

Clustering Prediction Techniques in Defining and Predicting Customers Defection: The Case of E-Commerce Context

Ait Daqud Rachid¹, Amine Abdellah², Bouikhalene Belaid³, Lbibb Rachid⁴

^{1,4}Department of Physics, FST, Sultan Moulay Slimane University, Morocco

²Department of Applied Mathematics, FST, Sultan Moulay Slimane University, Morocco

³Department of Mathematics and Informatics, PF, Sultan Moulay Slimane University, Morocco

Article Info

Article history:

Received Aug 31, 2017

Revised Feb 10, 2018

Accepted Feb 17, 2018

Keyword:

Churn prediction

Clustering

Customer churn

E-commerce

LRFM model

Prediction

ABSTRACT

With the growth of the e-commerce sector, customers have more choices, a fact which encourages them to divide their purchases amongst several e-commerce sites and compare their competitors' products, yet this increases high risks of churning. A review of the literature on customer churning models reveals that no prior research had considered both partial and total defection in non-contractual online environments. Instead, they focused either on a total or partial defect. This study proposes a customer churn prediction model in an e-commerce context, wherein a clustering phase is based on the integration of the k-means method and the Length-Recency-Frequency-Monetary (LRFM) model. This phase is employed to define churn followed by a multi-class prediction phase based on three classification techniques: Simple decision tree, Artificial neural networks and Decision tree ensemble, in which the dependent variable classifies a particular customer into a customer continuing loyal buying patterns (Non-churned), a partial defector (Partially-churned), and a total defector (Totally-churned). Macro-averaging measures including average accuracy, macro-average of Precision, Recall, and F-1 are used to evaluate classifiers' performance on 10-fold cross validation. Using real data from an online store, the results show the efficiency of decision tree ensemble model over the other models in identifying both future partial and total defection.

Copyright © 2018 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

Ait Daqud Rachid,

Department of Physics, FST,

Sultan Moulay Slimane University,

PB 523, Beni Mellal, Morocco.

Email: daoud.rachid@gmail.com

1. INTRODUCTION

According to the figures released by the Interbank Electronic Banking Centre (IEBC), the merchant sites have realized 3.1 million online payment transactions with credit cards for a total of MAD 1.3 billion in 2017. This year knew an increase of +81,7% in number and + 52,3% in the amount spent as opposed to 2016 [1]. Moreover, the growth of e-commerce means an increase in online competition. However, with the expansion of such Moroccan active e-commerce sites, it becomes difficult for a company to distinguish itself from other competitors – be they national or international.

The growing number of merchant sites have provided customers with more choices. In the first place, it incites them to divide their purchases amongst several e-commerce sites. Next, it enables them to compare as well as contrast competing products [2]. Nevertheless, the e-commerce sector suffers from the kind of relationship between companies and their customers (Non-contractual setting). In such a case, the customers have the possibility to change their purchasing patterns without informing sellers. Faced with these constraints, customers' retention necessitates disallowing them to switch some of their purchases to another

concurrent. This is a common challenge for any online company. We therefore address the following question to better retain customers: How can we avoid reducing customer retention rate? Put differently, how can we reduce customer churn in e-commerce context? According to Neslin *et al.*, [3] and Burez [4], two basic approaches exist for resolving this issue: On the one hand, ‘untargeted approaches’ that rely on superior product and mass advertising to increase brand loyalty and retain customers. A good example of this is AOL’s efforts to decrease churn through better software and content [5]. On the other hand, ‘targeted approaches’ that rely on identifying potential churners in order to avoid defection by targeting such customers with direct incentives [4], [6-9]. In this study, we are concerned with the second approach. For that alone, we investigate whether or not we are able to identify the moment when customers begin to discontinue their relationship with e-commerce website in order to target them by retention programs to avoid their total defection.

Customer relationship management, and customer churn prediction in particular, have received a growing attention during the last decade. Table 1 summarizes customer churn prediction models reported in the literature in recent years. The distinctive characteristics of each study in terms of the sectors, environment settings, defection types, and churn definitions are provided. As can be seen from Table 1, there are two major remarks: (1) Environment settings: great number of studies are in the contractual setting, that is characterized by the existence of a contract between the firm and the customer, in such a case, the date of churn is clearly known, and it matches up with the contract cancellation date. (2) Partial or total defection: most of those studies consider total defection, whilst only few studies use prediction models to identify partial defection [6], [10]-[12]. Moreover, each of those studies defines customer churn differently, this raises the following question: Which one is more appropriate?

Table 1 reveals that the churn issue has been under-researched in the e-commerce sector. Moreover, all analyses in this sector consider total defection (defection column). To discover both partial and total defection in e-commerce sector, this study contributes to the extant literature in two important ways. First, it combines LRFM model and clustering techniques during a calibration period (T1) to segment all customers into homogeneous clusters, then an LRFM pattern will be assigned to each cluster [13]. Change in the LRFM pattern (Moving a customer from one cluster with an important value in T1 to another group of less value in prediction period (T2)) may be a partial or total defection signal. Second, it introduces classification techniques for building prediction models to predict both partial and total defection in order to minimize the risk of churn.

On the other hand, contrary to research that seeks to retain only profitable customers [6], [7], [14], [15] or those that spend many efforts for the entire customer base [9], [16], [17], our study is centered not only on the customers who belong to the clusters representing the core customers, but also on those who demonstrate positive change in their purchase behavior even if they are grouped in clusters that do not contribute positively to profits.

The creation of a retention program that targets all types of customers will be very costly for the company. By adopting a method that focuses only on profitable customers, companies, especially those working in e-commerce field, can lose some customers. This could be ascribed to the lack of their engagement with the beneficiaries of the retention programs, which will lead to increased customer churn rate followed by a decrease in profits. These customers really deserve attention from the company; so they should not be eliminated, but they should be placed in another category. This is an important point because no company wants to miss the opportunity of converting a previously dissatisfied customer into a loyal customer. These customers are those that demonstrate positive change in their purchase behavior even if they are grouped in clusters that do not contribute positively to profits. The identification of these customers will be discussed in the following sections.

For example, in a situation where the goal of a company is to retain only profitable customers, the company should discover why customers leave and go to competitors. A churn analysis for their profitable customer’s segment shows that some customers leave the e-commerce website because delivery charges are not free. Subsequently, the company decides to reduce delivery costs for the most profitable customer in order to retain them. However, the less profitable customers are not served with this reduce; only profitable customers are satisfied. Therefore, targeting only profitable customers is not an optimal strategy for increasing retention rate because a group of customers was profitable in the past, doesn’t mean it will continue to be so in the future [18].

The rest of this paper is organized by 4 other sections, the research method including segmenting methods and data mining techniques used in this study are briefly described in Section 2 followed by an empirical study in Section 3 to demonstrate how this prediction approach works in practice, whereas in the fourth section we discuss the results. The paper finishes with the conclusion, limitation and some issues for future research.

