Journal of Systems and Information Technology

Comparison of supervised machine learning techniques for customer churn prediction based on analysis of customer behavior

Samira Khodabandehlou, Mahmoud Zivari Rahman,

Comparison of supervised machine learning techniques for customer churn prediction based on analysis of customer behavior

Samira Khodabandehlou
Master of Information Technology Engineering, Tehran, Iran, and
Mahmoud Zivari Rahman
Allameh Tabataba'i University, Tehran, Iran

Received 3 October 2016
Revised 26 February 2017
Accepted 17 April 2017

Abstract

Purpose – This paper aims to provide a predictive framework of customer churn through six stages for accurate prediction and preventing customer churn in the field of business.

Design/methodology/approach – The six stages are as follows: first, collection of customer behavioral data and preparation of the data; second, the formation of derived variables and selection of influential variables, using a method of discriminant analysis; third, selection of training and testing data and reviewing their proportion; fourth, the development of prediction models using simple, bagging and boosting versions of supervised machine learning; fifth, comparison of churn prediction models based on different versions of machine-learning methods and selected variables; and sixth, providing appropriate strategies based on the proposed model.

Findings – According to the results, five variables, the number of items, reception of returned items, the discount, the distribution time and the prize beside the recency, frequency and monetary (RFM) variables (RFMITSDP), were chosen as the best predictor variables. The proposed model with accuracy of 97.92 per cent, in comparison to RFM, had much better performance in churn prediction and among the supervised machine learning methods, artificial neural network (ANN) had the highest accuracy, and decision trees (DT) was the least accurate one. The results show the substantially superiority of boosting versions in prediction compared with simple and bagging models.

Research limitations/implications – The period of the available data was limited to two years. The research data were limited to only one grocery store whereby it may not be applicable to other industries; therefore, generalizing the results to other business centers should be used with caution.

Practical implications – Business owners must try to enforce a clear rule to provide a prize for a certain number of purchased items. Of course, the prize can be something other than the purchased item. Business owners must accept the items returned by the customers for any reasons, and the conditions for accepting returned items and the deadline for accepting the returned items must be clearly communicated to the customers. Store owners must consider a discount for a certain amount of purchase from the store. They have to use an exponential rule to increase the discount when the amount of purchase is increased to encourage customers for more purchase. The managers of large stores must try to quickly deliver the ordered items, and they should use equipped and new transporting vehicles and skilled and friendly workforce for delivering the items. It is recommended that the types of services, the rules for prizes, the discount, the rules for accepting the returned items and the method of distributing the items must be prepared and shown in the store for all the customers to see. The special services and reward rules of the store must be communicated to the customers using new media such as social networks. To predict the customer behaviors based on the data, the future researchers should use the boosting method because it increases efficiency and accuracy of prediction. It is recommended that for predicting the customer behaviors, particularly their churning status, the ANN method be used. To extract and select the important and effective variables influencing customer behaviors, the discriminant analysis method can be used which is a very accurate and powerful method for predicting the classes of the customers.

Journal of Systems and Information Technology
Vol. 19 No. 1/2, 2017
pp. 65-93

1328-7265
DOI 10.1108/JSIT-10-2016-0061

Originality/value – The current study tries to fill this gap by considering five basic and important variables besides RFM in stores, i.e. prize, discount, accepting returns, delay in distribution and the number of items, so that the business owners can understand the role services such as prizes, discount, distribution and accepting returns play in retraining the customers and preventing them from churning. Another innovation of the current study is the comparison of machine-learning methods with their boosting and bagging versions, especially considering the fact that previous studies do not consider the bagging method. The other reason for the study is the conflicting results regarding the superiority of machine-learning methods in a more accurate prediction of customer behaviors, including churning. For example, some studies introduce ANN (Huang *et al.*, 2010; Hung and Wang, 2004; Keramati *et al.*, 2014; Runge *et al.*, 2014), some introduce support vector machine (Guo-en and Wei-dong, 2008; Vafeiadis *et al.*, 2015; Yu *et al.*, 2011) and some introduce DT (Freund and Schapire, 1996; Qureshi *et al.*, 2013; Umayaparvathi and Iyakutti, 2012) as the best predictor, confusing the users of the results of these studies regarding the best prediction method. The current study identifies the best prediction method specifically in the field of store businesses for researchers and the owners. Moreover, another innovation of the current study is using discriminant analysis for selecting and filtering variables which are important and effective in predicting churners and non-churners, which is not used in previous studies. Therefore, the current study is unique considering the used variables, the method of comparing their accuracy and the method of selecting effective variables.

Keywords RFM model, Bagging algorithm, Boosting algorithm, Churn prediction, Supervised machine learning techniques

Paper type Research paper

1. Introduction

Nowadays, due to the advances of technology, the change in selling approaches and the creation of competitive markets, the markets are characterized by supply surplus, causing the customer to be considered as the real ruler of the market. Therefore, commercial enterprises must shift from focusing on the products toward focusing on customers and manage their behaviors to prevent them from leaving and to obtain the highest profits and revenues for their own organizations (Dick and Basu, 2003; Lai, 2009; Payne and Frow, 2004). Customer relationship management (CRM) is the process of strict management of information on the customers and the appropriate management of all the customers to maximize their loyalty (Kincaid, 2003; Ngai *et al.*, 2009; Parvatiyar and Sheth, 2001; Umayaparvathi and Iyakutti, 2012). Its main goal is to create satisfaction and happiness in the customers to prevent them from churning, as it is the gravest danger threatening all the organizations because a small change in the level of customer retention will lead to significant changes in the shares and the profits of the organization (Agarwal *et al.*, 2004; Kotler and Keller, 2006; Swift, 2001; Van den Poel and Larivie're, 2004).

