Customer Segmentation and Strategy Development Based on User Behavior Analysis, RFM Model and Data Mining Techniques: A Case Study

Conference Paper · October 2018

DOI: 10.1109/ICEBE.2018.00027

Mohammadreza Tavakoli
Forschungszentrum L3S

Mohammadreza Molavi Hajiagha
TIB – Leibniz Information Centre for Science and Technology

Vahid Masoumi
University of Tehran

Majid Eagle M
Babol Noshirvani University of Technology

Customer Segmentation and Strategy Development based on User Behavior Analysis, RFM model and Data Mining Techniques: A Case Study

Mohammadreza Tavakoli
dept. Data Science at Digikala
Tehran, Iran
mo.tavakoli@digikala.com

Mohammadreza Molavi
dept. Data Science at Digikala
Tehran, Iran
m.molavi@digikala.com

Vahid Masoumi
dept. Data Science at Digikala
Tehran, Iran
v.masoumi@digikala.com

Majid Mobini
dept. Data Science at Digikala
Tehran, Iran
m.mobini@digikala.com

Sadegh Etemad
dept. Data Science at Digikala
Tehran, Iran
sadegh.etemad@digikala.com

Rouhollah Rahmani
dept. Data Science at Digikala
Tehran, Iran
r.rahmani@digikala.com

***Abstract*—The RFM (Recency, Frequency and Monetary) model provides an effective analysis for decision makers in order to target their customers and develop appropriate marketing strategies according to their previous behaviors. Although the RFM model has been widely applied in various areas of marketing, its simplicity threatens its effectiveness since it does not consider the customers' relationship and changes in customers' behavior. In this paper, we propose an R+FM model which configures the segmentation according to the business changes and clusters customers using K-Means. We applied our model on Digikala company, the biggest E-Commerce in Middle East, and compared our model with the Digikala's previous RFM model which used Customer Quantile Method. Moreover, we built strategies for each segment and ran an SMS campaign according to those strategies. The results of the campaign showed that our Segmentation Model improved the number of purchase and average monetary of the baskets.**

***Index Terms*—CRM, Cutomer Segmentation, RFM, E-Commerce, K-Means**

I. Introduction

Nowadays, the relationship between companies and customers has become an undeniable element of businesses and thus, the existence of a mechanism to manage this relationship is essential. The controlling process of the interactions between organizations and customers is called Customer Relationship Management (CRM). Therefore, the concept of CRM includes a set of methods and strategies to develop long-term, profitable relationships with customers. Based on [14], the CRM is divided into four different dimensions: (1) Customer Identification, (2) Customer attraction, (3) Customer Retention and (4) Customer Development.

Customer Identification as the first step of CRM is extremely important since identifying customer properties helps companies select the proper strategies. Customer Segmentation is one the most significant part of Customer Identification and it includes division of the entire customers into several valuable segments. In other words, the customer segmentation process classifies the customers into several groups based on their purchase behavior, demographic and geographic information and also, psychographic attributes. One of the most commonly used methods for customer segmentation is the RFM analysis. The RFM analysis is known as a behavioral-based data mining technique which extracts customers' profiles by using their Recency, Frequency, and Monetary values [10], [12]. In RFM analysis, R shows the consumption interval, F provides the frequency of use, and M describes the monetary spending [9], [19].

In recent years, there are several researches which have used the RFM concept in order to build Customer Segmentation Model [7], [13], [18]. Moreover, some articles have tried to improve the RFM concept by adding extra features or applying data mining techniques [4], [9], [20]. In this paper, we proposed a novel RFM framework named R+FM for customer segmentation. To create our RFM model, we separated R from the two other features (F and M) since Recency shows only the time of the last purchase, but Frequency and Monetary indicate the loyalty of customers. Clustering method helped us assign customers to the segments which contain similar users with respect to their behavior. Also, we considered the relationship between Frequency and Monetary by using Frequency, Monetary and a linear combination of them as the features of our model. A significant point for our proposed model is that the Linear Combination of F and M helps managers create appropriate segments with respect to the similarity of users behavior. The proposed model has been evaluated based on real data from Digikala[1] company which is the fourth ranked of the most visited websites in Iran according to the Alexa Ranking. Based on our Exploratory Analysis and the need of business managers, we applied our segmentation model and we provided 3 segments of Active, Lapsing and Lapsed for our

[1] http://www.digikala.com/

Recency feature. Moreover, we applied K-Means and provided 4 segments of *High Value*, *Medium Value with High Monetary*, *Medium Value with High Frequency* and *Low Value* for the Active segment and 3 segments of *High Value*, *Medium Value* and *Low Value* for the Lapsing and Lapsed segments. We ran an SMS campaign for different segments of customers which were extracted from R+FM model, and the results showed that the proposed framework leads to increased purchase count and income compared with other Digikala campaigns.