Table 1. Literature Review

Literature review	Sectors	Environment setting	Defection	Churn definition
Buckinx and van den Poel (2005) [6]	Retailing	Offline environment / Non-contractual setting	Partial	Customer shift transaction pattern.
Shin-Yuan Hung, et al. (2006) [14]	Telecommunications	Offline environment / Contractual setting	Total	A churner is defined as a subscriber who is voluntary to leave
Jae-Hyeon Ahna, et al. (2006) [12]	Telecommunications	Offline environment / Contractual setting	Partial and Total	Change in customer's status.
Burez and van den Poel (2007) [4]	Pay-TV subscription	Offline environment / Contractual setting	Total	Subscriber does not renew or pay.
Xie, et al. (2009) [19]	Financial services	Offline environment / Contractual setting	Total	Not mentioned.
Xiaobing Yu, et al. (2011) [17]	E-commerce	Online environment / Non-contractual setting	Total	Customer stops doing business.
V.L. Miguéis, et al. (2012) [10]	Retailing	Offline environment / Non-contractual setting	Partial	Customers who, from a certain period, did not buy anything or those who in all subsequent periods spent less than 40% of the amount spent in the reference period.
Bingquan Huang, et al. (2012) [20]	Telecommunications	Offline environment / Contractual setting	Total	Not mentioned.
Kristof Coussemont, et al. (2013) [57]	Online gambling	Online environment / Non-contractual setting	Total	Gambler does not play during a period.
M. Clemente-Ciscar, al. (2014) [11]	Retailing	Offline environment / Non-contractual setting	Partial	Set of definitions of churn.
Faris (2014) [21]	Telecommunications	Offline environment / Contractual setting	Total	User leaves company.
A. T. Jahromi, et al. (2014) [7]	E-commerce (B2B)	Online environment / Non-contractual setting	Total	Inactivity in second half of the year.
K. Kyoungok et al. (2014) [22]	Telecommunications	Offline environment / Contractual setting	Total	Customers who quit the service or transfer to competitors at a date close to a specific date .
M.A.H. Farquad, et al (2014) [15]	Financial services	Offline environment / Contractual setting	Total	Shifting loyalties from one service provider to another
Ozden Gur Ali and Umut Arturk (2014) [8]	Financial services	Offline environment / Contractual setting	Total	Customer declared churned if her portfolio size falls below a specific threshold value and stays that way for six consecutive months.
Ssu-Han Chen (2016) [23]	E-commerce	Online environment / Non-contractual setting	Total	Customer with few sessions during the 8–12 months after registration.
Niccolò Gordinia, Valerio Vegliob (2017) [9]	E-commerce (B2B)	Online environment / Non-contractual setting	Total	When the company has been active (i.e. has at least one transaction in the year) but has no activity (i.e. purchase) in the year.
N. Holtrop, et al. (2017) [24]	Financial services	Offline environment / Contractual setting	Total	Customer with the insurer at the start of the year but is no longer at the end of the year.
This study	E-commerce	Online environment / Non-contractual setting	Partial and Total	Customer shifts LRFM pattern, changes in purchase behavior.

2. RESEARCH METHOD

The purpose of this study is to build a customer's churn prediction model in e-commerce sector by using clustering and prediction techniques to predict those customers who are likely to churn in the near future in order to minimize the risk of churn.

2.1. Customer profiling

Market segmentation is the process of identifying key groups within the general market that share specific characteristics and consuming habits [25]. RFM model, which was proposed by Hughes (1994) [26], is one of the most common methods for segmenting and identifying customer values in companies. Clustering techniques have been widely used to segment customers when using RFM model [13], [25], [27]-[29]. In this section, we discuss k-means as clustering technique and the LRFM model as the extended version of RFM model that consider customer relationship length (L) we use for customer profiling task.

2.1.1. RFM and LRFM models

RFM model is an effective method of segmenting and it is likewise a behavioral analysis that can be employed for market segmentation [30], [31]. A. Hughes [30] 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 [32]. A. Yang [32] 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. 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 [33]-[35]. For example, Chang and Tsay [36] 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. In addition, S. Chow and R. Holden [37] suggest 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.

2.1.2. K-means method

K-means clustering is the most common algorithm used to cluster n vectors based on attributes into k partitions, where $k < n$, depending on some measures. 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 [38].

However, in the k-means technique, the number of clusters is randomly selected, which means that the clustering result will become unreliable if the supposed number of the clusters is incorrect [39], [40], this raises the following fundamental question: How to choose the right number of expected clusters (k)?

Some types of efficient clustering quality indexes can help determine the best number. In this study, we have used two methods for determining the optimal number of clusters for k-means. These methods consist of optimizing a criterion, such as the within cluster sums of squares and the average silhouette. The corresponding methods are named elbow and silhouette methods, respectively. In this study, the sum of squared errors (SSE) and the average silhouette coefficient which are shown in the Equations (1) and (2) respectively, are combined to measure the quality of clustering and to determine the optimal clustering number. Specifically, we applied k-means technique under different k values and then we plot the curves of the SSE and average silhouette coefficient against the number of clusters to analyze the two curves and to identify the optimal number of cluster. The optimal clustering number can be found in a data set by looking for the number of clusters at which a knee, peak, or dip exists in the plot of the evaluation measure when plotted against the number of clusters [41].

$$SSE = \sum_{i=1}^k \sum_{x_j \in C_i} \|y_i - c_i\|^2 \quad (1)$$

Where k is the number of clusters, y_i is the j th object in cluster C_i and c_i is the center of cluster C_i .

$$S_i = \frac{b_i - a_i}{\max(a_i, b_i)} \quad (2)$$

Where a_i is the average distance of object I to all other objects in its cluster; for object I and any cluster not containing it, calculate the average distance of the object to all the objects in the given cluster, and b_i is the minimum of such values with respect to all clusters. The details of SSE and Silhouette can be found respectively in [42], [43].

This study combines K-means and LRFM model in e-commerce sector to divide the customer base up into homogeneous clusters according to their L, R, F and M values. Similarly to Chang and Tsay [36], we will use average LRFM values of each cluster to compare with the total average LRFM values of all clusters. If the average (L,R,F,M) value of a cluster is greater than the total average, an over bar appears. However, if the average (L,R,F,M) value of a cluster is less than the total average, an under bar appears. (i.e., \bar{R} : Higher R value; customer have recently made a purchase, \underline{R} : Lower R value; customer have not buy on the online store for a long time).

Chang and Tsay [36] based on Ha and Park [44] further proposed customer classification by summing the sixteen combinations of LRFM model to five kinds of customer groups according to their LRFM patterns, such as core customers, potential customers, lost customers, new customers, and resource-consumption customers. Specifically, core customers include $L\uparrow R\uparrow F\uparrow M\uparrow$, $L\uparrow R\uparrow F\uparrow M\downarrow$, and $L\uparrow R\uparrow F\downarrow M\uparrow$. Potential customers consist of $L\uparrow R\downarrow F\uparrow M\uparrow$, $L\uparrow R\downarrow F\uparrow M\downarrow$, and $L\uparrow R\downarrow F\downarrow M\uparrow$. Lost customers are composed of $L\downarrow R\downarrow F\uparrow M\uparrow$, $L\downarrow R\downarrow F\uparrow M\downarrow$, $L\downarrow R\downarrow F\downarrow M\uparrow$, and $L\downarrow R\downarrow F\downarrow M\downarrow$. New customers comprise $L\downarrow R\uparrow F\downarrow M\downarrow$, $L\downarrow R\uparrow F\uparrow M\downarrow$, $L\downarrow R\uparrow F\downarrow M\uparrow$, and $L\downarrow R\uparrow F\uparrow M\uparrow$. Finally, resource-consumption customers are $L\uparrow R\uparrow F\downarrow M\downarrow$ and $L\uparrow R\downarrow F\downarrow M\downarrow$.

When different LRFM combinations are identified during a period T, customers can be classified into appropriate groups such as core customers, potential customers, lost customers, new customers, and resource-consumption customers. First, we focus on customers belong to core customers, new customers (no company want to miss new customers), second, we take into account those belonging during the period T to other remaining groups, and which are subsequently converted into core customers in T+1. More specifically, the customers in our clusters of attention belong to the following patterns:

- \overline{LRFM} , \overline{LRFM} , \overline{LRFM} , \overline{LRFM} , \overline{LRFM} and \overline{LRFM} during a period T.
- Customers who do not belong in period T to the patterns listed in (1), but in the T+1 period, their LRFM pattern transformed into one of the patterns mentioned in (1).

Customers who were clustered during the period T with potential, lost or resource-consumption customers, and that are stayed in the same group or are transformed to a lower value group in T+1, they will be removed.

2.2. Partial and total churning

Among the first main hurdles which face on the customers churn prediction in the non-contractual businesses is the difficulty of defining churn because the characteristics that should be observed to saying that a customer has totally or partially defected are not clearly defined [11].