Maintaining the customers is the most basic and the most important issue for commercial entities, particularly, shopping centers. On the other hand, attracting new customers has a very high cost, sometimes five times the cost of retaining the current customers (Ali and Arıtürk, 2014; Hung and Wang, 2004; Marcus, 1998; Murakani and Natori, 2013; Reichheld and Sasser, 1990; Reichheld, 1993; Shoemaker and Bowen, 1998; Tamaddoni Jahromi *et al.*, 2014; Umayaparvathi and Iyakutti, 2012). However, many commercial organizations suffer from losing their valuable customers to the competition. This is known as customer churning (Huang *et al.*, 2012, 2010; Hung and Wang, 2004; Lu *et al.*, 2014; Umayaparvathi and Iyakutti, 2012; Yu *et al.*, 2011). Customer churning (customer attrition) is a marketing term in which the customer becomes interested in another organization or product (Chen, 2016; Coussement and Poel, 2009; Glady *et al.*, 2009; Hung and Wang, 2004; Keramati *et al.*, 2014; Yu *et al.*, 2011), leading to a reduction in sale revenues and an increase in the cost of attracting customers (Coussement and Poel, 2009; Haywood, 1988). Various studies have been carried out for predicting customer churn and its functions (Ali and Arıtürk, 2014; Chen, 2016; Hung and Wang, 2004; Keramati *et al.*, 2014; Kim *et al.*, 2014; Lu *et al.*, 2014; Tamaddoni Jahromi *et al.*, 2014; Tsai and Chen, 2010; Vafeiadis *et al.*, 2015; Yu *et al.*, 2011), indicating the significance of this issue. Based on the studies of marketing experts, many companies annually lose about 25 per cent of their customers on average (Chiang *et al.*, 2003). Sometimes, the statistics for customer churn have even been reported up to 36 per cent (Dursun and Caber, 2016). On the other hand, a 5 per cent reduction in customer churn will lead to 25 (Marcus, 1998; Reichheld and Sasser, 1990; Reichheld, 1993) to 85 per cent (Ivanovic *et al.*, 2011; Reichheld and Sasser, 1990) increase in revenues. These statistics indicate the role of "customer churn management" in ensuring the survival of the organization (Van den Poel and Larivie're, 2004). Accordingly, customer churn and its management is considered as a grave issue for various industries (Guo-en and Wei-dong, 2008).

One of the most important and basic industries is the food industry and the stores related to it because, nowadays, food stores play a significant role in countries' trade and are considered as one of the most important business centers in cities which attract a lot of customers. Therefore, the owners of these businesses are in a close competition to attract more customers and retain them. However, due to the presence of numerous stores and the intense competition among them, customer churn in these stores is higher than other businesses. As the majority of them do not have a clear understanding of their own customers, they cannot implement appropriate strategies and measures to obtain competitive advantage in the market. As they deal with every customer in a similar manner, they waste the lion's portion of their organizational resources and lose their revenues (Soudagar, 2012). Of course, one of the important reasons for this is the lack of strict analysis of customers' behaviors (Braun and Schweidel, 2011; Buckinx and Van den Poel, 2005; Hung and Wang, 2004). With the accurate prediction of customers' behaviors, the stores can have a better track record in retaining customers, providing services and providing better incentives to reduce the customer churn (Qiu *et al.*, 2015; Verbeke *et al.*, 2012).

If the companies do not distinguish clearly between the churning customers and those not churning, they will suffer great losses, as providing incentives to a non-churning customer instead of a real churning customer will waste organizational resources, and not only is the real churning customer not encouraged to stay but also the probability of their churning will increase. Therefore, companies must use models which accurately identify the customers who are in the risk of future churning (Tamaddoni Jahromi *et al.*, 2014). Hence, nowadays, business managers try to understand the important role of predicting customer churn in reaching success (Keramati *et al.*, 2014).

From a CRM point of view, managers must accurately and in a timely manner identify the reasons and the time of customer churn to define customer retention strategies and decide how much of the churn can be controlled (Braun and Schweidel, 2011; SAS Institute, 2000). Accordingly, nowadays in the competitive environment of the market and the shortening of the life cycle of businesses, the application of business intelligence (BI) for quicker and better decision-making and managing customer churn is increasing (Turban *et al.*, 2008). BI systems can provide business strategies to contribute to the organization's management of customer churn (Singh and Samalia, 2014). In this regard, data mining and machine learning are considered as important instruments in BI which can utilize the previous behaviors of customers in order to predict their patterns, behaviors and trends in order to better manage the customers (Huang *et al.*, 2010; Ivanovic *et al.*, 2011; Rygielski *et al.*, 2002) and provide commercial businesses with the possibility of knowledge-based decision making (Devi Prasad and Madhavi, 2012; Hung and Wang, 2004; Ngai *et al.*, 2009).