The rest of this paper is organized as follows: The reasons that motivated us to propose the new framework is described in Section II. Section III provides the related works and in Section IV and V, the proposed RFM Scoring and Segmentation Model are presented, respectively. In Section VI, the appropriate plan for each segment is built based on the customer segmentation results and the evaluation and experimental results are presented in Section VII. The technologies that were used to implement this article are described in Section VIII. Finally, the conclusion and future work are explained and elaborated.

II. Motivation

In this section, we briefly provide some of the reasons that motivated the need for the presence of our R+FM model which brings valuable Customer Segmentation model for managers, specifically in E-Commerce.

- *User behaviors change all the time.* Customers can show various behaviors in different situations and their behaviors depend on many factors such as their demographic properties, the new trends in societies, the marketing strategies of companies and even governments' policies. For instance, in the case of E-Commerce, if a company adds new category to its Product Catalog, it may change the patterns of purchases. Accordingly, having a segmentation model which updates itself based on the new patterns of behavior is extremely useful for the managers.
- *RFM models need supervision for providing meaningful segments.* Most of the time, managers should do some manual activities in order to build appropriate segmentation model and therefore, these manual activities can cause deficiency in their customer segmentation. For instance, in the models which depend on Quantile Analysis, managers should combine some segments at the end of the process in order to have meaningful segments. In clustering based customer segmentation, combining Recency, Frequency and Monetary in a correct manner and finding the best features for having high quality clustering can be problematic for managers. Moreover, sometimes, managers need some parameter of their segmentation to be static for their comparison. For instance, they need fixed segments in order to be able to compare their segment properties over time in order to measure their success. Thus, building a customer segmentation model that pays attention to the important factors for the managers without any manual activity is important for the companies.
- *RFM Model should pay attention to the Customer Behavior similarity and also the relationship between Frequency and Monetary.* Existing RFM models only cluster customers with static threshold and do not pay attention to the shape of the data and similarity in the user behaviors. Therefore, the results cannot provide high quality segments for users. Moreover, the RFM models behave Recency, Frequency and Monetary as 3 separate features and do not consider their relationship as an important factor. Therefore, building segmentation model which pays attention to the relationships helps managers have greater insight into their segments.

III. Related Work

For the first time, the concept of RFM was proposed by Hughes in 1994 [11]. After defining RFM, several works used this concept. An initial study [21] determined that customers with higher R, F, and M scores are more probable to make a new purchase. Thus, several efforts have been done to segment customers based on RFM values.

For example, [1] have computed RFM values for bank customers and have used the K-Means clustering method for their segmentation. Huang et al. [7] have applied K-Means, Fuzzy C-means and Bagged Clustering algorithms on the RFM analysis of customers for an outfitter in Taipei. In [9], an expanded RFM model has been introduced by an additional parameter which is named Weight (W); they have also used R, F, M and W features in K-means clustering technique to segment customer of Sapco company in Iran. Namvar et al. [16] have proposed a new customer segmentation method based on RFM and demographic data. They have applied Self Organizing Map (SOM) to get demographic variables and used K-Means clustering algorithm to segment customers of an Iranian bank. The soft clustering method which uses a latent mixed-class membership clustering approach has been proposed in [22]. Actually, they have proposed Latent Dirichlet Allocation (LDA) based model to create the customer segmentation in E-Commerce area. Wei et al [20] have developed an extended RFM model, namely LRFM (Length RFM) by using self-organizing maps (SOM) technique for children's dental clinic in Taiwan. Dursun et al [6] have investigated the profiles of the profitable customers from the customers of the five-star hotel at Antalya in the content of CRM, by the use of RFM analysis and K-means algorithm. Recently, Aghdaie et al [2] have proposed a novel integrated Fuzzy Group Multiple Attribute Decision Making (FGMADM) and Fuzzy C-mean clustering (FCM) as a data mining tool for segmenting customers based of RFM model.

IV. Building RFM Scoring Model

To build a Customer Segmentation based on RFM model, managers need to create an RFM Scoring for their business. After the process of RFM Scoring, each of the users has meaningful values for their Recency, Monetary and Frequency. In this section, at first, we introduce the data set which has been used in our analysis and model building. After that, we define our scoring model and finally, we try to show some insights about our data and the applied RFM scoring model.

A. Data Collection

Digikala is the biggest E-Commerce and one of the significant technological companies in the middle east. Moreover, it has around 3 million customers (users with at least one purchase) and around 10 million purchases since its first day. We collected information about customers and their purchases between January 1st, 2014 to December 30th, 2017 from Digikala's database. The resulting dataset contains simple demographic information (age, gender) and purchase information (size of baskets, date of purchase) for all of the 3 million customers.