For solving the problems above, (definition of customer churn) LRFM model and clustering technique (k-means) are combined. This study proposes a new procedure by joining the quantitative values of the LRFM attributes, extracted during a period T, into K-means algorithm to identify the different types of customer profiles (different LRFM patterns). We then define a customer's LRFM pattern change from a core (\overline{LRFM} , \overline{LRFM} , \overline{LRFM}) or a new customer (\underline{LRFM} , \underline{LRFM} , \underline{LRFM}) to potential customer (\overline{LRFM} , \overline{LRFM} , \overline{LRFM}) or to low consuming resource customer groups (\overline{LRFM}) as partial defection. By the same token, if a customer changes her LRFM model from one of two following types of customer: core customers (\overline{LRFM} , \overline{LRFM} , \overline{LRFM}) or new customer (\underline{LRFM} , \underline{LRFM} , \underline{LRFM}) to the lost customers (\underline{LRFM} , \underline{LRFM} , \underline{LRFM} , \underline{LRFM}) or to high consuming resource customer groups (\overline{LRFM}), in this case, we are talking about total defection. This would indicate that a customer's change in LRFM patterns is an early signal of either partial or total defection. Whereas customers who staying true to their existing positive patterns are likely to stay. For this purpose, as shown in Figure 1, we consider two equal sub-periods T1 and T2. T1 is used to determine the different customer groups (different LRFM patterns) and assign each customer to its appropriate group. The period T2 is used to determine partial or total defection. Figure 2 illustrates our proposed approach to defining partial and total defection, and the full process is summarized in Figure 3.

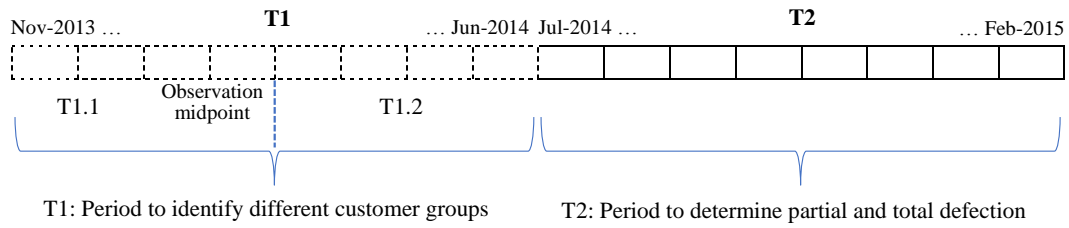


Figure 1. Period of observation. T1 this period of eight months (from November 2013 to June 2014) was also used to derive the independent variables (calibration period) of the model. T2 this period of eight months (from July, 2014 to February, 2015) was used to derive the dependent variable (prediction period)

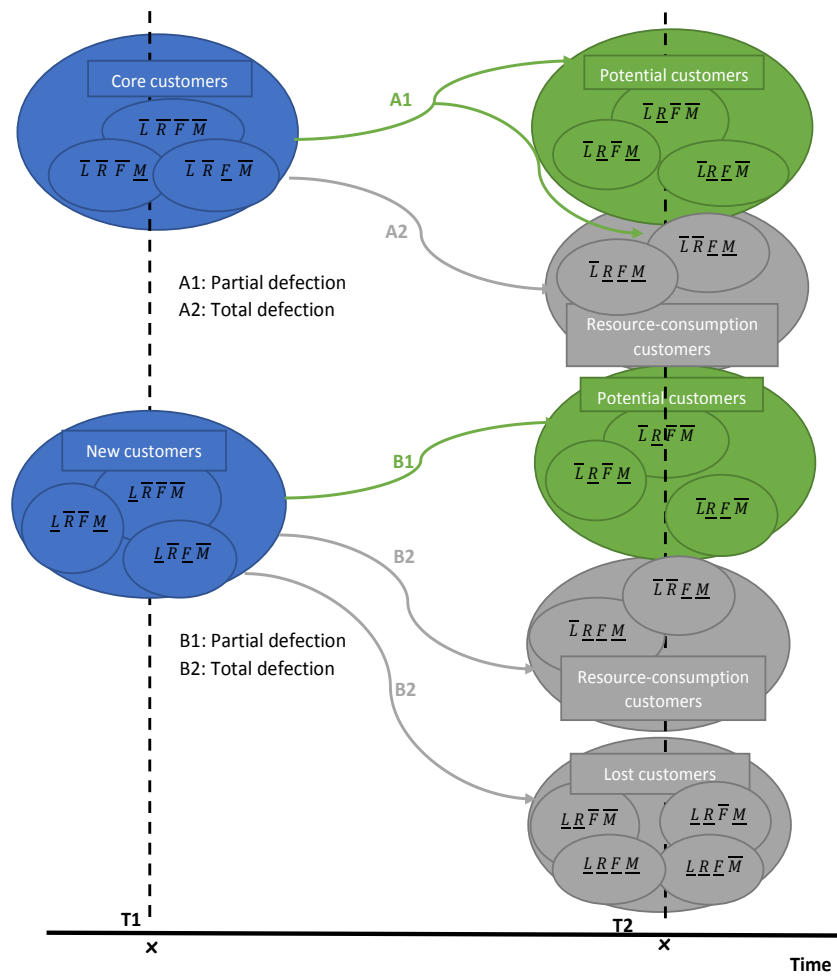


Figure 2. Deviation of customers LRFM pattern over time to define customer churn

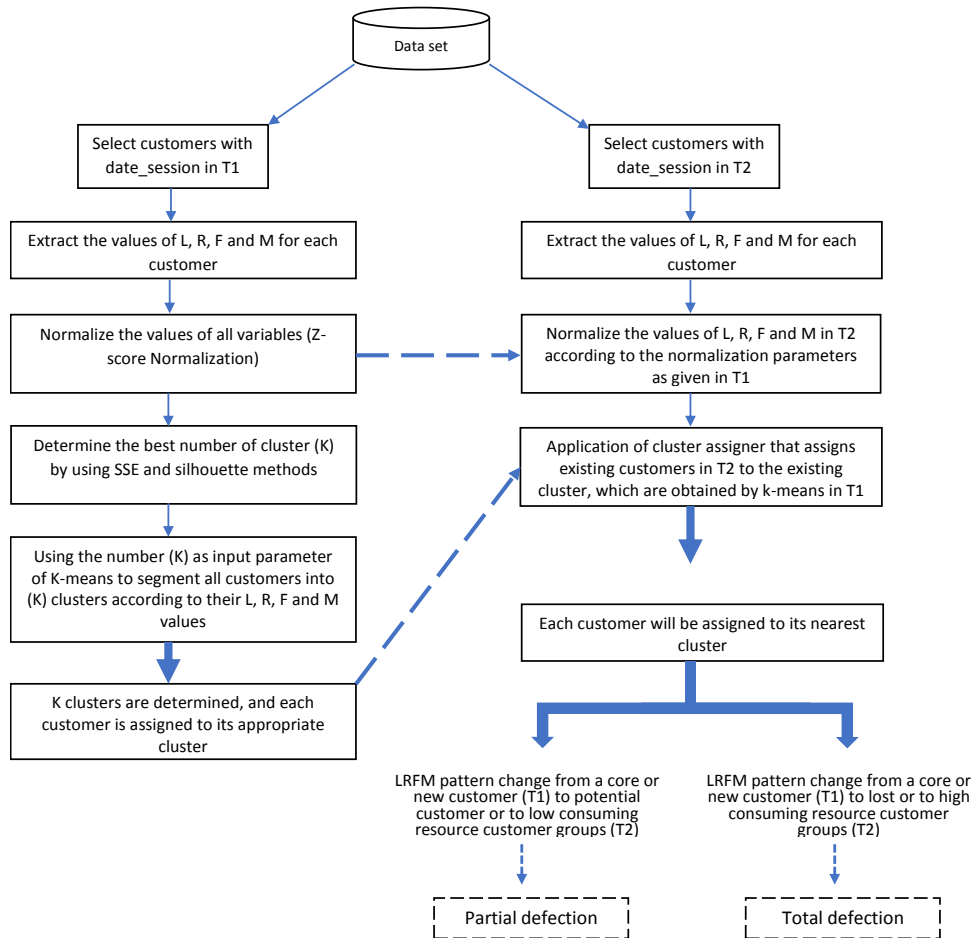


Figure 3. Defining churn in non-contractual settings methodology based on LRFM model and K-means technique

2.3. Classification techniques

The objective of this research was to develop a predictive model for customer churn in a non-contractual setting which would be able to distinguish between customers who are likely to partially or totally churn in the near future and the ones who are likely to stay with the company based on historical transactions and characteristics of a customer. To reach this goal three models are proposed, the first is based on decision tree techniques (DT), the second on artificial neural networks (ANN) and the third based on an ensemble of decision trees. We note that all our models are constructed using KNIME Analytics Platform 3.3.2. The following is the short description for these known data mining techniques used for this task.

