JOURNAL OF APPLIED INTELLIGENT SYSTEMS & INFORMATION SCIENCES
VOL. 1, ISSUE. 1, PP. 35-43, FEBRAURY 2020.
Available at: **www.journal.research.fanap.com**
DOI: 10.22034/jaisis.2020.102488

USING RFM MODEL AND MARKET BASKET ANALYSIS FOR SEGMENTING CUSTOMERS AND ASSIGNING MARKETING STRATEGIES TO RESULTED SEGMENTS

Mohsen Maraghi[1,*], Mohammad Amin Adibi[2] , Esmaeil Mehdizadeh[3]
[1] *Faculty of Industrial and Mechanical Engineering, Islamic Azad University, Qazvin Branch, Qazvin, Iran; mohsen.maraghi@gmail.com*
[2] *Faculty of Industrial and Mechanical Engineering, Islamic Azad University, Qazvin Branch, Qazvin, Iran.*
[3] *Faculty of Industrial and Mechanical Engineering, Islamic Azad University, Qazvin Branch, Qazvin, Iran.*

ABSTRACT

Customer relationship management (CRM) in supermarkets is willing to interact with customers appropriately with the aim of making strong relationship and resultantly gaining maximum profits. Customers consist of various groups of people and have different needs, styles and expectations. Marketing management of a supermarket segments customers to respond their different demands correctly. Another important concern of supermarket managers is to detect profitable customers. These customers supply main profit of the company and keeping these profits guarantees existence of the supermarket. This research presents a model completing CRM process from understanding customers to assigning marketing strategies. Profitable customers will be distinced as a result of correct understanding of all customers. Present research is comprised of two phases. At phase one, dataset with recency, frequency and monetary (RFM) measures is constructed and clustered using *K*-means algorithm. Six segments of customers are detected based on the results of clustering. All segments are comprehensively analyzed and marketing strategies for them are described in phase two. Transactions of every segment of customers are separated and association rules are extracted using market basket analysis and Apriori algorithm. Consequent and also antecedent product items are proposed to customers who purchased antecedent product iems. So, dedicated marketing proposals are developed for some special customers.

***KEYWORDS*:** RFM Model, Segmentation, Association Rules, *K*-means, Market Basket Analysis, Machine Learning.

1. INTRODUCTION

Today, many supermarkets and hypermarkets have emerged all over the cities, where people go in order to meet their everyday needs. Chain stores are formed in larger sizes and provide all of the families' needed goods with better and typically different discounting policies. Most of these stores are equipped with a customer club system and record their customers' information. Customers in the stores have different needs. They come from different social classes and vary in many characteristics, such as age, gender, place of residence, education, income, etc. Identifying different customer groups (valuable, regular, new, etc.) and determining the appropriate marketing approach for each group is an important issue in each store. Customer relationship management (CRM) section in stores segments customers in order to understand them better and increase sales and the effectiveness of marketing proposals. The more carefully customer segmentation is conducted, the more effective the assigned marketing strategies are.

Researches have shown that 15% of customers provide 45% of revenue and 70% of company profits (Ivanovic et al., 2011). On the other hand, customer loyalty and profitability are obviously correlated (Payne et al., 1999). Therefore, the identification of profitable customers and retaining them is an important concern for

[*] *Corresponding Author*

RECEIVED: 21 DECEMBER, 2019; ACCEPTED: 1 FEBRUARY, 2020; PUBLISHED ONLINE: 3 FEBRUARY, 2020

any business enterprise. After segmenting customers, marketing strategies such as special discount offers, etc. can be applied to the resultant segments and especially to profitable customers using Apriori algorithm. The obtained association rules are used to cross-sell products to customers who have done shopping in the past. In addition, the products that are related and usually sold together are placed on the baskets next to each other so that when the customer purchases one s/he faces the others.