The main objective of predicting churn using BI tools (e.g. data mining and machine learning) is to classify the customers into two categories of churners and non-churners and provide targeted and efficient marketing strategies for potential churners (Singh and Samalia, 2014). Various machine-learning techniques such as logistic regression (LR), artificial neural networks (ANN), decision trees (DT) and support vector machine (SVM) have successfully been used to predict customer churn. In some cases, the results have been compared (Buckinx and Van den Poel, 2005; Devi Prasad and Madhavi, 2012; Guo-en and Wei-dong, 2008; Huang *et al.*, 2010,2012; Hung and Wang, 2004; Keramati *et al.*, 2014; Qureshi *et al.*, 2013; Vafeiadis *et al.*, 2015), some results indicate the similarity of their functionalities (Buckinx and Van den Poel, 2005) and some other results indicate the difference of these techniques (Devi Prasad and Madhavi, 2012; Guo-en and Wei-dong, 2008; Keramati *et al.*, 2014; Qureshi *et al.*, 2013; Vafeiadis *et al.*, 2015) in accurate prediction of customer churn. This shows the need for more research to obtain more reliable results.

Among behavioral variables, recency, frequency and monetary (RFM variables) are the best predictors for distinguishing loyal and non-loyal (potential churners) customers (Buckinx and Van den Poel, 2005; Coussement and De Bock, 2013; Dursun and Caber, 2016; Tamaddoni Jahromi *et al.*, 2014; Wei *et al.*, 2010) in a way that the RFM analysis is one of the best-known data mining methods for predicting customer behavior in the field of direct marketing (Glady *et al.*, 2009; Hughes, 1996; Madani, 2009; Sohrabi and Khanlari, 2007; Wei *et al.*, 2010, 2013). However, considering the results of previous studies, this model cannot accurately and completely determine the extent of loyalty (McCarty and Hastak, 2007; Miglautsch, 2000; Wang, 2010; Wei *et al.*, 2010). Accordingly, some studies have tried to expand or improve the RFM model to increase its prediction capability including RFMTC, time since first purchase and churn (Yeh *et al.*, 2009); RFC, cost (King, 2007); eRFM-EMO, demographic and emotional variables (Coussement and Poel, 2009); LRFM, length of customer relation (Chang and Tsay, 2004; Li *et al.*, 2011); RFM + I, influence (Murakani and Natori, 2013); and RFMP, period of product activity (Hosseini *et al.*, 2010). However, there are still other factors that can be effective along with RFM in identifying and evaluating customer behaviors such as loyalty or customer churn which have not been evaluated. These factors can vary based on the type of the company and its databases.

In this study, using the type of services and incentives provided for attracting customer in food stores registered in a database, in a case study we focus on expanding RFM. Accordingly, five factors including the number of purchased items, the number of returned items, the discount amount, the distribution time and the number of prizes have been added to the RFM parameters, creating the RFMITSDP model. Therefore, this study proposes a new framework for the process of identifying and predicting churners in the field of commercial businesses in which three classification methods of supervised machine learning including ANNs, SVM and DT are extensively used for the problem of customer churn, and their ability to predict the churners will be compared. Particularly, we compare the performance of these methods with their boosting and bagging versions to improve their performance in predicting the churning of customers. The main objective of the current study is to evaluate the appropriateness of the most advanced machine-learning methods for the churning problem where the effects of the selected parameters on customer churn will be compared based on these methods. This framework can identify the most accurate prediction method and the most effective variables for the future researchers as well as the business owners to be able to accurately predict the behaviors of the customers.

2. Related work

Hosseini *et al.* (2010) proposed a method for better identifying and classifying customers cased on an augmented RFM model (period of product activity). They showed that this augmented method provided better results for CRM compared to other common models. Yeh *et al.* (2009) expanded the RFM model into RFMTC (time since first purchase and churn probability) to predict the future behavior of the customer and the state of the churning and showed that the proposed method provides better accuracy than RFM model in predicting customer churn. Buckinx and Van den Poel (2005) studied retail stores using ANN and LR machine-learning methods to show that the RFM variables were the best predictors for customer churn. Moreover, they showed that besides the RFM model, the length of customer relationship, mode of payment, buying behavior across categories, usage of promotions and brand purchase behavior have an effective impact on predicting customer churn. Coussement and Poel (2009) tried to add the information on the emails of the call center to the traditional system of predicting customer churn, i.e. expanding RFM into eRFM-EMO to increase their accuracy in identifying churning customers. They used the SVM, LR and random forests classification methods to compare the accuracy of these two models. The results showed that eRFM-EMO provided a better performance compared to RFM, and the ensemble learner provided a better performance compared to other methods. Noyan and Simsek (2014) carried out a study to create a conceptual model for better understanding the customer behavior and the level of loyalty. In this study, structural equation modeling (SEM) was used for analyzing the data obtained from 1,530 customers in four large chain stores. The results of the study showed that customer loyalty depended on price, discount, the quality of the products, the quality of the services, value perception and the satisfaction, and among these variables, customer satisfaction was the most important factor influencing customer loyalty. Li *et al.* (2011) used the measures of the LRFM (relation length) model and clustering methods to analyze the customer characteristics for better management for improving CRM in textile industry. Based on the results, LRFM provides a better understanding of the customer for the organization to determine marketing strategies. Zakaria *et al.* (2014) studied the relationship between loyalty programs, customer satisfaction and customer loyalty in the retail industry. They showed that the store's partnership program, prizes, insurance and price had a significant impact on customer satisfaction and the store's partnership program, everyday programs of the members, discount and price had significant impact on customer loyalty. Murakani and Natori (2013) tried to improve CRM through the RFM + I (influence) analysis and showed that traditional customer management was focused on maximizing the profit obtained from customers, while using RFM + I analysis not only maximized profits but also contributed to attracting new customers and preventing the churning of current customers. Chan (2008) proposed an approach for better understanding of customer behaviors and classifying them based on RFM model, LTV (life time value) and genetic algorithm for managing the customers. Based on the obtained results, considering the relationship between the value of customers and recommendations, the customers with higher value could be identified for recommendation programs. Aggelis and Christodoulakis (2005) used RFM analysis to classify the customers and showed that the RFM model could easily identify the most important valuable customers.