B. Finding Best Definition for Recency, Frequency and Monetary

We applied different existing definitions for Recency, Frequency and Monetary and tried to find the best one. Our process of finding the best definition for each of the R, F and M elements has been clarified as follows:

- *Recency:* We applied various definitions for R such as assigning score from 5 to 1 to the users who have had purchase in last 7 days, 30 days, 90 days, 365 days and more than 365 days, respectively. Moreover, we used scoring from 3 to 1 for the users who have done their last purchase in recent 6 months, recent 1 year and more that 1 year. As we could expect, these definitions could not pay attention to the business and the user behaviors since they are rigid and have low accuracy. Thus, we tried to define Recency in a way that managers easily score the customers based on the state of the business. Accordingly, we defined R as the number of days from the last purchase. This definition brings flexibility for decision makers since it shows the most accurate value for recency of users.
- *Frequency:* Again, we tried to use different definitions for Frequency. For each customer, we assigned the number of purchases from the first day of starting Digikala and the number of purchases from the first day of the current year as two possible values for Frequency. Although these definitions are reasonable and simple, they cannot consider the starting date of becoming customer and therefore, comparison between the customers with these definition is not fair. Accordingly, we tried to pay attention to the date of the first purchase for customers and thus, we set Frequency of each customers as the number of their purchases divided by the number of days from their first purchase. An important point for our Frequency calculation is that according to the importance of customers' recent behavior, we applied weighted scoring in a way that frequency of recent years has higher weight in comparison with the previous years. There are several researches such as [8] which have used (1) as *Exponential Decay Function* in order to apply weighting over time. Since we aimed at having much stronger effect on F value for recent purchases, we used the function to have weighted Frequency. We considered yearly time slices with using *t* parameter. For instance, parameter t will be set to 1 in current year and it will be set to 2 for the previous year. In our model, empirically, we set the *Decay Rate* (k) to *0.75* in order to weight each year around two times more than the previous year. Accordingly, Frequency for $User_i$ has been calculated as (2) where $Purchase_{i,t}$ is the purchase count of $User_i$ in period slice of *t*, and $\Delta_{i,t}$ is the length of the period slice *t* for $User_i$ which is calculated according to (3).

$$W_t = Exp(-k \times t) \tag{1}$$

$$F_i = \frac{\sum_{t=1}^{CurrentYear-StartYear} W_t \times (\frac{Purchase_{i,t}}{\Delta_{i,t}})}{\sum_{t=1}^{CurrentYear-StartYear} W_t} \tag{2}$$

$$\Delta_{i,t} = \begin{cases} 365 - (User_i\text{'s Day of Becoming Customer}) \\ \quad \textbf{if t = Period Slice of Becoming} \\ \quad\quad \textbf{Customer for user i} \\ \\ \text{Day of Year} \\ \quad \textbf{if t = Period Slice of Current Year} \\ \\ 365 \\ \quad \textbf{otherwise} \end{cases} \tag{3}$$

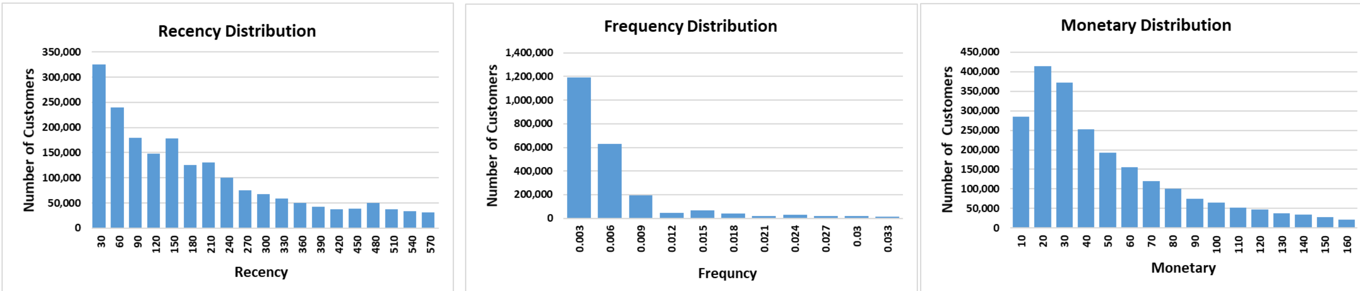
- *Monetary:* We applied two definitions for Monetary; one of them was sum of purchases and the second one was the average of purchases. Customers with higher number of purchases have greater scores in Frequency and if we use *sum of purchases* for Monetary, these customers will be encouraged twice which is not a proper decision for our model. Therefore, we used average of the purchases as Monetary for each customer.