2.3.1. Artificial neural networks (ANN)

Unlike to conventional statistical methods, artificial neural networks do not need any hypothesis on the variables, they are well-suited to handle unstructured complex problems, i.e. issues on which there is no a priori specify the form of relationships between variables.

Neural networks can be distinguished into single-layer perceptron and multilayer perceptron (MLP), in this paper, we use the MLP structure that allows realizing the most diverse applications. An MLP network is generally composed of a finite set of cells (neurons), organized in successive layers. The first layer comprising several neurons is called the input layer, the last layer is the output layer, and the intermediate layers (if any) are the hidden layers. Neurons in different layers are connected by sigmoid or hyperbolic tangent functions that are used as activation functions in Multi-layer perceptron. The details of MLP can be found in [45].

2.3.2. Simple decision tree (DT)

Decision tree (DT) is one of the most data mining techniques for knowledge discovery and it is used usually for the purpose of classification and prediction [46]. The simplicity and ease of interpreting the

observed results by decision makers are the main reasons for its popularity in business compared to other prediction techniques [47]. DT development usually consists of two distinct stages, tree building, and tree pruning. At first, the techniques start to search in the training set an attribute offering the best information gain at the root node level, and then dividing the tree into sub-trees. The same procedure is used to recursively partitioned the sub-tree following the same rule, then the partitioning stops when the leaf node is reached. Once the tree is created, rules can be extracted by traversing through the tree until a leaf node is reached. Several algorithms such as C4.5, C5.0, CHAID and CART are used to produce the trees, in this study we consider C4.5 algorithm. The details of DT can be found in [48], [49].

2.3.3. Decision tree ensemble (DTE)

Despite the advantages of the decision tree method mentioned above, it also has some disadvantages. For example, Dudoit, Fridlyand, and Speed [50] note some of its disadvantages; e.g. its suboptimal performance and the lack of robustness. Among the best ways to solve them is the creation of the ensemble of trees followed by a vote for the most popular class [51]. This solution is the result of some researchers who optimized the Decision tree technique.

In this regard, we use the both Tree Ensemble Learner and the Tree Ensemble Predictor nodes of Knime to build our third model that is based on decision tree ensemble.

The Tree Ensemble Learner node builds an ensemble of decision trees, as a variant of the random forest. Each of the decision tree models is trained on a different subset of rows and/or on a different subset of columns, randomly selected at each iteration. The output model is then an ensemble of differently trained decision tree models. The decision trees learning parameters are similar to the Random Forest classifier described by Leo Breiman [51]. The Tree Ensemble Predictor node applies all decision trees to each data row and uses the simple majority vote for prediction.

3. EMPIRICAL STUDY

3.1. General

The data analyzed in this research have been provided from one of the biggest online retailers specialized in electronics, fashion, home appliances and children's items in Morocco. When customers visit the website, the system records their login, logout, shopping process and the final state of each session. A customer can make four types of events, namely "Session with Product Views", "Session with Add to Cart", "Session with Check-Out", and "Session with Transactions". The dataset consists of 2783 customers who visited the e-commerce website. Specifically, the dataset consists of information at the individual customer level, such as customer register, login, session, transaction and web log in the e-commerce website. Transactional records of customers for the period November 1, 2013 through February 28, 2015 have been utilized.

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

The Transactional records 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.

The definition of LRFM model used in this study is shown in Table 2.

The descriptive statistics for the variables (LRFM) in T1 are provided in Table 3.

Table 2. The Definitions of LRFM Model

Attribute name	Data content
Length (L)	Refers to the number of days from the first to the last purchase
Recency (R)	Refers to the number of days between the first day of study period and the day of the last purchase.
Frequency (F)	Refers to the number of transaction observed in the period analyzed
Monetary (M)	Refers to the total amount spent by customers in the period analyzed. (Moroccan dirhams)

Table 3. The Descriptions of Length, Recency, Frequency and Monetary in T1

Variables	Max	Min	Average	Standard deviation
Length (L)	813	2	656.68	192.87
Recency (R)	241	1	164.77	76.05
Frequency (F)	17	1	8.67	4.99
Monetary (M)	13,723.00	87.00	4431.15	4327.72

3.2. Clustering by K-means based on LRFM variables

The first eight-months period of the available data, from November, 2013 to June, 2014 (T1), is used to identify the different customer groups (different LRFM patterns). Consequently, 2692 customers who visited the e-commerce website in this period are selected.

According to the proposed model described in Section 3, KNIME Analytics Platform 3.3.2 is used. Consequently, we find seven clusters of customers that have a different LRFM behavior. The optimal number of cluster ($k=7$) is obtained based on elbow and silhouette methods. Figure 4 shows the plots of the SSE and average silhouette coefficient versus the number of clusters for k-means. A distinct knee in the SSE and a distinct peak in the silhouette coefficient are present when the number of clusters is equal to 7.

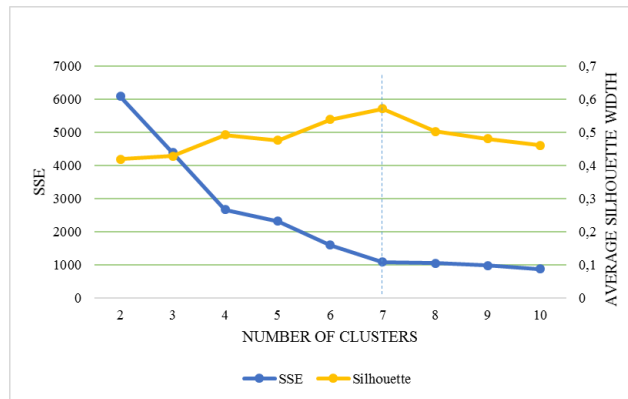


Figure 4. Elbow and Average silhouette methods for determining the optimal number of cluster

Table 4 is a summary of the clustering of these seven clusters, each with the corresponding number of customers, average length (L), average recency (R), average frequency (F), average monetary (M) and the last column shows the LRFM pattern for each cluster. Most of the customers are in Clusters 1, 3 and 5. Whereas, cluster 6 includes the minimum number customers (only 77 customers).

As mentioned earlier, we focus our study on customers belong to core customers \overline{LRFM} , \overline{LRFM} and \overline{LRFM} (Cluster 2, 3 and 4) and the high-value new customers \underline{LRFM} (Cluster 0), the both represent 51.23 % of the total available customer database.

Table 4. Descriptive Statistics of Seven Clusters based on K-Means Method in T1

Cluster	Count	Mean(L)	Mean(R)	Mean(F)	Mean(M)	Pattern
cluster_0	332	282.31	211.26	12.41	8204.28	\underline{LRFM}
cluster_1	760	752.87	48.31	4.47	857.62	\overline{LRFM}
cluster_2	375	707.71	206.59	13.34	2187.08	\overline{LRFM}
cluster_3	509	742.81	210.91	14.65	10817.28	\overline{LRFM}
cluster_4	210	741.10	209.95	6.97	8266.72	\overline{LRFM}
cluster_5	428	699.51	212.89	3.93	1063.69	\overline{LRFM}
cluster_6	77	35.52	214.66	2.91	405.12	\underline{LRFM}

In the second period T2 (from July, 2014 to February, 2015), we introduce the cluster assigner node (that assigns existing customers in T2 to the existing groups, which are obtained by k-means in T1) to determine customer who has moved from the core customer in T1 to the defector customer during the subsequent period of eight months. Applying our partial-total churn definition described in 3.2 section results in 254 partial defections (17.81% = 254/1426) and 363 total defections (25.45% = 363/1426), where 1426 represents the number of customers under investigation (cluster_2 + cluster_3 + cluster_4 + cluster_0 = 1426).