Qureshi *et al.* (2013) used machine-learning techniques such as ANN, DT and linear and LR to predict the churning of customers of customer DNA website. The results showed that these methods, including DT, provide a good performance for identifying potential churners. To improve the performance of churning prediction models, Huang *et al.* (2010) propose a new set of variables with three input window techniques and use machine-learning methods including SVM, DT and ANN as the predictors. The results show that these new variables increase the accuracy of churning prediction and ANN and SVP provide a better performance compared to DT. Runge *et al.* (2014) try to compare the performance of four classification methods (SVM, DT, NN and LR) for churning prediction and show that ANN provides the best performance for churning prediction, and mutual relation can significantly impact the churning of players. Crone *et al.* (2006) showed that data preparation methods including sampling and coding of categories and attribute scaling had a significant impact on the prediction accuracy of ANN, SVM and DT classifiers. Tsai and Chen (2010) propose a method for selecting important variables for developing customer churning prediction models using ANN and DT machine-learning techniques. Based on the criteria of accuracy, precision, recall and *F*-measure, the results of the study show that selecting more effective variables causes DT and ANN models to provide a better performance in predicting the churning of customers. Umayaparvathi and Iyakutti (2012) investigate the application of ANN and DT techniques in predicting churning customers and the effects of variable selection on identifying churners. They showed that these techniques, particularly DT, provided a good accuracy for prediction and selecting related variables improved the performance of prediction methods. Guo-en and Wei-dong (2008) tried to improve the prediction capabilities of machine-learning methods using SVM to minimize the risks in predicting customer churn. The results show that compared to ANN, DT, LR and naive Bayesian (NB) classification, SVM provides the best performance and gives an effective measurement for predicting the churning of customers. Tamaddoni Jahromi *et al.* (2014) carried out a study to propose a method based on data mining for modeling the churning of customers based on RFM variables. They compared a number of modeling techniques based on their ability to predict churning customers. Their results show that the boosting method is the best method for churning prediction and developing better strategies for retaining customers in management decisions. Hung and Wang (2004) used machine-learning techniques for comparing the rate of "propensity-to-churn" for the customers. The results show that compared to DT, the ANN method can provide a more accurate model for predicting the churning of the customer using their demographic and transaction data. Huang *et al.* (2012) proposed a new set of variables for predicting the churning of customers and then used seven prediction techniques (LRs, linear classifications, NB, DT, ANN, SVMs and the evolutionary data mining algorithm) as the churning predictors based on the new variables. These show that these new variables are more effective than the current variables for predicting the churning of customers. Coussement and De Bock (2013) predicted the churning of customers based on RFM, length of relationship, inter-purchase time, gender and location variables using DT, generalized additive model and ensemble methods. The results show that predicting customer churn is a valuable strategy for identifying and profiling the set of at-risk customers. Moreover, the performance of ensemble methods is better and more powerful than single models. Keramati *et al.* (2014) use DT, ANN and SVM classification techniques for classifying the customers and show that using these techniques gives a 95 per cent accuracy of classification for recall and precision, and ANN has a significantly better performance compared to the other three methods, KNN, DT and SVM. Vafeiadis *et al.* (2015) try to compare machine-learning methods for predicting the churning of customers to identify the best machine learning technique in predicting the churning of customers in the telecommunication industry. The results show that the boosting versions are better than non-boosted (simple) methods, and SVM-POLY is the best classifier. Lu *et al.* (2014) use the boosting algorithm for clustering and LR as a basic learner for predicting the churning of customers. The results show that compared to LR, boosting models provide a better understanding of the churning state of customers.

A summary of the results of these studies is presented in Table I. Based on the above material, the studies carried out on the field of customer churning can be divided into the following categories:

- Regarding the type of business, some of the studies predict churning in the retail (Buckinx and Van den Poel, 2005; Chan, 2008; Chen, 2016; Coussement and Poel,