C. Exploratory Data Analysis

In this part, we provide some figures and tables in order to create appropriate insights about our data and scoring model. Fig. 1 shows the histogram of Recency, Frequency and Monetary of customers in our dataset. The important point is that we have cut our plots in some points in order to have high quality histograms and therefore, we added Table I to show the detailed statistics for our plots. Moreover, we have added statistics of Purchase Count for our customers in Table I.
As it can be seen in the figures, we have Long Tail in all of our plots except for Monetary [3], [17]. The reason is that Digikala provides Free Delivery for purchases more than 20 USD and thus, in Monetary plot, the Long Tail starts from 20 USD.

V. Building Customer Segmentation Model

In this section, we explain our Customer Segmentation model. At first, we clarify our goals for building the model and afterward, we provide the process of building our proposed model.



(a) Recency (Days) (b) Frequency (Purchase Count / Days) (c) Monetary (USD)

Fig. 1: Histogram of Purchase Count, Recency, Frequency and Monetary in our Dataset

TABLE I: The Description of Recency, Frequency, Monetary and Purchase Count in our Dataset.

RFM and Purchases Element	Minimum	Maximum	Average	Standard Deviation
Recency (*Number of Days from Last Purchase*)	1	1,753	343	372
Weighted Frequency of Purchases	7e-5	11.55	0.007	0.02
Monetary (*Average of Purchase Spending in USD*)	1	208,106	76	287
Purchase Count	1	11,388	5.09	15.5

A. Model Goals

There are several goals which should be considered in order to build a useful Customer Segmentation model. Managers should pay attention to these goals in order to propose a model which can be widely applied in various parts of companies. We briefly describe these goals in Digikala company which provided an incentive for conducting the current study.

- *Managers' team needs to have static or semi-dynamic definition for R.* This goal helps managers have appropriate insight for their Active, Lapsing and Lapsed customers that is extremely significant for the business. Moreover, they can monitor their success or failure of a campaign or plan since according to the static definition of Recency, the segments before a campaign or plan can be compared with the segments after that campaign or plan.
- *Relationship between Frequency and Monetary should be considered.* The relationship between Frequency and Monetary is highly important in a successful Customer Segmentation model since they have several effects on each other which should be handled by decision makers. Therefore, in the case of low quality model, managers may merge or manipulate some segments to handle this relationship.
- *Changes in business and User Behavior should affect Customer Segmentation model.* As we said earlier, e-businesses, especially E-Commerce, are faced with several changes which have effects on Customer Segmentation models and therefore, the models need to be improved in order to handle the changes. Accordingly, if managers have a high quality Customer Segmentation model, they would not need to improve their model by a manual process.
- *Building Meaningful Segments.* In order to have effective segments which help managers build appropriate strategies, we need to have meaningful segments. The important point is that the mean of the segments should not be changed frequently, but the behavior of their users may change.

B. Customer Segmentation Model

1) Separating R from the other factors: To create a Customer Segmentation model which satisfies our goals, at first, we tried to give managers the ability of comparison over time by separating Recency from the two other factors (Frequency and Monetary) and having static segmentation for it. Moreover, since Recency only shows the time of the last purchase for each customer and does not show their loyalty, the type of Recency differs from Frequency and Monetary and our separation would make sense. Accordingly, this is the reason that we call our model R+FM. In order to find the best segments for Recency, we collected information from project managers and marketing team and understood that we need 3 segments of Active, Lapsing and Lapsed in that not only are the segments simple and usable, but also they help managers have different strategies for their segments. The reason is that there are three different segments for customers:

1) The ones who have purchased recently and therefore, they are active customers.
2) The ones who have not purchased recently, but they have had purchase in close past and thus, they are in the process of becoming churned.
3) The ones who have not purchased for a long period of time and therefore, they are churned customers.

These segments can cover the existing states for the customers in our company (and several other companies) and thus, managers can build all of their Recency-Related strategies according to these three simple and meaningful segments. In order to find the thresholds for our Recency, we used our quantile method plot (which divides our customers into 3 same-size segments) and managers' points of view and finally, set them to 90 and 365. Thus, a customer is Active if he has purchased in recent 90 days; he is Lapsing if his last purchase is between 90 and 360 days and he is Lapsed if he has not purchased in recent 365 days.

2) Feature Selection for K-Means: In the second step, we tried to segment our customers according to their Frequency and Monetary. We used clustering methods for building segments since they assign customers with similar behavior into one group. Moreover, as we said before, Frequency and Monetary are related to each other and if we find appropriate feature for the clustering step, we can consider this relationship.

To find the best features for our clustering step, we selected Frequency and Monetary as the two main features. Furthermore, we wanted to add a feature which contains both Frequency and Monetary. There are several cases in which researchers and managers have used combination of R, F and M as a 3-digit number [15]. Accordingly, we used a linear combination of F and M ($F \times W_F + M \times W_M$) in order to build our third feature for clustering in that not only is this feature simple and can be calculated fast, but also it is a meaningful feature which provides flexibility for establishing a balance between Frequency and Monetary. Accordingly, we use Frequency, Monetary and (F $\times$ W_F + M $\times$ W_M) as the three features for our clustering. An important point is that these features and our definition for Frequency and Monetary help us have a dynamic segmentation over time since in case of changes in user behaviors, new clusters are made and the best segments will be identified according to the new values of Frequency, Monetary and their linear combination.