Recency, frequency and monetary (RFM) model is one of the most commonly used techniques in recent years for segmenting customers. This model uses three measures including recency, frequency and monetary value of purchases. RFM model was developed by Hughes (1994) and is widely believed to be the most powerful and, at the same time, the easiest technique for generating knowledge from CRM data (Khajvand et al., 2011; Kahan, 1998; McCarty & Hastak, 2007). The RFM model is valuable in predicting the response and can greatly increase the companys' profits in a short time (Baecke & Van den poel, 2011). The RFM model is also effective in identifying valuable customers (Frederick, 1997). The algorithms presented in this study are implemented on Ta Fang dataset (a store in China) storing purchase transactions and the results are analyzed and discussed (Yu et al., 2016; Rpubs, 2017).

The contribution of this research can be discussed in two aspects. As far as the as far as the authors are aware, the integrated CRM model in the form of a combination of RFM analysis and clustering and market basket analysis is a new issue studied here. This model completes the process of CRM from customer understanding to assigning their marketing strategy. This hybrid model identifies customers who are likely to do shopping in the future at first and then identifies the products that will be purchased by these customers. Secondly, in order to optimize marketing proposals and mine higher-quality association rules (greater support and confidence), ABC analysis of product inventory management is inspired, which is also a new issue. Therefore, a two-way optimization is designed and implemented to identify the more desirable customers (segments of customers) and the more desirable products (to be proposed to each segment).

In next section, previous researches on the proposed framework are elaborated. The research methods and executive phases are explained in Section three. Followingly, the results of implementing RFM model and clustering at phase one and defining marketing proposals at phase two are presented. The obtained segments of customers are then comprehensively analyzed and marketing strategies in each segment including dedicated marketing proposals are discussed in last section.

2. Research Literature

Miglautsch (2000) used different weights for recency, frequency, and monetary (RFM) measures for the first time. Wang et al. (2004) developed the problem of mining association rules from transactions of a store and assigned different weights to items to express their importance. Liu & Shih (2005) combined the RFM model and the Customer Lifetime Value (CLV) value to analyze their customers buying behavior. In their research, Clustering was done with weighted RFM criteria (WRFM). Sohrabi & Khanlari (2007) suggested the use of the CLV concept to measure the profitability of a bank's customers. The customer lifetime value can be defined as "the present value of the total profit to be obtained from the customer in the future" (Gupta & Lehmann, 2003). Shim et al. (2012) presented a model for classifying customers into VIP and non-VIP classes using RFM criteria. They used artificial neural networks, decision tree and classification for this purpose. Then, association rules and sequence patterns were discovered in VIP class transactions and were used to determine CRM strategies. Hu & Yeh (2014) presented an efficient algorithm for discovering RFM patterns in customer-unknown datasets. Instead of measuring the value of the patterns from the perspective of customers, they rank them only with RFM features. These RFM patterns were approxmitations of the customers' RFM patterns. Dursun & Caber (2016) used RFM model in the CRM field. They clustered them to identify the profitable customers of a hotel in Antalya. *R*, *F*, and *M* criteria were determined for hotel customers and then clustering was performed using SOM and *K*-Means algorithms. Customers who scored high on all three criteria were identified as profitable customers. Some clusters had a small number of members which made it difficult to analyze them.