Table I. Customer churn prediction models

References	Dataset	Methods	Churn def.
Ali and Arıtürk (2014)	Banking	Independently trained binary, Multinomial and Ordinal LR, and Survival analysis	Customer with portfolio size below a specific threshold value during a period
Buckinx and Van den Poel (2005)	Retailing (FMCG)	ANN (ARD), LR, and ensemble learner (Random forest)	Customer changes purchasing patterns during a period
Chen (2016)	Retailing	gamma CUSUM chart	Customer without a single login during a period
Coussement and De Bock (2013)	Online gambling	DT (CART), GAM, and ensemble learners (random forests, GAMens)	Customer not having played during a period
Coussement and Poel (2009)	Retailing (newspaper)	SVM (RBF), LR, and ensemble learner (random forests)	Customer does not renew their product during a period
Crone *et al.* (2006)	Publishing	SVM (RBF), ANN (MLP), and DT (C4.5)	Not mentioned
Devi Prasad and Madhavi (2012)	Banking	DT (CART, C5.0)	Customer closes one of accounts
Glady *et al.* (2009)	Banking	DT (Cost-sensitive), ANN (MLP), LR, and ensemble learner (AdaCost)	Customer with a CLV decreasing over time
Guo-en and Wei-dong (2008)	Telecommunication	SVM (linear), ANN (BPANN), DT (C4.5), LR, and NB	Customer does not enjoy all services of company
Hosseini *et al.* (2010)	Retailing (automotive)	K-Means	Not mentioned
Huang *et al.* (2010)	Land-line telecommunication	ANN (MLP), DT (C4.5), and SVM (Polynomial, RBF)	Customer leaves company
Huang *et al.* (2012)	Telecommunication	DT(C4.5), ANN (MLP), LR, NB, SVM (RBF), Linear Classifiers (LC), and Data Mining by Evolutionary Learning (DMEL)	Customer shifts to competitors
Hung and Wang (2004)	Wireless telecommunication	DT (C5.0), and ANN (BPN)	Subscribers switches to a competitor
Keramati *et al.* (2014)	Telecommunication	ANN, KNN, SVM (Polynomial, RBF, MLP), and DT	Not mentioned
Li *et al.* (2011)	Wholesale (textile)	K-Means	Not mentioned
Lu *et al.* (2014)	Mobile telecommunication	LR, and ensemble learner (boosting)	Subscriber switches to another service provider during a period

(*continued*)

Table I.

References	Dataset	Methods	Churn def.
Noyan and Simsek (2014)	Retailing (supermarket)	Structural Equation Modeling (SEM)	Not mentioned
Qureshi *et al.* (2013)	Online Retailing	DT (CHAID, Exhaustive CHAID, CART, QUEST), ANN (MLP), and linear and LR	Customer terminates to use the network
Runge *et al.* (2014)	Online games	ANN, SVM (RBF), DT, LR, and HMM	Player permanently leaves game
Tamaddoni Jahromi *et al.* (2014)	Retailing (FMCG)	DT (simple, cost-sensitive, CART), LR, and ensemble learner (boosting)	Customer has no purchase in prediction period
Tsai and Chen (2010)	Mobile telecommunication	ANN (MLP), DT (C5.0), and Association rules	Customer terminate services of company
Umayaparvathi and Iyakutti (2012)	Telecommunication	DT, and ANN	Customer switches to another service provider
Vafeiadis *et al.* (2015)	Telecommunication	ANN (MLP), SVM (Polynomial, RBF), DT (C5.0), NB, LR, and ensemble learner (boosting)	Not mentioned
Yeh *et al.* (2009)	Blood Transfusion service	Bernoulli sequence	Customer discontinues his use of a service for ever after a period
Yu *et al.* (2011)	Online retailing	ANN (BP), DT (C4.5), and SVM (Linear, Polynomial, RBF)	Customer shifts to a competitor

2009; Hosseini *et al.*, 2010; Noyan and Simsek, 2014; Qureshi *et al.*, 2013; Tamaddoni Jahromi *et al.*, 2014; Yu *et al.*, 2011; Zakaria *et al.*, 2014), telecommunication (Guo-en and Wei-dong, 2008; Huang *et al.*, 2010, 2012; Hung and Wang, 2004; Keramati *et al.*, 2014; Lu *et al.*, 2014; Qureshi *et al.*, 2013; Runge *et al.*, 2014; Tsai and Chen, 2010; Umayaparvathi and Iyakutti, 2012; Vafeiadis *et al.*, 2015), banking (Aggelis and Christodoulakis, 2005; Ali and Arıtürk, 2014; Devi Prasad and Madhavi, 2012; Glady *et al.*, 2009), wholesale (Hosseini *et al.*, 2010; Li *et al.*, 2011), online games (Runge *et al.*, 2014), online gambling (Coussement and De Bock, 2013) and publishing (Crone *et al.*, 2006). There has been a little focus on customer churning in food stores.