3) Preprocessing & Model Tuning: In this part, we explain two steps which was needed in order to create our model: (1) Preprocessing step: In order to provide clean and high quality data and (2) Tuning step: In order to find the best parameters for our model. We provide the results of these two steps as follows:

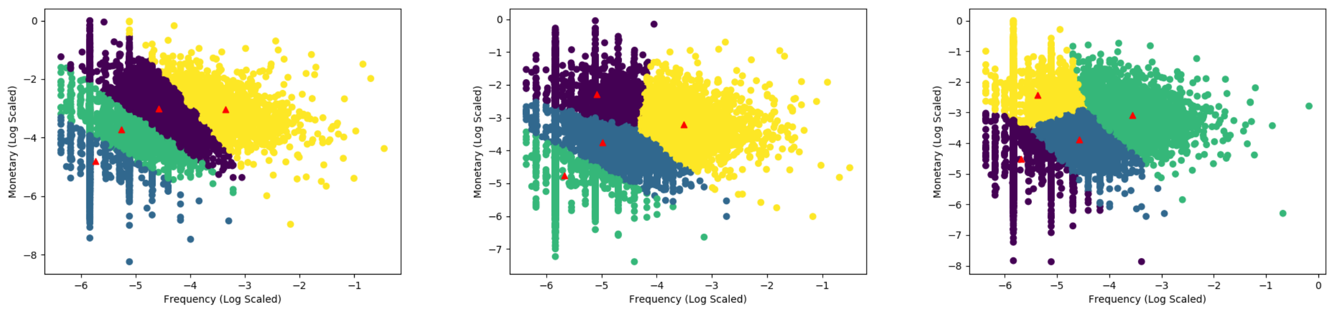
- *Removing the Outliers:* In almost all businesses, some customers show exceptional behavior. For instance, in Digikala, there are customers who buy each day and also there are other customers who have Monetary around 20,000 USD. In order to have high quality data analysis and apply machine learning methods, we had to remove these customers who might be buyers for other companies and should be behaved differently. An important point is that we used *Interquartile Ranges (IQR)* to remove the outliers. Therefore, we used (4) in order to find the threshold for upper bound and lower bound of our data in features of Frequency and Monetary.

$$\begin{cases} \text{IQR = Q3 - Q1} \\ \text{lower\_bound = Q1 - 1.5 } \times \text{ IQR} \\ \text{upper\_bound = Q3 + 1.5 } \times \text{ IQR} \end{cases} \tag{4}$$

 Accordingly, in each of the Frequency and Monetary, we removed the values below the *lower_bound* or above the *upper_bound*.

- *Scale the range of Frequency and Monetary to have appropriate data for Clustering Methods:* Since the range of values in our Frequency was completely different from Monetary values, we had to apply normalization step in our data of Frequency and Monetary. In order to do this step, we used Max-Min Scaling which helped us scale our data in an appropriate range for clustering.
- *Solving the Long Tail problem in Frequency and Monetary:* As we see in Fig. 1, we have Long Tail in our data which can threaten our clustering step. Therefore, according to [5], we use Log Transformation to change the distribution of our data from a Long Tail shape to a more *Normal* distribution.
- *Finding the best Number of Segments (best K for K-Means):* We use K-Means, as one the most applied clustering method, in order to find the best segments for our customers. For finding the best K for our K-Means model, we hold several meetings with project managers and marketing team. According to the meetings, we understood that E-Commerce companies need 3 segments of High-Value, Medium-Value and Low-Value based on the loyalty of their customers. Moreover, they need to be able to find the difference between the Medium-Value users who have high Frequency and low Monetary with the users who have low Frequency and high Monetary. Accordingly, we decided to have 4 segments of *High Value*, *Medium Value with High Monetary* (cusotmers with high Monetary but low Frequency), *Medium Value with High Frequency* (cusotmers with low Monetary but high Frequency) and *Low-Value* for our Active customers. Since the difference between the two segments of *Medium Value with High Monetary* and *Medium Value with High Frequency* is not useful for the users who have not purchased recently, we have 3 segments of *High Value*, *Medium Value* and *Low Value* for the Lapsing and Lapsed Customers. The reason is that the goal for the Lapsing segment is avoiding them from being churned and the goal for the Lapsed segment is encouraging them to return and there is no meaningful difference between *Medium High Monetary* and *Medium High Frequency*. As a result, we need only one segment for our Medium Value customers in Lapasing and Lapsed Segments.
- *Finding the best Weights for the feature of (F $\times$ W_F + M $\times$ W_M):* In order to find the best values for W_F and W_M, we checked the values from 0 to 10 with step size of 0.1 for each of the Frequency and Monetary and tried to select the best values according to the result of clustering. In this process, we found 3 general clustering shapes and Fig. 2 shows them with their weights for the Active segment. We will show the selected segmentation model in the following part.