Weng (2017) used the RFM model and weighted transactions of a store with the *F* and *M* criteria. In classic Apriori algorithm, the weight (importance) of all transactions are the same, and does not affect its approach to know which customer owns transaction. Weng devised a new algorithm for mining frequent itemsets in FM-weighted transactions, and showed that it works more successfully than classic algorithm in predicting sales revenue. Christy et al. (2018) presented a new approach for segmenting customers with RFM criteria. They initially performed RFM analysis on purchases transactions and then clustered with well-known algorithms such as *K*-means and Fuzzy *C*-means. They also changed the definition of RFM variables. Recency variable (*R*) was defined the number of days between two purchases of a customer. Their idea was to put the median value of each variable *R*, *F*, and *M* as the initial center of the *K*-means algorithm. Stormi et al. (2019) examined the functionality of recency, frequency and monetary variables of RFM analysis in customer segmentation of product-oriented services in the industry of the original equipment manufacturers. Tahanisaz & Shokuhyar (2020) employed a multi-method approach to evaluate passenger satisfaction regarding service quality on the Iranian airlines. They used ICF model for clustering which is based on RFM model. Flight intention, cabin class and frequency of flights were three variables considered for clustering. They could propose some managerial insights and marketing strategies based on the results. Zhou et al. (2020) combined RFM model with sparse *K*-means clustering algorithm of Witten and Witten & Tibshirani (2010) to invent a novel methodology for market segmentation. They showed that their methodology would produce robust results in dealing with a large high-dimentional and sparse customers' dataset. They identified eight segments of customers and made a demoghraphical analysis to have a better understanding of them.

This study firstly, combines RFM model and simple *K*-means clustering for market segmentation and secondly, performs market basket analysis to define marketing strategy for every segment. Zhou et al. (2020) figured out the resultant segments by demographical characteristics and produced general marketing proposals on that basis. This study analyzes the obtained segments by recency, frequency and monetary criteria and produces both general and dedicated marketing proposals based on the past purchase transactions of customers. The obtained association rules would help decision makers to improve marketing effectiveness and thus to gain more profit along with higher customers' satisfaction.

3. Methodology

1. Understanding the problem and data: At this stage, research problem is explained and variables are defined. Types of data stored in dataset including binary, continuous, categorical, etc. are determined and coding and features are defined.
2. Preparing data: Incomplete, noisy and inconsistent data are modified and data are ordered in a standard format ready for processing in data mining software.
3. Selecting model: Depending on the type of problem and the data collected, the appropriate technique and its optimal settings are determined.
4. Analyzing findings and Interpreting the model: The results are analyzed and interpreted for correct decision making based on research problem.

This research consists of two executive phases. The first phase is dedicated to customer segmentation using *K*-means clustering technique and RFM model. Purchase transactions in the primary dataset are scored with RFM measures and quarters are defined. Then, the new dataset is clustered using *K*-means and it is analyzed. Segments of customers are now specified.

The second phase is the stage of defining the marketing strategy for each group (segment) of customers. The primary dataset is transformed to employ binary variables which are necessary for running Apriori algorithm. To get better results, group *C* products are deleted which will be fully discussed in "Results" section. Then, the obtained segments of customers in first phase are used and association rules are mined for each segment separately by the means of market basket analysis and Apriori algorithm. The stages of this research are shown in Fig. 1.

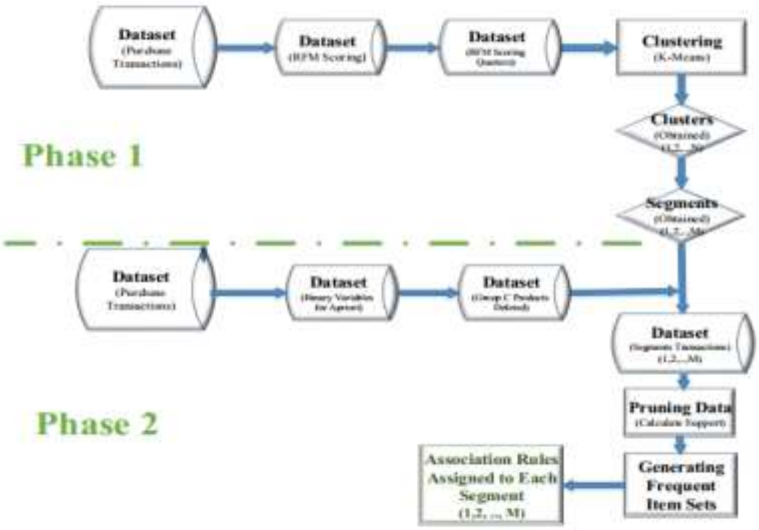

Fig. 1. The stages of research process

One of the most practical techniques for customer segmentation is customer value analysis via RFM model. The model was first introduced by Hughes in 1994. Using this technique, we can identify the most profitable customers and divide them into different segments of platinum, gold, silver, bronze and so on. As a result, we will be able to set a specific and customized pricing and marketing strategy for every segment of customers. The recency (*R*), frequency (*F*) and monetary (*M*) variables for each customer are defined as follows:

(Recency) *R*: Number of days since last purchase.