- The majority of studies use various machine-learning methods including ANNs (Buckinx and Van den Poel, 2005; Crone *et al.*, 2006; Glady *et al.*, 2009; Guo-en and Wei-dong, 2008; Huang *et al.*, 2010,2012; Hung and Wang, 2004; Keramati *et al.*, 2014; Qureshi *et al.*, 2013; Runge *et al.*, 2014; Tsai and Chen, 2010; Umayaparvathi and Iyakutti, 2012; Vafeiadis *et al.*, 2015; Yu *et al.*, 2011), SVM (Coussement and Poel, 2009; Crone *et al.*, 2006; Guo-en and Wei-dong, 2008; Huang *et al.*, 2010,2012; Keramati *et al.*, 2014; Runge *et al.*, 2014; Vafeiadis *et al.*, 2015; Yu *et al.*, 2011), DT (Coussement and De Bock, 2013; Crone *et al.*, 2006; Devi Prasad and Madhavi, 2012; Guo-en and Wei-dong, 2008; Huang *et al.*, 2010; Huang *et al.*, 2012; Hung and Wang, 2004; Keramati *et al.*, 2014; Qureshi *et al.*, 2013; Runge *et al.*, 2014; Tsai and Chen, 2010; Umayaparvathi and Iyakutti, 2012; Yu *et al.*, 2011), LR (Buckinx and Van den Poel, 2005; Coussement and Poel, 2009; Glady *et al.*, 2009; Guo-en and Wei-dong, 2008; Huang *et al.*, 2012; Qureshi *et al.*, 2013; Tamaddoni Jahromi *et al.*, 2014; Tsai and Chen, 2010; Vafeiadis *et al.*, 2015), ensemble learners (Buckinx and Van den Poel, 2005; Coussement and De Bock, 2013; Coussement and Poel, 2009; Glady *et al.*, 2009; Lu *et al.*, 2014; Tamaddoni Jahromi *et al.*, 2014; Vafeiadis *et al.*, 2015) and NB (Guo-en and Wei-dong, 2008; Huang *et al.*, 2012; Vafeiadis *et al.*, 2015). As can be seen, the majority of studies use a number of methods simultaneously.
- These studies vary based on the variables used for investigating the churning or loyalty of customers. Some of them only use behavioral variables, RFM, (Aggelis and Christodoulakis, 2005; Chan, 2008; Coussement and De Bock, 2013; Tamaddoni Jahromi *et al.*, 2014), and some others use augmented RFM (Buckinx and Van den Poel, 2005; Coussement and De Bock, 2013; Coussement and Poel, 2009; Hosseini *et al.*, 2010; Li *et al.*, 2011; Murakani and Natori, 2013; Yeh *et al.*, 2009).
- The majority of studies define customer churning differently based on the type of the business, the revenues and the profits obtained from the business. For instance, people who leave the organization (Huang *et al.*, 2010; Qureshi *et al.*, 2013; Runge *et al.*, 2014; Tsai and Chen, 2010; Yeh *et al.*, 2009), closing one of the accounts (Devi Prasad and Madhavi, 2012), a customer with decreasing customer lifetime value (CLV) (Glady *et al.*, 2009), switching to the competitors (Huang *et al.*, 2012; Hung and Wang, 2004; Lu *et al.*, 2014; Umayaparvathi and Iyakutti, 2012; Yu *et al.*, 2011) and changing the purchasing pattern (Buckinx and Van den Poel, 2005).

Based on these results, it can be said that studies on data mining and customer churning have been approached from different directions by the researchers, e.g. prediction and classification methods, type of business, variables affecting the churning or the loyalty of the customer and the type of churning definition. However, there is a gap in these studies; studies trying to improve the performance of the RFM model and expanding it only evaluate a limited number of variable (one or two variables) (Hosseini *et al.*, 2010; Li *et al.*, 2011;

Murakani and Natori, 2013; Yeh *et al.*, 2009), while there are other effective variables which are not considered. The current study tries to fill this gap by considering five basic and important variables besides RFM in stores, i.e. prize, discount, accepting returns, delay in distribution and the number of items so that the business owners can understand the role services such as prizes, discount, distribution and accepting returns play in retraining the customers and preventing them from churning. Another innovation of the current study is the comparison of machine-learning methods with their boosting and bagging versions, especially considering the fact that previous studies do not consider the bagging method. The other reason for the study is the conflicting results regarding the superiority of machine-learning methods in more accurate prediction of customer behaviors, including churning. For example, some studies introduce ANN (Huang *et al.*, 2010; Hung and Wang, 2004; Keramati *et al.*, 2014; Runge *et al.*, 2014), some introduce SVM (Guo-en and Wei-dong, 2008; Vafeiadis *et al.*, 2015; Yu *et al.*, 2011) and some introduce DT (Freund and Schapire, 1996; Qureshi *et al.*, 2013; Umayaparvathi and Iyakutti, 2012) as the best predictor, confusing the users of the results of these studies regarding the best prediction method. The current study identifies the best prediction method specifically in the field of store businesses for researchers and the owners. Moreover, another innovation of the current study is using discriminant analysis for selecting and filtering variables which are important and effective in predicting churners and non-churners, which is not used in previous studies. Therefore, the current study is unique considering the used variables, the method of comparing their accuracy and the method of selecting effective variables.

3. Supervised machine learning methods

Selecting an appropriate algorithm for prediction is one of the most important steps (Buckinx and Van den Poel, 2005). In this section, three important and practical techniques for predicting the churning of customers are briefly introduced, whose reliability, performance and functionality have been proved in a great number of studies (Buckinx and Van den Poel, 2005; Freund and Schapire, 1996; Hung and Wang, 2004; Keramati *et al.*, 2014; Qureshi *et al.*, 2013; Runge *et al.*, 2014; Vafeiadis *et al.*, 2015; Yu *et al.*, 2011).

3.1 Artificial neural network

ANN is one of the most successful machine-learning methods for solving complex problems including the prediction of customer churning (Huang *et al.*, 2010; Hung and Wang, 2004; Vafeiadis *et al.*, 2015). Neural networks such as multi-layer perceptron (MLP) and radial basis function (RBF) are among the most popular and the most commonly used feed forward neural networks with supervised learning (Huang *et al.*, 2010, 2012; Vafeiadis *et al.*, 2015) which consist of a number of layers which are completely connected and the information moves from a layer to another layer (from left to right) (Crone *et al.*, 2006; Heykin, 1999). For the problem of predicting customer churning, Keramati *et al.* (2014) show that ANN has a significantly better performance compared to DT and SVM methods. Other studies also show the performance superiority of ANN compared to DT (Huang *et al.*, 2010; Hung and Wang, 2004; Yu *et al.*, 2011) and sometimes SVM (Keramati *et al.*, 2014; Runge *et al.*, 2014) for predicting churning.