4) Final Customer Segmentation Model: In this part, we provide our final Customer Segmentation model. Fig. 2c shows the final Segmentation model for our Active customers. The reason of selecting 1.3 for W_F and 0.7 for W_M is that, as you



(a) Segmentation with W_F=1.5, W_M=1 (b) Segmentation with W_F=1, W_M=1 (c) Segmentation with W_F=1.3, W_M=0.7

Fig. 2: Segmentation Models based on different W_F and W_M

can see in the figure, these weights help us have our 4 segments in a meaningful manner. Indeed, we have a segment for our High Value customers (the **Green Part**), two parts for our two Medium Value segments (the **Yellow Part** for our *Medium Value with High Monetary* and the **Blue Part** for our *Medium Value with High Frequency*) and also a segment for Low Value Customers (the **Purple Part**). We did the same process for our Lapsing and Lapsed segments and selected Weights to generate clusters similar to Fig. 2a but with 3 segments of High Value, Medium Value and Low Value. Moreover, Table II provides detailed explanation for our final segments. For instance, our *Active High Value* segment has 110,818 customers who have done their last purchase in last 26 days (in average) and have purchased 1.65 times in a month and whose average monetary is 105 USD.

VI. Appropriate Strategies for each Segment

In order to target the customers based on their previous behavior, we have built appropriate plan for each segment according to the customer segmentation results. Therefore, we have made the following strategies for each of the segments.

- *Active High Value:* This segment contains the best customers who have bought recently and with high frequency and monetary. Thus, the plan for this segment is encouraging them to continue their behavior and also increase their value. For instance, companies should monitor the behavior of their best customers and recommend appropriate offers to maintain them in their states.
- *Active Medium with High Monetary:* This segment includes the customers who buy few times but spend high. Therefore, the plan for this segment is encouraging them to buy more often. Since this segment buy expensive products, increasing the Frequency of its customers leads to a dramatic benefit gain.
- *Active Medium with High Frequency:* The customers who buy with an appropriate frequency but spend low belong to this segment. Thus, the strategy for this segment is encouraging them to spend more. Since this segment has second ranked in frequency between the Active Customers, increasing their spends help companies improve their benefit.
- *Active Low Value:* This segment contains the lowest value of active customers who buy few times and spend low. Accordingly, the plan for this segment is converting them to more loyal customers by increasing their purchase count and money spending.
- *Lapsing High Value, Lapsing Medium Value, Lapsing Low Value:* The customers who have not bought recently are assigned to these segments. An important point is that, these customers have not been churned yet and the process of returning them is not hard. Thus, the strategy for this segment is encouraging them to return by recommending interesting offers. It should be noted that the offers for each of the lapsing segments should be different regarding their values (High Value, Medium Value and Low value) of their customers.
- *Lapsed High Value, Lased Medium Value, Lapsed Low Value:* These segments contain the customers who have not bought for a long period of time. An important point is that, these customers have been churned and the process of returning them may be hard. Therefore, the plan for this segment is encouraging them to return by recommending interesting offers. It should be mentioned that the cost of the offers should be meaningful according to the Customer Acquisition Cost (CAC) and the value of the customers.

VII. Evaluation and Experimental Results

In this section, at first, we compare our model with the previous RFM model in Digikala (based on Quantile Method which is one of the most common and efficient type of RFM) to show our improvement in the RFM method. Afterward, we evaluate our model and the strategies in a real practice.

A. Improving the Customer Quantile Method

Previous RFM method in Digikala was created based on Customer Quantile Method. This method only divided customers into the segments with the same size. Thus, it cannot find any relationship between customers. Moreover, this type of RFM do not pay attention for the relationship between

TABLE II: Customer Segmentation Overview

Segment		Details			
Recency	Monetary and Frequency	No. of Customers	Avg. Recency (Days)	Avg. Frequency (Monthly)	Avg. Monetary (USD)
Active	High Value	110,818	26	1.65	105
	Medium Value with High Monetary	217,309	40	0.45	112
	Medium Value with High Frequency	214,666	46	0.3	40
	Low Value	120,270	55	0.18	14
Lapsing	High Value	297,677	159	0.27	139
	Medium Value	410,987	188	0.15	46
	Low Value	256,177	265	0.03	26
Lapsed	High Value	396,525	617	0.03	126
	Medium Value	363,833	875	0	34
	Low Value	59,138	1,490	0	105