3.3. Variables operationalization

3.3.1. Predictors (independent variables)

A major part of the existing studies related to the prediction of customer churn focuses on the incorporation of two groups of information: behavioral information and customer demographics. According to several studies such as Coussement and Van den Poel [52], Guadagni and Little [53]; Rossi *et al.*, [54], and Tamaddoni Jahromi *et al.*, [7] demographic data (i.e gender, age, address, profession, etc) have less impact on churn prediction. For this, our study will be based only on behavioral information at the level of the individual customer (independent variables), this will allow us to keep the models in their simplest form and, on the other hand, to maximize their predictive power.

Compared with traditional transaction methods, the biggest advantage of e-commerce is that all the navigation data of all the visits made by customers on the e-commerce site are stored in the servers. From this behavioral and transactional information at the level of the individual customer (page viewed, sequence of visits, purchase process, number of transactions, etc ...) and in addition to RFM variables, many indicators can be extracted [55], and used as predictor variables by our models to improve their distinction power between customers totally churn and those who partially defect and those who remain loyal. An overview of all extracted variables used in this study is presented in Table 5.

Table 6 summaries all behavioural independent variables supported by previous research in the both offline and online environments. The recency, frequency and monetary variables are those that have more popularity in predicting customer churn in the online environments. The variables that describe the dropout rates at each step of the buying process, the length of relationship (L), average interpurchase time (ITP) and mode of payment (Mopayment) are variables infrequently used in prior research. Therefore, in order to assess their importance in predicting customer churn, we will take them into account.

Table 5. Predictor Variables

Variable type	Variable name	Description
Recency	R	Number of days between the first day of the study period and the day of the last purchase in calibration period ($0 \leq R \leq T1$).
Frequency	F	Number of purchases observing during the calibration period (T1).
	R_change.F	Relative change in number of purchases in the second half of the calibration period F.T1.2 when compared with the first half of the calibration period F.T1.1, i.e. $R_change.F = (F.T1.2 - F.T1.1) / F.T1.1$
Monetary	M	Total monetary amount of purchases in calibration period (T1).
Length of relationship	L	Number of days from the first to the last purchase.
Interpurchase time	ITP	Average number of days between purchases.
	R_change.ITP	Relative change in interpurchase time in the second half of the calibration period ITP.T1.2 when compared with the first half of the calibration period ITP.T1.1, i.e. $R_change.ITP = (ITP.T1.2 - ITP.T1.1) / ITP.T1.1$
Mode of payment	Mopayment	Indicates the most mode of payment used in the last three transactions.
Dropoutrate	Last_session_abandoned	Indicates whether the last session was abandoned at checkout step (yes,non).
	aband_rate(allvisit to productviews)T1.2	The percent of sessions that abandoned the buying process at the "Product views" step in the second half of the calibration period T1.2.
	aband_rate(productviews to addcart) T1.2	Abandonment rate for a customer when moving from "Product views" step to "Add to cart" step in the second half of the calibration period T1.2.
	aband_rate(addcart to checkout)T1.2	Abandonment rate for a customer when moving from "Add to cart" step to "Check-out" step in the second half of the calibration period T1.2.
	aband_rate(checkout to transaction)T1.2	Abandonment rate for a customer when moving from Check-out step to Transaction step in the second half of the calibration period T1.2.
	aband_rate(allvisit to transaction)T1.2	The percent of sessions that abandoned the buying process at Check-out step in T1.2 period.

Table 6. Behavioural Predictors of Defection and Type of Target Variable in Prior Research

		Predictor variables							Target variable Variable type		
		L	R	F	M	IPT	Product categories	Mode of payment		Failure	Dropoutrate
Offline environment	Buckinx and van den Poel (2005) [6]	X	X	X	X	X	X	X			Binary (Churner, Non-churner)
	V.L. Miguéis, <i>et al</i> (2012) [10]	X	X	X	X						Binary (Partially churned, Non-churner)
	Mozer, <i>et al.</i> (2000) [16]	X		X	X	X	X	X			Binary (Churner, Non-churner)

		L	R	F	M	IPT	Predictor variables				Target variable Variable type
							Product categories	Mode of payment	Failure	Dropoutrate	
Online environment	Keaveney and Parthasarathy (2001) [56]			X	X						Binary (switchers, continuers)
	K. Coussement and, K. W. De Bock (2013) [57]	X	X	X	X	X					Binary (Churner, Non-churner)
	A.T. Jahromi, et al. (2014) [7]		X	X	X						Binary (Churner, Non-churner)
	Ssu-Han Chen (2016) [23]		X			X					Binary (Churner, Non-churner)
	N. Gordini and V. Vegliob (2017) [9]	X	X	X	X		X		X		Binary (Churner, Non-churner)
	This study	X	X	X	X	X		X		X	Multi-class (Partially-churned, Totally-churned, Non-churner)

3.3.2. Target variable (dependent variable)

The target variable in the current study is ‘status’, a categorical variable which has three values: Partially-churned, Totally-churned and Non-Churner, and which is predicted based on customer's event history on the e-commerce website.

3.4. Performance Measures

Table 6 reveals that all existing churn studies focus on binary classification models. This study contributes to the literature by not focusing on the binary classification models where the model predicts the status of a customer as churner or non-churner. Our study, however, will address the case of Multi-Class Classification problems where the dependent variable classifies a particular customer either as a customer continuing his loyal buying pattern (Non-churned) or as a partial defector (Partially-churned) or as a total defector (Totally-churned).

In terms of multi-class classification problems, Micro-average and Macro-average measures are commonly used in evaluating performance. However, the Micro averaging does not provide an accurate measure of performance when the instances are not equally distributed over the classes (most instances belong to one class). Unlike Micro averaging, Macro averaging provides meaningful performance measure despite that data is not equally representative of all classes (imbalanced classes) [58]. Therefore, Macro averaging is used as a measure to evaluate the multi-class model performance in this study.

Table 7 gives a typical resulting confusion matrix (is a table that shows each class in the test set and the number of correct predictions and incorrect predictions) for a problem with three classes, where N^{ij} represents the number of instances with actual class i which are predicted as a class j . ($i = 1, 2, 3, j = 1, 2, 3$).

Table 7. A Typical Resulting Confusion Matrix

		Predict		
		Class _i	Class _j	Class _k
Actual	Class _i	N^{ii}	N^{ij}	N^{ik}
	Class _j	N^{ji}	N^{jj}	N^{jk}
	Class _k	N^{ki}	N^{kj}	N^{kk}

Table 8 presents the most often used measures for multi-class classification based on the values of the confusion matrix.

Besides, we consider 10-fold cross validation in which the initial data are randomly divided into 10 equal parts, 9 parts are used as training data to build the prediction model, while the remaining one part is reserved as the test set. Therefore, each part will be trained and tested ten times and the average accuracy rate can be obtained consequently.

Table 8. Measures for multi-class classification used in this study based on TP_i: instances of Class_i correctly predicted, TN_i: instances of other Class_(j,k) correctly predicted, FP_i: instances of other Class_(j,k) predicted as instances of class_i and FN_i: instances of Class_i incorrectly. M indice represent macro-averaging [59]

Measure	Formula	Description
Average Accuracy	$\sum_{i=1}^l \frac{TP_i+TN_i}{TP_i+TN_i+FP_i+FN_i}$	The average per-class effectiveness of a classifier
Error Rate	$\sum_{i=1}^l \frac{FP_i+FN_i}{TP_i+TN_i+FP_i+FN_i}$	The average per-class classification error
Precision _M	$\sum_{i=1}^l \frac{TP_i}{TP_i+FP_i}$	An average per-class agreement of the data class labels with those of a classifiers
Recall _M	$\sum_{i=1}^l \frac{TP_i}{TP_i+FN_i}$	An average per-class effectiveness of a classifier to identify class labels
F-1 _M	$\frac{l}{((1+\beta^2) * Precision * Rappel) + ((\beta^2 * Precision) + Rappel)}$, <i>avec</i> $\beta^2 = 1$	Relations between data's positive labels and those given by a classifier based on a per-class average

4. RESULTS AND DISCUSSION

The actual fraction of partially churned, totally churned, and non-churned in the dataset is 17.81%, 25.45%, and 56.73% respectively. In order to evaluate the quality of the predictions of the churn prediction models including decision tree, artificial neural network and Decision Tree Ensemble, Macro average measures and 10-fold cross validation method are considered.

