaluation, Journal of Information Management, Vol.11, No.4, 2004, pp. 161–203.
[8] W. J. Reinartz, V. Kumar, On the profitability of long-life customers in a noncontractual setting: An empirical investigation and implications for marketing. Journal of Marketing, Vol.64, No.4, 2000, pp. 17–35.
[9] A. M. Hughes, Boosting response with RFM. Marketing Tools, Vol.3, No.3, 1996, pp. 4-7.
[10] G. M. Marakas, Decision Support Systems in the 21st Century, Second Edition. Prentice Hall, Upper Saddle River, NJ, 2003.
[11] A. X. Yang, How to develop new approaches to RFM segmentation, Journal of Targeting, Measurement and Analysis for Marketing, Vol.13, No.1, 2004, pp. 50-60.
[12] S. Golsefid, M.Ghazanfari, S. Alizadeh,'Customer Segmentation in Foreign Trade based on Clustering Algorithms Case Study: Trade Promotion Organization of Iran'. World Academy of Science, Engineering and Technology, International Science Index 4, International Journal of Mechanical, Aerospace, Industrial, Mechatronic and Manufacturing Engineering, Vol.1, No.4, 2007, pp. 230 - 236.
[13] C. H. Wang, Apply robust segmentation to the service industry using kernel induced fuzzy clustering techniques, Expert Systems with Applications, Vol.37, No.12, 2010, pp. 8395-8400.
[14] J. T. Wei, S. Y. Lin, and H. H. Wu, A review of the application RFM model, African Journal of Business Management, Vol.4, No.19, 2010, pp. 4199–4206.
[15] A. M. Hughes, Strategic database marketing. Chicago, Probus Publishing Company, 1994.
[16] R. Kahan, Using database marketing techniques to enhance your one-to-one marketing initiatives, Journal of Consumer Marketing, Vol.15, No.5, 1998, pp. 491-493.
[17] E. C. Chang, H. C. Huang, H. H Wu, Using K-means method and spectral clustering technique in an outfitter's value analysis, Quality & Quantity, Vol.44, No.4, 2010, pp. 807–815.
[18] S. M. S. Hosseini, A. Maleki, M. R. Gholamian, 2010, Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty, Journal of Expert Systems with Applications, Vol.37, No.7, 2010, pp. 5259–5264.
[19] I. C. Yeh, K. J. Yang, T. M. Ting, Knowledge discovery on RFM model using Bernoulli sequence, Expert Systems with Applications, Vol.36, No.3, 2009, pp. 5866–5871.
[20] H. C. Chang, H. P. Tsai, Group RFM analysis as a novel framework to discover better customer consumption behavior, Expert Systems with Applications, Vol.38, No.12, 2011, pp.14499–14513.
[21] L. H. Li, F. M. Lee, W. J. Liu, The timely product recommendation based on RFM method, Proceedings of the International Conference on Business and Information, 2006, Singapore.
[22] S. Chow, R. Holden, Toward an understanding of loyalty: The moderating role of trust, Journal of Management issues, Vol.9, No.3, 1997, pp. 275–298.
[23] S. Huang, E. C. Chang, H. H. Wu, A case study of applying data mining techniques in an outfitter's customer value analysis, Expert Systems with Applications, Vol.36, No.3, 2009, pp 5909–5915.
[24] J. T. Wei, S. Y. Lin, C. C. Weng, H. H. Wu, Customer relationship management in the hairdressing industry: An application of data mining techniques, Expert Systems with Applications, Vol.40, No.18, 2013, pp. 7513–7518.
[25] S. Wang, Cluster analysis using a validated self-organizing method: Cases of problem identification, Intelligent Systems in Accounting, Finance and Management, Vol.10, No.2, 2001, pp. 127–138.
[26] K. E. Fish, P. Ruby, An artificial intelligence foreign market screening method for small businesses, International Journal of Entrepreneurship, Vol.13, 2009, pp. 65–81.
[27] D. Ordonez, C. Dafonte, M. Manteiga, B. Arcay, Hierarchical Clustering Analysis with SOM Networks, World Academy of Science, Engineering and Technology, International Science Index 45, Vol.4, No.9, 2010, pp. 213 - 219.
[28] L. Churilov, A. Bagirov, D. Schwartz, K. Smith, D. Michael, Data mining with combined use of optimization techniques and self-organizing maps for improving risk grouping rules: Application to prostate cancer patients, Journal of Management Information Systems, Vol.21, No.4, 2005, pp. 85–100.
[29] A. Sindhuja, V. Sadasivam, Automatic Detection of Breast Tumors in Sonoelastographic Images Using DWT, World Academy of Science, Engineering and Technology, International Science Index 81, International Journal of Biological, Biomolecular, Agricultural, Food and Biotechnological Engineering, Vol.7, No.9, 2013, pp. 590-596.
[30] D. Birant, Data Mining Using RFM Analysis, Knowledge-Oriented Applications in Data Mining, ISBN: 978-953-307-154-1, 2011, InTech, DOI: 10.5772/13683, Available from: http://www.intechopen.com/books/knowledge-oriented-applications-in-data-mining/data-mining-using-rfm-analysis
[31] G. Punj, D. W. Steward, (1983), Cluster analysis in marketing research: review and suggestions for applications, Journal of Marketing Research, Vol.20, No.2, 1983, pp. 134-148.
[32] R. J. Kuo, L. M. Ho, C. M. Hu, Integration of self-organizing feature map and K-means algorithm for market segmentation, Computers & Operations Research, Vol.29, No.11, 2002, pp. 1475-1493.
[33] J. Han, M. Kamber, Data Mining: Concepts and Techniques, CA: Morgan Kaufmann, San Francisco, 2006.
[34] P. N. Tan, M. Steinbach, V. Kumar, Introduction to data mining, Pearson education, 2005.
[35] O. Krivobokova, 'Evaluating Customer Satisfaction as an Aspect of Quality Management'. World Academy of Science, Engineering and Technology, International Science Index 29, Vol.3, No5, 2009, pp. 482 - 485.
[36] R. M. Schindler, T. M. Kibarian, Increased Consumer Sales Response through Use of 99 Ending Prices, Journal of Retailing, Vol.72, No.2, 1996, pp. 187-199.
[37] N. GUEGUEN, 100 petites expériences en psychologie du consommateur pour mieux comprendre comment on vous influence, Paris: Dunod, 2005.
[38] H. Isaac, P. Volle, E-Commerce De la stratégie à la mise en œuvre opérationnelle, Pearson Education, 2008.
[39] S. Bellman, G. Lohse, E. Johnson, Predictors of online buying behavior, Communications of the ACM, Vol.42, Vo.12, 2000, pp. 32-38.