Recency, Frequency and Monetary. Since these features have tightly related to each others in several cases, RFM analysis should pay attention to their relationships to provide meaningful segments. For instance, there are several customers with high Frequency and low Monetary (the customers who bought from FMCG [2] category) and analyzing them without investigating the relationship between F and M is not convincing.
Moreover, the RFM models based on Quantile Method need to be handled by managers since after the analysis, the segmentation provides several segments, which is calculated by (**??**), and thus, managers have to merge most of them manually. The reason is that most of the segments have no meaningful difference with each other and decision makers cannot apply appropriate strategies for them. In our R+FM model, since we use clustering method, we provide final 10 segments for managers which are meaningful and be able to be assigned Marketing Strategies.

$$\begin{cases} \text{Total No. of Segments} = R_{Segment} \times F_{Segment} \times M_{Segment} \\ \quad \text{where } X_{Segment} \text{ is number of segments based on X} \end{cases} \tag{5}$$

B. Evaluation of R+FM model by Running an SMS Campaign

In order to evaluate our segmentation with a real practice, we designed an SMS campaign according to our proposed strategies. Moreover, we analyzed the results of the campaign and compared them with the previous Digikala's campaigns.

1) Design the experimental campaign: Based on the strategies which are provided in Section VI, we designed and ran an SMS campaign. In the campaign, we gave a Voucher to each customer according to his segment. The important point is that we only ran the campaign for Active Segments in order to build complete analysis and evaluate our proposed model. The reason is that in Digikala, they had ran several campaigns for their previous Active Segments which helped us do a valid comparison. We assigned the vouchers to each segment as follows:

- *Active High with Value:* We offered the customers of this segment, 10 percent discount up to 20 USD. We tried to maintain the state of these users with an interesting offer since they are our most loyal customers.
- *Active Medium with High Monetary:* We offered these customers 10 percent discount up to 10 USD. The reason is that the users of this segment spend high and we only tried to encourage them to buy more from Digikala in order improve their Frequency.
- *Active Medium with High Frequency:* For the customers of this segment, we offered 10 percent discount up to 20 USD for the purchases more than 50 USD (which is the average Monetary of this segment) in order to increase their purchase spending.
- *Active Low Value:* We offered the customers of this segment, 10 percent discount up to 20 USD. in this offer, converting these customers to more loyal customers was our goal.

2) Analysis of the Results of the Campaign: For analyzing the results of the campaign, we select 20 percent of customers in each segment randomly as our control customers and we did not send any vouchers to them. The monetary comparison between the control customers and the campaign customers (the ones who bought with given vouchers) before and after the campaign is shown in Table III. The purpose of the campaign for *Active Medium with High Frequency* was increasing their Monetary Spending. As shown in Table III, after running the campaign, the average monetary of the campaign customers in *Active Medium with High Frequency* segment was about 14.3 USD more than before running campaign whereas this increment was 3.2 USD for Control customers. As you can see in Table III, we have enhanced the average monetary of the campaign customers, compare with themselves and control customers.
Moreover, the previous SMS campaigns in Digikala had purchase rate of about 0.1 percent whereas we had 1 percent purchase in our campaigns in similar situation. The reason is that Digikala's previous segmentation could not help managers target their customer accurately and they had to send low value vouchers for them, but in our proposed RFM model, managers can target each of the segments according to their strategies and therefore, provide more interesting vouchers to the customers.

[2]Fast Moving Consumer Goods

TABLE III: Experimental Campaign Result Analysis

Segment		Average Monetary (USD)			
Recency	Monetary and Frequency	Control Users		Campaign Users	
		Before Campaign	After Campaign	Before Campaign	After Campaign
Active	High Value	74.2	73.8	88.2	89.2
	Medium Value with High Monetary	100.2	97.6	104.6	105.2
	Medium Value with High Frequency	32	35.2	35.4	49.7
	Low Value	50.7	53.2	56.4	65.2

VIII. Implementation

In order to do our Exploratory Analysis, we have used *R 3.4.4* and thus, R helped us gain several pieces of information and statistics about our data. Moreover, to implement our building model steps and providing services for managers, we have used *Python 3.6.5* and *Django 2.0*. Finally, we have benefited from Elastic Search as our Database in order to store our results and users information. It should be mentioned that the whole process of building our Customer Segmentation model takes around 75 minutes.

IX. Conclusions and Future Work

In recent years, the significance of behaving with customers according to their history has been growth dramatically. Customer Segmentation has been used as the basis for understanding and grouping customers. One of the most widely applied method for segmenting customers is RFM model. RFM model is extremely useful since it builds customer segmentation model effectively. However, simplicity threaten the power of RFM and the models need to be made and be improved by a manual process. Moreover, RFM models cannot confront with changes in the business and managers should handle them by ad-hoc decisions. In this paper, we found the best definitions for R, F and M to have a dynamic RFM model and also using K-Means in order to propose R+FM model which builds customer segmentation model dynamically. We applied our model on Digikala Online Retail Company and tried to create meaningful segments with appropriate strategies. Finally, we ran an SMS campaign in order to validate our approach and found that our segmentation model will improve the effectiveness of campaigns.