Table 9. Prediction performances of the three models with 10-fold cross validation

	DT	ANN	DTE
fold 1	93.71	90.21	93.71
fold 2	95.10	95.10	98.60
fold 3	95.80	97.20	98.60
fold 4	93.01	90.91	95.10
fold 5	92.31	91.61	95.10
fold6	96.50	97.90	99.30
fold7	96.48	93.66	97.18
fold8	95.07	92.25	97.18
fold9	92.25	93.66	97.18
fold10	95.07	92.25	96.48
Avg	94.53	93.48	96.84

Table 9 shows the prediction performance of three models based on 10-fold cross-validation. On average, the prediction models provide higher than 93% accuracy. When comparing the different classification techniques, the Decision Tree Ensemble offers the best results in terms of accuracy in all the test folds. Accuracy alone is sometimes quite misleading to confirm the prediction quality [60]. For this, additional measures of models' performance such as Recall, and Precision are required to identify the better performing churn prediction model. Therefore, based on confusion matrix tables we calculate the Recall_i, Precision_i and F-1_i values for each class to assess the performance with respect to every of three classes in our dataset. The detailed results are presented in Table 10, Table 11 and Table 12.

Table 10, Table 11 and Table 12 summarize the overall accuracy, recall, precision and F-1 values for each of the three classes for the three classification techniques with 10-fold cross validation.

Table 10. The overall accuracy, Recall, Precision and F-1 values for each of the three classes for the ANN classifier with 10-fold cross validation

ANN	Overall accuracy	TP	FP	TN	FN	Recall	Precision	F-1
Partially-churned	93.48 %	192	29	1143	62	0.756	0.869	0.808
Totally-churned		350	27	1036	13	0.964	0.928	0.946
Non-churned		791	37	580	18	0.978	0.955	0.966

Table 11. The overall accuracy, Recall, Precision and F-1 values for each of the three classes for the DT classifier with 10-fold cross validation

DT	Overall accuracy	TP	FP	TN	FN	Recall	Precision	F-1
Partially-churned	94.53 %	218	42	1130	36	0.858	0.838	0.848
Totally-churned		346	13	1050	17	0.953	0.964	0.958
Non-churned		784	23	594	25	0.969	0.971	0.97

Table 12. The overall accuracy, Recall, Precision and F-1 values for each of the three classes for the DTE classifier with 10-fold cross validation

DTE	Overall accuracy	TP	FP	TN	FN	Recall	Precision	F-1
Partially-churned	96.84 %	218	9	1163	36	0.858	0.960	0.906
Totally-churned		357	14	1049	6	0.983	0.962	0.973
Non-churned		806	22	595	3	0.996	0.973	0.985

Table 13. Macro-averaging measures for the three classifiers

	Average Accuracy	Recall _M	Precision _M	F-1 _M
Artificial Neural Networks (ANN)	0.957	0.899	0.917	0.907
Simple Decision Tree (DT)	0.964	0.927	0.924	0.925
Decision Tree Ensemble (DTE)	0.979	0.946	0.965	0.955

The results shown in Table 10, Table 11 and Table 12 indicate that for the three classes of our database the DTE model offers the better results in terms of precision, recall and F-1.

Moreover, as stated in Table 13, the DTE shows better predictive performance than the other models in terms of Macro-averaging measures. Compared to DT and ANN is + 1.90, + 4.63 points respectively in terms of Recall_M, whilst the improvement in Precision_M is + 4.06, + 4.76 respectively and + 2.93, + 4.80 respectively in terms of F-1_M. Based on these results, we conclude that the DTE model performs the best in identifying customers totally defect, partially defect and those who remain loyal. Consequently, we are able to follow both partial and total defection in contrast with past research that focused either on a total or partial defect. This contribution is important due to several reasons. First, since we consider the two types of defection (partial and total), the degree of risk related to partial defection is different from that of total defection. Therefore, due to the costs associated with retention strategies, it is advisable not to concentrate churn management efforts in the same way on both [61]. In other words, a customer predicted by the model as a customer that will likely to (partially) churn in the future, should not be targeted by the same incentive program dedicated to those predicted that will likely to leave the company definitely in the future and vice versa. This will help the managers to make right interactions at the right time for retaining these customers without wasting resources. Secondly, the ability to check whether the total churn is always preceded by a partial attrition or that there are cases where the customer definitely leaves without leaving any signs of dissatisfaction. This will allow companies to think about solutions to such situations.

Finally, we consider which predictors contribute more to predicting partial and total customer defection using the three models. For this, we create a knime workflow that allows calculating the variable importance, the basic idea is: for calculating the importance of the variable k, we exclude it out for the prediction. If the prediction accuracy of the model decreases to the absence of this variable, this indicates that the latter is important in the prediction process. The same procedure is used for all variables and finally, the importance variable for each predictor is normalized from 0 to 1 to get an indicator how important each variable. The results are presented in Table 14. Therefore, the knowledge of the most important churn predictors in the e-commerce sector would be of particular interest to marketing manager because they provide actionable information to target the customers that are most likely to churn in the near future with tailored incentives to minimize the risk of churn [9].

Table 14 shows the importance of each predictor for DT, ANN and DTE models, where the first column indicates the rank of importance, the second shows the name of the variable, and the last column shows the normalized importance of a variable based on the accuracy of the model without it.

Table 14. Importance of Variables

Rank	Decision tree ensemble		Decision tree		Artificial Neural Networks	
	Variable	NormImp	Variable	NormImp	Variable	NormImp
1	aband_rate(checkout-transaction)T1.2	1.000	aband_rate(checkout-transaction)T1.2	1.000	aband_rate(checkout-transaction)T1.2	1.000
2	aband_rate(productviews-addcart)T1.2	0.562	aband_rate(productviews-addcart)T1.2	0.287	aband_rate(productviews-addcart)T1.2	0.652

Rank	Decision tree ensemble		Decision tree		Artificial Neural Networks	
	Variable	NormImp	Variable	NormImp	Variable	NormImp
3	aband_rate(allv-transaction)T1.2	0.210	aband_rate(addcart-checkout)T1.2	0.181	aband_rate(allv-transaction)T1.2	0.277
4	aband_rate(addcart-checkout)T1.2	0.181	aband_rate(allv-productviews)T1.2	0.070	aband_rate(addcart-checkout)T1.2	0.241
5	L	0.038	R	0.064	L	0.143
6	aband_rate(allv-productviews)T1.2	0.038	Last_session_abandoned	0.058	R	0.098
7	ITP	0.029	aband_rate(allv-transaction)T1.2	0.053	R_change.F	0.089
8	R_change.ITP	0.029	L	0.053	ITP	0.071
9	F	0.019	ITP	0.053	R_change.ITP	0.071
10	M	0.010	R_change.F	0.053	F	0.071
11	Mode-ofpayment	0.010	M	0.047	aband_rate(allv-productviews)T1.2	0.063
12	Last_session_abandoned	0.010	F	0.035	M	0.027
13	R	0.010	Mode-ofpayment	0.029	Mode-ofpayment	0.027
14	R_change.F	0.000	R_change.ITP	0.000	Last_session_abandoned	0.000

It is clear from the rankings of variable importance that variables which describe the dropout rates in the buying process steps like: `aband_rate(checkout-transaction)T1.2`, `aband_rate(productviews-addcart)T1.2` and `aband_rate(addcart-checkout)T1.2` are indeed powerful predictors of partial and total churn in the e-commerce field. The confirmation of this view is given by the similarity in the ranking of the importance of these variables, where we find them at the four of top-ten variables of all models and seem to outperform other variables. The main difference between DTE model and the other models is, most clearly related to the variable that describes the recency of the last purchase (R) which appears at the bottom of the ranking for DTE and quite important in the other two models. However, when comparing the results of the study with the recent research that deals with the prediction of churn in online environments A.T. Jahromi *et al* [7], N. Gordini and V. Vegliob [9] and K. Coussement and K.W. De Bock [57], it appears on the one hand, that in terms of the importance of variables used in the phase learning, recency and frequency variables seem less important for predicting churn. This goes in utter contrast with the expectations we have formulated from existing research, which strongly emphasize the predictive power of the RF variables of the RFM models. This result occurs mainly because the clients involved in the prediction models are the ones that represent the core customers and new high-value customers, more precisely, customers belonging to clusters \overline{LRFM} , \overline{LRFM} , \overline{LRFM} and \overline{LRFM} . Considering Table 4, one observes that the average F and R values of these four clusters are almost close. In addition, the descriptive statistics presented in Table 3 indicate that the standard deviation of F and R is low, which means that the values of these two variables are little dispersed around the average. Therefore, this makes their contribution less important in the distinction between total defectors customers, partial defectors and loyal ones.