(Frequency) *F*: The total number of customer purchases in the time period.

(Monetary) *M*: The average cost of money spent on purchases.

In order to reduce the correlation between *F* and *M*, the monetary value criterion is considered as the average cost of purchases (rather than total cost) (Marcus, 1998). So, a dataset of customers with *R*, *F*, and *M* features could be created.

4. Results

The results of the research are presented in two phases: 1- Customer segmentation; 2- Defining marketing proposals for each segment.

4.1. *Customer segmentation (phase one)*

We use RFM model and K-means clustering to segment our customers. The recency (*R*), frequency (*F*) and monetary (*M*) variables for each customer ID need to be calculated and scaled in order to form the desirable RFM dataset. Then, we can run *K*-means and identify segments of customers through analyzing the resultant clusters.

4.1.1. *Understanding the problem and data*

The dataset contains 32266 customers, who own three features/variables *R*, *F*, and *M*. All three variables (*R*, *F*, and *M*) are quantitative and continuous. The variable *R* has been calculated from the date of last purchase of each customer. The used dataset contains the transactions during the four months from November 2000 to February 2001. The variable *R* for the last day of this period is one, and the day before the last is two and eventually for the beginning of the time period is 120. Therefore, the variable *R* has a negative nature. This means that a customer with a lower *R* value is more desired. The variable *F* is calculated from the number of purchases made by each customer in these four months. To calculate the variable *M*, the total cost of each customer purchases is divided by the number of times it was purchased. The statistical characteristics of these variables are given in Table 1:

Table 1. Statistical characteristics of variables *R*, *F* and *M*

Variable	Min	Max	Mean	Standard devision
Recency (*R*)	1	120	83.57	33.66
Frequency (*F*)	1	86	3.71	4.84
Monetary (*M*)	8	43917	1033.57	1174.59

4.1.2. Preparing data

The used dataset does not have missing data. There were a number of inconsistent data meaning some products with the same Product ID had different but close prices. These data are replaced by the mean value. Nothing is done for noisy data because the used model is not sensitive to them.

4.1.3. Selecting model

The model presented is implemented in SPSS Modeler software. Since these variables have a scale, they have an undesirable effect on clustering results; therefore, we need to scale them out or scale them in the same way. Distributions of the variables *R*, *F*, and *M* are quartered and scored from one to four according to their class data. The obtained scores are shown in Table 2.

Table 2. Scaling the variables *R*, *F* and *M*

Scores	Variable *R*	Variable *F*	Variable *M*
1	[59,120]	1	[8,366.3]
2	[26,59)	2	(366.3,710.1]
3	[9,26)	3,4	(710.1,1306]
4	[1,9)	[5,86]	(1306,43917]