In future, we have a plan to improve our R definition in order to be both semi-dynamic (to be adjusted during the time) and simple (to be useful for marketing team). Furthermore, we want to calculate Customer Lifetime Value (CLV) for each segment and also each customer according to our segmentation.

References

[1] V. Aggelis and D. Christodoulakis. Rfm analysis for decision support in e-banking area. *WSEAS Transactions on Computers*, 4(8):943–950, 2005.

[2] M. H. Aghdaie and P. F. Tafreshi. A new perspective on rfm analysis. pages 1–20, 2017.

[3] E. Brynjolfsson, Y. Hu, and D. Simester. Goodbye pareto principle, hello long tail: Modeling and measuring the effect of search cost on product sales distribution. In *workshop on Information Systems and Economics (WISE), University of California, Irvine (December 10–11)*, 2005.

[4] H.-C. Chang and H.-P. Tsai. Group rfm analysis as a novel framework to discover better customer consumption behavior. *Expert Systems with Applications*, 38(12):14499–14513, 2011.

[5] N. J. Cox. Transformations: an introduction. http://fmwww.bc.edu/repec/bocode/t/transint.html, 2007.

[6] A. Dursun and M. Caber. Using data mining techniques for profiling profitable hotel customers: An application of rfm analysis. *Tourism management perspectives*, 18:153–160, 2016.

[7] S. H. Han, S. X. Lu, and S. C. Leung. Segmentation of telecom customers based on customer value by decision tree model. *Expert Systems with Applications*, 39(4):3964–3973, 2012.

[8] Y. He, C. Lin, W. Gao, and K.-F. Wong. Dynamic joint sentiment-topic model. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 5(1):6, 2013.

[9] S. M. Hosseini, A. Maleki, and M. R. Gholamian. Cluster analysis using data mining approach to develop crm methodology to assess the customer loyalty. *Expert Systems with Applications*, 37(7):5259–5264, 2010.

[10] Y.-H. Hu and T.-W. Yeh. Discovering valuable frequent patterns based on rfm analysis without customer identification information. *Knowledge-Based Systems*, 61:76–88, 2014.

[11] A. M. Hughes. *Strategic database marketing: the masterplan for starting and managing a profitable, customer-based marketing program*, volume 12. McGraw-Hill New York, 2000.

[12] U. Kaymak. Fuzzy target selection using rfm variables. In *IFSA World Congress and 20th NAFIPS International Conference, 2001. Joint 9th*, volume 2, pages 1038–1043. IEEE, 2001.

[13] S.-Y. Kim, T.-S. Jung, E.-H. Suh, and H.-S. Hwang. Customer segmentation and strategy development based on customer lifetime value: A case study. *Expert systems with applications*, 31(1):101–107, 2006.

[14] A. H. Kracklauer, D. Q. Mills, and D. Seifert. Customer management as the origin of collaborative customer relationship management. pages 3–6, 2004.

[15] J. R. Miglautsch. Thoughts on rfm scoring. *Journal of Database Marketing & Customer Strategy Management*, 8(1):67–72, 2000.

[16] M. Namvar, M. R. Gholamian, and S. KhakAbi. A two phase clustering method for intelligent customer segmentation. In *Intelligent Systems, Modelling and Simulation (ISMS), 2010 International Conference on*, pages 215–219. IEEE, 2010.

[17] D. D. Phan and D. R. Vogel. A model of customer relationship management and business intelligence systems for catalogue and online retailers. *Information & management*, 47(2):69–77, 2010.

[18] A. Z. Ravasan and T. Mansouri. A fuzzy anp based weighted rfm model for customer segmentation in auto insurance sector. In *Intelligent Systems: Concepts, Methodologies, Tools, and Applications*, pages 1050–1067. IGI Global, 2018.

[19] J.-T. Wei, M.-C. Lee, H.-K. Chen, and H.-H. Wu. Customer relationship management in the hairdressing industry: An application of data mining techniques. *Expert Systems with Applications*, 40(18):7513–7518, 2013.

[20] J.-T. Wei, S.-Y. Lin, C.-C. Weng, and H.-H. Wu. A case study of applying lrfm model in market segmentation of a childrens dental clinic. *Expert Systems with Applications*, 39(5):5529–5533, 2012.

[21] J. Wu and Z. Lin. Research on customer segmentation model by clustering. In *Proceedings of the 7th international conference on Electronic commerce*, pages 316–318. ACM, 2005.

[22] R.-S. Wu and P.-H. Chou. Customer segmentation of multiple category data in e-commerce using a soft-clustering approach. *Electronic Commerce Research and Applications*, 10(3):331–341, 2011.