Another explanation could reside in the fact that these studies have not fully exploited the large amount of data generated by online environments, but they have remained limited to the variables that characterize offline environments such as recency, frequency and the monetary. Indeed, from the events made by customers on the merchant site, we can easily extract many predictor variables that have an explanatory power in the understanding of customers' behavior, and in the analysis of their buying experience that starts with product consultation and ends with validation of the transaction. For example, we can retrieve variables that describe for each customer the rate of dropping sessions at different stages of the buying process.

5. CONCLUSION

In order to address the crucial problem of churn definition in the non-contractual (e-commerce) settings, LRFM model and clustering technique (k-means) are combined in the first stage to identify different types of customer profiles (different LRFM patterns) based on the first sub-period (T1). Consequently, we find seven clusters of customers that have a different LRFM behavior (\overline{LRFM} , \overline{LRFM} , \overline{LRFM} , \overline{LRFM} , \overline{LRFM} , \overline{LRFM} , \overline{LRFM}) and then we define a customer's LRFM pattern change over time as an early signal of either partial or total defection. In our opinion, the proposed methodology for churn definition can be a useful decision tool for companies operating in non-contractual settings, where customers and companies do not have any contracts between them. After resolving the problem of churn definition, we have proposed three

predictive models (Artificial Neural Networks, Simple Decision Tree, and Decision Tree Ensemble) for partial/total customer churn in thee-commerce sector.

In order to test the proposed models in a real context, we used as a case study an online store, where the click stream behavior records of customers for the period November 1, 2013 through February 28, 2015 have been utilized. The results reported reveal that three proposed models can provide an individual-level prediction of the probability to partially or totally defect in the future, that would enable us to follow the both partial and total defectors. A comparative analysis of different models is also presented, the results of this comparative analysis show the beneficial impact of Decision Tree Ensemble over other models (simple decision tree and artificial neural networks) in terms of prediction quality.

This prediction is very useful for marketing managers because will greatly help them to implement new tailored incentives solutions (retention actions) according to the degree of the defection (Partial or total) to convince them to stay.

Finally, the variables that can be contributing more to predicting partial and total customer's defection in e-commerce sector have been identified.

Our findings indicate, also, some limitations and issues for further research.

Firstly, this study is limited to e-commerce sector, and it is difficult to apply it in the offline world, because it's based on analyzing Web browsing behavior (page views, sequence of visits, buying process, session dropout rates at each stage of the buying process).

Secondly, we have used only a few numbers of variables in the clustering phase. However, further studies may utilize additional variables such as variables related to product category.

Finally, the predictive power of the model is significantly influenced by the choice of classification technique. However, in future work other classification techniques, such as genetic algorithms, naïve Bayes tree (NBTree), rough set approaches and fuzzy logic, will be used.

REFERENCES

- [1] Interbank Electronic Banking Center, Morocco, "Activité monétaire 1^{er} semestre 2017 au Maroc", [<https://www.cmi.co.ma/>]
- [2] Hongsheng Xu, *et al.*, "Construction of Ecommerce Recommendation System based on Semantic Annotation of Ontology and User Preference", *TELKOMNIKA (Telecommunication Computing, Electronics and Control)*, vol. 12, no. 3, pp. 2028-2035, 2014.
- [3] Neslin, S.A., Gupta, S., Kamakura, W., Lu, J., & Mason, C., "Defection detection: improving predictive accuracy of customer churn models", *Working Paper*, Teradata Center at Duke University, 2004.
- [4] Burez, J., and Van den Poel, D., "Crm at a pay-TV company: Using analytical models to reduce customer attrition by targeted marketing for subscription services", *Expert Systems with Applications*, vol. 32, pp. 277-288, 2007.
- [5] Y. Catherine, "AOL: Scrambling to Halt the Exodus," *Business Week*, 62, August 4, 2003.
- [6] Buckinx, W. and Van den Poel, D., "Customer base analysis: partial defection of behaviorally loyal clients in a non-contractual FMCG retail setting", *European Journal of Operational Research*, vol. 164, no. 1, pp. 252-268, 2005.
- [7] A.T. Jahromi, *et al.*, "Managing B2B customer churn, retention and profitability", *Industrial Marketing Management*, vol. 43, no. 7, pp. 1258-1268, October 2014
- [8] Özden Gür Ali, Umüt Arıtürk, "Dynamic churn prediction framework with more effective use of rare event data: The case of private banking", *Expert Systems with Applications*, vol. 41, no. 17, pp. 7889-7903, 2014.
- [9] N. Gordini and V. Veglio, "Customers churn prediction and marketing retention strategies. An application of support vector machines based on the AUC parameter-selection technique in B2B e-commerce industry", *Industrial Marketing Management*, vol. 62, pp. 100-107, April 2017.
- [10] V.L. Miguéis, *et al.*, "Modeling partial customer churn: On the value of first product-category purchase sequences", *Expert Systems with Applications*, vol. 39, no. 12 and 15, pp. 11250-11256, September 2012.
- [11] M. Clemente-Ciscar, *et al.*, "A methodology based on profitability criteria for defining the partial defection of customers in non-contractual settings", *European Journal of Operational Research*, vol. 239, no. 1, 16 November 2014.
- [12] Jae-Hyeon Ahna, Sang-Pil Hana, Yung-Seop Lee., "Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry", *Telecommunications Policy*, vol. 30, no. 10 and 11, pp. 552-568, 2006.
- [13] R. Ait daoud, *et al.*, "Customer Segmentation Model in E-commerce Using Clustering Techniques and LRFM Model: The Case of Online Stores in Morocco", *World Academy of Science, Engineering and Technology International Journal of Computer, Electrical, Automation, Control and Information Engineering*, vol. 9, no. 8, 2015.
- [14] Shin-Yuan Hung, *et al.*, "Applying data mining to telecom churn management", *Expert Systems with Applications*, vol. 31, pp. 515-524, 2006.
- [15] M.A.H. Farquad, Vadlamani Ravi, S. Bapi Raju, "Churn prediction using comprehensible support vector machine: An analytical CRM application", *Applied Soft Computing*, vol. 19, pp. 31-40, 2014