Despite *F* and *M* varaiables, lower values of *R* variable gained higher scores due to its negative nature. Now clustering could be performed by the resultant recency score, frequency score and monetary score variables. These three variables are all positive and their higher values are more desirable. An appropriate RFM dataset is now available for clustering. We use the *k*-means algorithm for clustering. In this algorithm, the number of clusters as the initial parameter must be determined. Based on previous researches, the number of clusters is considered to be eight ($2 \times 2 \times 2 = 8$) according to the fact that each of the variables *R*, *F*, and *M* can be higher or lower than their mean (Liu & Shih, 2005; Sohrabi & Khanlari, 2007; Dursun & Caber, 2016). To be more confident in this research, we use self organizing maps (SOM). SOM is a type of artificial neural network (ANN) that maps a continuous space of high-dimensional input variables into an output discrete space, called a map, with lower dimensions (usually two-dimensional) through an unsupervised learning process. The output map is also known as Kohonen map in respect to its inventor. In this research, we use the Kohonen node in the SPSS modeler software to implement this technique. The average Silhouette refers to a method for measuring the quality of clustering. It is a measure of how similar an object is to its own cluster (cohesion) compared to other clusters (separation). The Silhouette ranges from -1 to $+1$, where higher values indicate better quality of clustering (Rousseeuw, 1987). In this problem, the highest average Silhouette obtained score is 0.4 for eight clusters. Therefore, eight clusters are approved as the initial parameter of the *K*-means algorithm.

4.1.4. Analyzing findings and interpreting the model

After performing *K*-means algorithm, the value of average Silhouette = 0.3 was obtained, which falls within acceptable limits. The obtained cluster centers, the percentage of members of each cluster and the RFM analysis are presented in Table 3. In the RFM analysis column, for variables holding a value which is higher than mean value, the High sign (*H*) is marked and for variables holding a value which is lower than mean value, the Low sign (*L*) is marked. As an example, the cluster whose RFM analysis is HLH means that whose members gained higher than mean values in *R* and *M* variables and lower than mean value in *F* variable.

Table 3. Obtained clusters (*K*-means algorithm)

Clusters	Number of members	Percentage of members	Recency score	Frequency score	Monetary score	RFM analysis
Cluster 1	7402	22.9 %	1	1	1	LLL
Cluster 2	3177	9.8 %	4	2	3	HLH
Cluster 3	3425	10.6 %	3	4	2	HHL
Cluster 4	4477	13.9 %	4	4	2	HHL
Cluster 5	4749	14.7 %	2	2	4	LLH
Cluster 6	3267	10.1 %	3	3	3	HHH
Cluster 7	3703	11.5 %	2	1	4	LLH
Cluster 8	2066	6.4 %	2	3	4	LHH

4.2. *Defining marketing proposals for each segment (Phase II)*

We now consider the transactional dataset. There are 119578 rows of transactions in this dataset. The total number of products is 2012. We want to mine patterns in purchasing products by market basket analysis.

4.2.1. *Understanding the problem and data*

The issue is composed of 119578 transactions that occurred in 2012 different products. The problem variables are all of binary types; if there is a product in a transaction, it is number one, otherwise it is number zero.

4.2.2. *Preparing data*

The dataset does not conatin missing, noisy and inconsistent data. The dataset is inputed in SQL server to form appropriate format for entering the SPSS modeler.

4.2.3. *Selecting the model*

The total four-month customer transactions equals to 119578. We want to use the Apriori algorithm to mine frequent itemsets and association rules. The total number of products is 2012. Different prices and sales rates have made different values of products for the store. We use ABC analysis of inventory management in order to improve the efficiency and quality of the association rules. We keep "very important" (*A*) and "important" (*B*) products and eliminate "not important" (*C*) products. The value of the products in ABC analysis is defined as follows:

$$\text{Product price} \times \text{Product demand in period} = \text{Product value in period} \quad (1)$$

Types of product items are shown in Fig. 2. Sum of items in group *A* and group *B* is 805, accounting for 95% of the total value of the products. The remaining 1207 group *C* products are deleted from the dataset and then Apriori algorithm is run. After deleting these products, the total number of transactions reaches from 119578 to 118574. By deleting 60% of the total number of products, only less than one percent of dataset transactions is eliminated.