3.2 Support vector machine

SVM is a supervised learning method which is capable of solving linear and on-linear classification problems (Crone *et al.*, 2006; Guo-en and Wei-dong, 2008). This method tries to maximize the distance between super planes and classes to move the classes apart as much as possible (Tomar and Agarwal, 2015; Vapnik, 1995). The core function of SVM is a key factor for the decision-making function and its functionality (Vapnik, 1995; Yu *et al.*, 2011). Polynomial and RBF cores are usually selected as the core functions (Huang *et al.*, 2010, 2012; Vafeiadis *et al.*, 2015; Yu *et al.*, 2011). SVM is considered as a very promising method for predicting customer churning (Huang *et al.*, 2010; Yu *et al.*, 2011), and it has gained very successful experimental experiences in this field in a way that many studies show that this method provides a better performance compared to DT (Guo-en and Wei-dong, 2008; Huang *et al.*, 2010; Vafeiadis *et al.*, 2015; Yu *et al.*, 2011).

3.3 Decision trees

Decision tree is one of the most popular and common classification techniques (Juan *et al.*, 2007; Keramati *et al.*, 2014), and one of the most important reasons behind its popularity is the fact that it is flexible and easy to understand (Tamaddoni Jahromi *et al.*, 2014). This method is known as the "divide and conquer" method which is used for creating a binary tree (Huang *et al.*, 2012). DT C5.0 is a popular and widely used supervised machine-learning machine for creating DT (Juan *et al.*, 2007; Williams *et al.*, 2012) because it provides the best results with high accuracies (Juan *et al.*, 2007). C5.0 is widely used in predicting the customer churning (Devi Prasad and Madhavi, 2012; Hung and Wang, 2004; Tsai and Chen, 2010), and considering the type of the data, it can provide a good performance as well as accurate models for churning prediction (Devi Prasad and Madhavi, 2012; Hung and Wang, 2004; Qureshi *et al.*, 2013; Vafeiadis *et al.*, 2015).

4. Ensemble learners

In ensemble learners, the predictions (results) of classifiers are combined, which increases their efficiency (Coussement and De Bock, 2013; Freund and Schapire, 1996; Simidjievski, Todorovski, and Džeroski, 2015). There are a lot of popular ensemble learners including boosting and bagging methods. Boosting is an ensemble method used for improving the functionality of learning algorithm which tries to significantly increase the accuracy of learning algorithms (Freund and Schapire, 1996; Lu *et al.*, 2014; Simidjievski *et al.*, 2015). Boosting is a general method for obtaining an accurate prediction by combining a number of predictions (with lower accuracies) in which the models are trained using completely different training data (in a way that the samplings are conducted without replacement meaning that the samples which are chosen in the first time are removed from the original data and the next sampling is done without the presence of first samples and the third without the presence of the first two samples and so on) (Freund and Schapire, 1996; Simidjievski *et al.*, 2015). Bagging (bootstrap aggregation) is an ensemble method for creating multiple versions of the predictor; it, then, uses these versions to obtain an integrated prediction. (It is also called "sampling with replacement", in which each time the sampling is done, the population elements are returned to the original data and next time the new random sampling is done with the previous selected elements included again in the sampling process.) (Freund and Schapire, 1996; Quinlan, 1996). This method increases the stability of the models (Quinlan, 1996; Simidjievski *et al.*, 2015). For problems of predicting customer churning, ensemble learners have been successfully used due to their ability to increase the accuracy and stability of the models (Bock and Van, 2011; Coussement and De Bock, 2013; Glady *et al.*, 2009; Lu *et al.*, 2014; Tamaddoni Jahromi *et al.*, 2014; Vafeiadis *et al.*, 2015).

5. Evaluation measures

One of the most important steps for ensuring that a churning prediction model can be properly generalized is to evaluate its performance (Huang *et al.*, 2012). In fact, evaluation criteria measure the capability of a prediction model for accurate ranking of customers based on the probability of churning (Coussement and De Bock, 2013). There are well-known criteria for evaluating the accuracy or the performance of a predictor model; criteria including accuracy, precision, recall and the *F*-measure, which can successfully show the performance of the predictor (Coussement and De Bock, 2013; Keramati *et al.*, 2014). To evaluate the performance of machine-learning methods in predicting churning, we have used these criteria, which are calculated based on the confusion matrix shown in Table II. There are four terms which should be defined to get familiar with these criteria:

(1) *TP (True positives)*: the number of customers that should be in the churner category and the prediction algorithm has determined their category correctly as churner.
(2) *TN (True negatives)*: the number of customers that should be in the non-churner category and the prediction algorithm has determined their category correctly as non-churner.
(3) *FP (False positives)*: the number of customers who are non-churners but the algorithm incorrectly categorized them as churners.
(4) *FN (False negatives)*: the number of customers who are churners but the algorithm incorrectly categorized them as non-churner.