- [16] Mozer, *et al.*, “Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry”, *IEEE Transactions on Neural Networks*, vol. 11, no. 3, pp. 690-696, 2000.
- [17] Xiaobing Yu, *et al.*, “An extended support vector machine forecasting framework for customer churn in e-commerce”, *Expert Systems with Applications*, vol. 38, pp. 1425-1430, 2011.
- [18] Reinartz W1, Kumar V, “The mismanagement of customer loyalty”, *Harv Bus Rev*, vol. 80, no. 7, 125, pp. 86-94, 2002.
- [19] Xie, *et al.*, “Customer churn prediction using improved balanced random forests”, *Expert Systems with Applications*, vol. 36, no. 3, pp.5445-5449, 2009.
- [20] Bingquan, *et al.*, “Customer churn prediction in telecommunications”, *Expert Systems with Applications*, vol. 39, no. 1, pp. 1414-1425, January 2012.
- [21] Faris, H., “Neighborhood cleaning rules and particle swarm optimization for predicting customer churn behavior in telecom industry”. *Int. J. Adv. Sci. Technol.*, vol. 68, pp. 11-22, 2014.
- [22] K. Kyoungok, *et al.*, “Chi-Hyuk Jun, Jaewook Lee, Improved churn prediction in telecommunication industry by analyzing a large network”, *Expert Systems with Applications*, vol. 41, no. 15, 2014, pp. 6575-6584, 2014.
- [23] Ssu-Han Chen, “The gamma CUSUM chart method for online customer churn prediction”, *Electronic Commerce Research and Applications*, vol. 17, pp. 99-111, May–June 2016.
- [24] Niels Holtrop, Jaap E. Wieringa, Maarten J. Gijzenberg, Peter C. Verhoef, “No future without the past? Predicting churn in the face of customer privacy”, *International Journal of Research in Marketing*, vol. 34, no. 1, pp. 154-172, 2017.
- [25] H. H. Wu, *et al.*, “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, pp. 7, 2014.
- [26] A.M. Hughes, “Strategic database marketing”, *Probus Publishing*, 1994.
- [27] R. Ait daoud, *et al.*, “Combining RFM model and clustering techniques for customer value analysis of a company selling online”. *IEEE/ACS 12th International Conference of Computer Systems and Applications (AICCSA)*, 1-6, 2015.
- [28] Jo-Ting Wei, *et al.*, “Customer relationship management in the hairdressing industry: An application of data mining techniques”, *Expert Systems with Applications*, vol. 40, no. 18, pp. 7513-7518, 15 December 2013.
- [29] Der-Chiang Li, *et al.*, “A two-stage clustering method to analyze customer characteristics to build discriminative customer management: A case of textile manufacturing business”, *Expert Systems with Applications*, vol. 38, no. 6, pp. 7186-7191, June 2011.
- [30] A.M. Hughes, “Boosting response with RFM”. *Marketing Tools*, vol. 3, no. 3, pp. 4-7, 1996.
- [31] G.M. Marakas, “Decision Support Systems in the 21st Century”, *Second Edition. Prentice Hall, Upper Saddle River, NJ*, 2003.
- [32] A.X. Yang, “How to develop new approaches to RFM segmentation”, *Journal of Targeting, Measurement and Analysis for Marketing*, vol. 13, no. 1, pp. 50-60, 2004.
- [33] S.M.S. Hosseini, *et al.*, “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, pp. 5259-5264, 2010.
- [34] I.C. Yeh, *et al.*, “Knowledge discovery on RFM model using Bernoulli sequence”, *Expert Systems with Applications*, vol. 36, no. 3, pp. 5866-5871, 2009.
- [35] H.C. Chang and H.P. Tsai, “Group RFM analysis as a novel framework to discover better customer consumption behavior”, *Expert Systems with Applications*, vol. 38, no. 12, pp.14499-14513, 2011.
- [36] H.H. Chang and S.F. Tsay., “Integrating of SOM and K-mean in data mining clustering: an empirical study of CRM and profitability evaluation”, *Journal of Information Management*, vol. 11, no. 4, pp. 161-203, 2004.
- [37] S. Chow. and R. Holden., “Toward an understanding of loyalty: The moderating role of trust”, *Journal of Management issues*, vol. 9, no. 3, pp. 275-298, 1997.
- [38] D. Birant, “Data Mining Using RFM Analysis, Knowledge-Oriented Applications in Data Mining”, *InTech*, ISBN: 978-953-307-154-1, 2011.
- [39] Se-Hoon Jung, *et al.*, “ Prediction Data Processing Scheme using an Artificial Neural Network and Data Clustering for Big Data”, *International Journal of Electrical and Computer Engineering*, vol. 6, no. 1, pp. 330-336, February 2016.
- [40] R.J. Kuo, *et al.*, “Integration of self-organizing feature map and K-means algorithm for market segmentation”, *Computers & Operations Research*, vol. 29, no. 11, pp. 1475-1493, 2002.
- [41] Tan P.N., Steinbach M., Kumar V. “Introduction to Data Mining”, *Pearson Addison Wesley*; Boston, MA, pp. 487-556 ,USA: 2006.
- [42] Jiawei Han and Micheline Kamber, “Data Mining: Concepts and Techniques, 2nd ed”, *Morgan Kaufmann Publishers*, ISBN 1-55860-901-6, March 2006.
- [43] Peter J. Rousseeuw, “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis”, *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53-65, 1987.
- [44] S.H. Ha and S.C. Park., “Application of data mining tools to hotel data mart on the Intranet for database marketing”, *Expert Systems with Applications*, vol. 15, no. 1, pp. 1-31, 1998.
- [45] Rumelhart, *et al.*, “Learning internal representations by error propagation”, (vol. 1). *MA: MIT Press Cambridge*, 1986.
- [46] Y.L.Chen, *et al.*, “Constructing a multivalued and multi-labeled decision tree”, *Expert Systems with Applications*, vol. 25, no. 2, pp. 199-209, 2003.

- [47] Wei, C. -P., & Chiu, I. -T., "Turning telecommunications call details to churn prediction: A data mining approach". *Expert Systems with Applications*, vol. 23, no. 2, pp. 103-112, 2002.
- [48] Quinlan, J.R., "C4.5: Programs for machine learning", *Morgan Kaufman Publishers*, 1993.
- [49] Quinlan, J.R., "Improved use of continuous attributes in c4.5", *Journal of Artificial Intelligence Research*, vol. 4, pp. 77-90, 1996.
- [50] Dudoit, *et al.*, "Comparison of discrimination methods for the classification of tumors using gene expression data", *Journal of the American Statistical Association*, vol. 97, no. 457, pp. 77-87, 2002.
- [51] L. Breiman, "Random forests", *Machine Learning*, vol. 45, no. 1, pp. 5-32, 2001.
- [52] K. Coussement and D. Van den Poel, "Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques", *Expert systems with applications*, vol. 34, no. 1, pp. 313-327, 2008.
- [53] P.M. Guadagni, and J.D.C. Little., "A Logit Model of Brand Choice Calibrated on Scanner Data", *Marketing Science*, vol. 2, no. 3, pp. 203-238, 1983.
- [54] P. Rossi *et al.*, On the value of household purchase history information in target marketing, *Marketing Science*, vol. 15, no. 4, pp. 321-340, 1996.
- [55] Guofang Kuang and Yuanchen Li, "Using Fuzzy Association Rules to Design Ecommerce Personalized Recommendation System", *TELKOMNIKA (Telecommunication Computing, Electronics and Control)*, vol. 12, no. 2, pp. 1519-1527, 2014
- [56] Keaveney, S.M., & Parthasarathy, M. Customer switching behavior in online services: An exploratory study of the role of selected attitudinal, behavioral, and demographic factors. *Journal of the Academy of Marketing Science*, vol. 29, no. 4, pp. 374-390, 2001.
- [57] K. Coussement and K.W. De Bock, "Customer churn prediction in the online gambling industry: The beneficial effect of ensemble learning", *Journal of Business Research*, vol. 66, pp. 1629-1636, 2013.
- [58] Ligang Zhou, *et al.*, "One versus one multi-class classification fusion using optimizing decision directed acyclic graph for predicting listing status of companies", *Information Fusion*, vol. 36, pp. 80-89, 2017.
- [59] M. Sokolova and G. Lapalme., "A systematic analysis of performance measures for classification tasks", *Information Processing & Management*, vol. 45, no. 4, pp. 427-437, 2009.
- [60] A Sturm and Bob L, "Classification accuracy is not enough", *Journal of Intelligent Information Systems*, vol. 41, no. 3, December 2013, pp. 371-406, 2013.
- [61] J. Hadden, *et al.*, "Computer assisted customer churn management: State-of-the-art and future trends", *Computers & Operations Research*, vol. 34, no. 10, pp. 2902-2917, 2007.