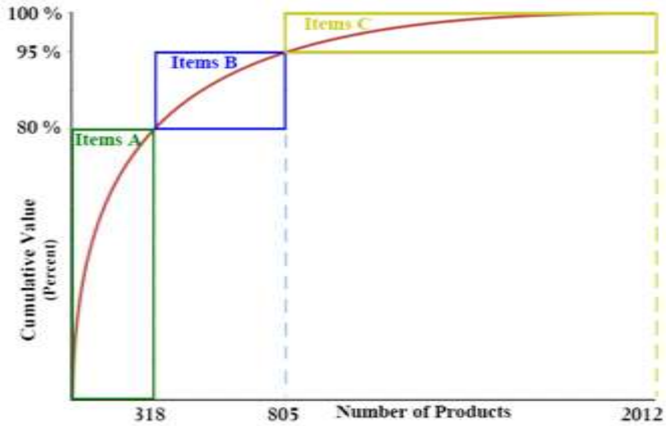

Fig. 2. ABC analysis on dataset products

So the new dataset specifications changes according to Table 4.

Table 4. New dataset specifications

Dataset name	Total number of products	Number of transactions
Ta Feng dataset	805	118574

In order to understand whether deletion of group *C* products has improved the results of the Apriori algorithm, we run the algorithm in two modes: before and after removal of group *C* products. The result of running the Apriori algorithm in SPSS modeler 18 is shown in Table 5. The computer system used has a Core *i*-7 processor and 8 MB of memory.

Table 5. Running apriori algorithm before and after the removal of group *C* products

Dataset with 119578 transactions and 2012 products									
Number of association rules	**Support**			**Confidence**			**Min lift**	**Max lift**	**Running time (min)**
	Min	**Avg**	**Max**	**Min**	**Avg**	**Max**			
79	0.1%	0.15%	0.51%	60%	63.4%	75.97%	4.89%	15.17%	2:56
Dataset with 118574 Transactions and 805 products (Group *C* Products Removed)									
Number of association rules	**Support**			**Confidence**			**Min lift**	**Max lift**	**Running time (min)**
	Min	**Avg**	**Max**	**Min**	**Avg**	**Max**			
82	0.1%	0.15%	0.52%	60%	64.11%	75.98%	4.85%	17.05%	1:10

After examining the table it becomes evident that the results of the algorithm have been improved by removing the group *C* products. Except the min lift criterion, seven other criteria show improvements. (Min support and min confidence are set in advance). Therefore; the removal of group *C* products is approved. In the first phase of this research, customers were divided into six segments based on recency, frequency and monetary criteria. Since different groups of customers have different characteristics and desires, the transactional dataset is divided into six sections (based on six segments of customers) to make the proposals from the association rules more effective. Then, association rules are mined.

4.2.4. Analyzing the findings and interpreting the model

We implemented the Apriori algorithm on six segments of customers. In all runs, the minimum support is set 0.1% and the minimum confidence is set 70%. Table 6 shows the number of obtained association rules in every segment of customers.

Table 6. Result of apriori algorithm on six segments of customers

Segments	Number of transaction	Number of features (Products)	Number of obtained association rules	Running time
One	8213	785	575	4”
Two	16866	796	46	9”
Three	59928	802	11	35”
Four	16480	800	267	10”
Five	10211	789	258	6”
Six	6876	792	1499	4”
Total	118574	-	2656	1’8”

Using the obtained association rules for each segment of customers, we provide dedicated marketing proposals for them. An association rule is defined as follows (Agrawal et al., 1993):

$$X \rightarrow Y \tag{2}$$

$I = \{i_1, i_2, ..., i_n\}$

A set of n binary variables showing presence/ absence of items in shopping cart.

$D = \{t_1, t_2, ..., t_m\}$

A set of m transactions making the dataset.

$X \cap Y = \varphi \ \& \ X, Y \subset I$

Applying the obtained association rules, we propose products of set Y and also set X to customers who have bought products of set X in the past. It is clear that marketing proposals are only available to customers who have purchased the products of set X; therefore, a specific proposal may not be available to a number of customers.

5. Discussion and Conclusion

In this research, customers were scored using the RFM model and the quarters of recency (R), frequency (F), and monetary (M) criteria and then clustered using K-means algorithm. Based on previous researches, the number of clusters was considered eight. To be more confident, Kohonen networks was used to determine the optimal number of clusters and the highest Silhouette was achieved for eight clusters. After analyzing the clusters, it was revealed that six different segments of customers existed in the dataset.

In segment one of customers, the RFM analysis was LLL. These customers comprised 22.9% of the dataset and were named "lost customers". Due to resources constraints, no particular plan was considered for these customers. Segment two customers had the RFM analysis as HLH. These customers comprised 9.8% of the dataset and were known as "profitable new customers." Due to the new and high potential of these customers, the VIP membership of Shop Customer Club was determined as a marketing plan. Segment three customers had the RFM analysis as HHL. These customers comprised 24.5% of the dataset and were known as "economical loyal customers". Two approaches were proposed for the marketing of these customers. In the first approach, the regular membership of Shop Customer Club was proposed to them due to their loyalty; and in the second approach, no marketing plan was considered for them due to marketing costs and their low profitability.

The fourth segment of customers had the RFM analysis as LLH. These customers comprised 26.2% of the dataset and were known as "high potential customers". Due to significant amount of money spent by these customers, it is important to attract them. As a marketing plan, they were offered to be a VIP member of Shop Customer Club if they are not and to get the best discount offerings. The fifth segment of customers had the RFM analysis as HHH. These customers comprised 10.1% of the dataset and were known as "profitable (golden) customers". They were the best segment of the store's customers and retaining them is the first priority for the store management. Continuous constructive interaction and awarding on occasions (birthday, anniversary of marriage, etc.) were considered as marketing plans of these golden customers. The sixth segment of customers had the RFM analysis as LHH. These customers comprised 6.4% of the dataset and were known as "valueable customers at risk". These customers had a good history, but have recently changed their minds and were leaving the store. To re-enable these customers, sending personalized messages (with their names) and marketing proposals based on past purchases, as well as discounted offerings with a limited time considered as marketing plans.

Dedicated marketing proposals were set up for every segment of customers. In order to improve the quality of these proposals (higher support and confidence), ABC analysis was used. The 2012 products decreased to 805. Results showed improvements on the quality of marketing proposals. After removing group (C) products, the transactional dataset was split into six different sections (based on the obtained six segments of customers) and the Apriori algorithm was implemented on each one. In each segment of customers, association rules were considered and consequent and also antecedent products were proposed to customers who had purchased antecedent products. Therefore, a number of dedicated marketing proposals for those who had purchased antecedent products were developed.

The implications of applying the results of this research for stores include increasing profitability, brand promotion and enhancing customer satisfaction resulting from a comprehensive understanding of customers' needs and expectations and assigning appropriate marketing strategies to them. For future researches, the concept of customer lifetime value in the form of RFM can be used to rank different segments of customers. In addition, the framework of the proposed hybrid model can be applied to other contexts such as bank customers, insurance, and so on. Besides, considering different weights for *R*, *F* and *M* criteria and using AHP for setting them may produce more desirable results. Finally, the clustering quality (average Silhouette) of the present method can be increased using other clustering algorithms such as two-step clustering.

REFERENCES

Agrawal, R., Imieliński, T. & Swami, A. (1993). Mining association rules between sets of items in large databases. *In Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, Washington, USA.

Baecke, P., & Van den Poel, D. (2011). Data augmentation by predicting spending pleasure using commercially available external data. *Journal of Intelligent Information Systems*, 36(3), 367-383.

Christy, A.J., Umamakeswari, A., Priyatharsini, L. & Neyaa, A. (2018). RFM ranking-An effective approach to customer segmentation. *Journal of King Saud University-Computer and Information Sciences*, https://doi.org/10.1016/j.jksuci.2018.09.004.

Dursun, A., & Caber, M. (2016). Using data mining techniques for profiling profitable hotel customers: An application of RFM analysis. *Tourism Management Perspectives*, 18, 153-160.

Frederick, N. (1997*). The New Rules of Marketing: How to use one-to-one relationship marketing to be the leader in your industry*. New York: McGraw-Hills Companies Inc.

Gupta, S., & Lehmann, D.R. (2003). Customers as assets. *Journal of Interactive Marketing*, 17(1), 9-24.

Hu, Y.H., & Yeh, T.W. (2014). Discovering valuable frequent patterns based on RFM analysis without customer identification information. *Knowledge-Based Systems*, 61, 76-88.

Hughes A.M. (1994). *Strategic Dataset Marketing*. Chicago: Probus Publishing Company.

Ivanovic, S., Mikinac, K., & Perman, L. (2011). CRM development in hospitality companies for the purpose of increasing the competitiveness in the tourist market. *UTMS Journal of Economics*. 2(1), 59-68.

Kahan, R. (1998). Using database marketing techniques to enhance your one-to-one marketing initiatives. *Journal of Consumer Marketing*, 15(5), 491-493.

Khajvand, M., Zolfaghar, K., Ashoori, S., & Alizadeh, S. (2011). Estimating customer lifetime value based on RFM analysis of customer purchase behavior: Case study. *Procedia Computer Science*, 3, 57-63.

Liu, D.R., & Shih, Y.Y. (2005). Integrating AHP and data mining for product recommendation based on customer lifetime value. *Information & Management*, 42(3), 387-400.

McCarty, J.A., & Hastak, M. (2007). Segmentation approaches in data-mining: A comparison of RFM, CHAID, and logistic regression. *Journal of Business Research*, 60(6), 656-662.

Marcus, C. (1998). A practical yet meaningful approach to customer segmentation. *Journal of Consumer Marketing*, 15(5), 494-504.

Miglautsch, J.R. (2000). Thoughts on RFM scoring. *Journal of Database Marketing & Customer Strategy Management*, 8(1), 67-72.

Payne, A., Christopher, M., Clark, M., & Peck, H. (1999). *Relationship Marketing for Competitive Advantage*. Butterworth Heinemann: Oxford, UK.

Rousseeuw, P.J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53-65.

Shim, B., Choi, K., & Suh, Y. (2012). CRM strategies for a small-sized online shopping mall based on association rules and sequential patterns. *Expert Systems with Applications*, 39(9), 7736-7742.

Sohrabi, B., & Khanlari, A. (2007). Customer lifetime value (CLV) measurement based on RFM model. *Iranian Accounting & Auditing Review*, 14(7), 7-20.

Stormi, K., Lindholm, A., Laine, T., & Korhonen, T. (2019). RFM customer analysis for product-oriented services and service business development: an interventionist case study of two machinery manufacturers. *Journal of Management and Governance*, 1-31.

Tahanisaz, S. (2020). Evaluation of passenger satisfaction with service quality: A consecutive method applied to the airline industry. *Journal of Air Transport Management*, 83, 101764.

Wang, W., Yang, J., & Yu, P. (2004). WAR: Weighted association rules for item intensities. *Knowledge and Information Systems*, 6(2), 203-229.

Weng, C.H. (2017). Revenue prediction by mining frequent itemsets with customer analysis. *Engineering Applications of Artificial Intelligence*, 63, 85-97.

Witten, D.M., Tibshirani, R. (2010). A framework for feature selection in clustering. *Journal of the American Statistical Association*, 105(490), 713–726.

Yu, F., Liu, Q., Wu, S., Wang, L., & Tan, T. (2016). A dynamic recurrent model for next basket recommendation. *SIGIR '16: Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Pisa, Tuscany.

Zhou, J., Zhai, L., & Pantelous, A.A. (2020). Market segmentation using high-dimensional sparse consumers' data. *Expert Systems with Applications*, 145, 113136.