Recall is the ratio of real churners which are correctly identified, and it is calculated using equation (1):

$$\text{Recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \tag{1}$$

Precision is the ratio of predicted churners which are correct, and it is calculated using equation (2):

$$\text{Precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})} \tag{2}$$

Accuracy is the number of all the correct predictions, and it is calculated using Equation (3):

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{FP} + \text{TN} + \text{FN})} \tag{3}$$

F-measure is the harmonic average of precision and recall, and it is calculated using Equation (4):

$$\text{F-measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precisio} + \text{Recall}} \tag{4}$$

Table II. Confusion matrix for binary classification of churn

Actual	Predicted Churner	Non-churner
Churner	TP	FN
Non-churner	FP	TN

6. Methodology

6.1 Data set description

The presence of real and high-quality data plays an important role in indicating the reasons for customer churning (Hadden *et al.*, 2005). The data set used in this study includes the real data from a food store in Iran from March 25, 2013 to March 11, 2015 (about 24 months). This data set involves 1,050 customers (761 non-churners and 287 churners) and includes 577,200 records. In this store, churner refers to a customer whose recency is more than two months; this measure is based on the opinion of 12 experts, owners and stakeholders of the store whose rate of agreement in underlying this criterion to identify the churning customers has been 0.92. Table III presents the data set fields of customer transactions and their descriptions.

6.2 Data preparation

Data preparation (a.k.a. data preprocessing) is sometimes necessary in predicting churning, and it requires 60-70 per cent of the total time (Devi Prasad and Madhavi, 2012; Umayaparvathi and Iyakutti, 2012). This step includes three stages of cleaning, integration and transformation of the data.

6.2.1 Data cleaning. The process of cleaning the data involves completing the missing values, identifying and eliminating the outlier – data, and resolving conflicts among the data (Huang *et al.*, 2012). This cleaning process is carried out by eliminating the entire record, eliminating the data or replacing it with another value. After this step, 5,050 records were identified as missing data and deleted from the data set, reducing the total number of records in the data set from 577,200 records to 572,150 records.

Variables	Description
Customer ID	An ID given to each one of the customers
Order ID (OI)	An ID given to the each order (transaction) of the customer
Order Date (OD)	The date the customer orders a product or carries out a transaction
Item ID	The ID given to each product item
Total number (TN)	The total number of a certain product item purchased by the customer in a transaction
Total amount (TA)	The total amount paid for a certain product purchased by the customer in a transaction
Discount amount (DA)	The cash discount given to the customer during the purchase of a certain item through a transaction
Prize number (PN)	Based on the pre-defined calculations of the store, for buying a certain number of some of the items, a number of items will be given for free to the customer as a prize
Return number (RN)	The number of a certain item returned by the customer and accepted by the seller in a transaction
Distribution date (DD)	As the store distributes the items, this date is the date the items are distributed and delivered to the customers
Distance	The distance between the customer and the store
Gender	The gender of the customer
Group	There are three groups for customers in this data set: (1) consumer (the customer is not a seller), (2) retailer (the customer is a retail seller), (3) wholesaler (the customer is a wholesale seller)
Debt	In case the customer owes some money to the store and has not paid the entire bill
Education level	The education level of the customer
Status	Being a churner or a non-churner

Table III. Definition of data set variables

6.2.2 Data integration. To better understand the data and managing them more scientifically in each data set, data integration (combining or integrating two or more data sets together) is necessary, which is considered the second stage of data preparation (Soudagar, 2012). The data set of customers' transactions was integrated into the demographic data set (age, gender, education level, distance and group).

6.2.3 Data transformation. In this stage, the data will be transformed into another shape which is more appropriate for data mining and creating the prediction models (Soudagar, 2012). In this stage, string variables in the data set were transformed into numerical variables and numbers.

6.3 Derived variables

Derived variables are new variables created after data preparation based on the main variables of the dataset; they indicate the purchasing behavior of the customer. These variables contain important information and explain the customer's behavior better than the main variables. These variables are used more effectively by the prediction models in predicting the customer behavior (Umayaparvathi and Iyakutti, 2012). Table IV shows the newly created variables and the method of their calculation.

In Table IV, the variable i indicates the ith item; the variable j shows the number of times the ith item has been bought by the customer; TN indicates the total number of items; RN indicates the number of returned items; DA shows the discount amount, PN indicates the number of prizes; DD indicates the distribution date; OD indicates the order date; LOD indicates the last order date for the customer; CD indicates the current date; OI indicates the order ID; and finally, TA indicates the total amount.

6.4 Training and testing data sets

To show the real capabilities of a prediction algorithm, it is best to repeat the training and testing phases multiple times to reduce the risk of false identification in a certain model (Coussement and De Bock, 2013). Considering the fact that there is no exact formula or resource regarding the ratio of the training data and the test data for creating an optimal and accurate prediction model, based on Table V, we selected four different ratios for training and testing the model and selected the training and test data randomly (Tsai and Chen, 2010); in a way that 40 per cent of the data were selected randomly for testing in the first phase and the remaining 60 per cent were selected for training, and in the second phase 30 per cent were selected randomly for testing and the remaining 70 per cent for training, and so on. As the ratio of churner to non-churner in the training and test data must be similar to that of the main data set, we used chi-squared test to make sure of this ($p > 0.05$).

6.5 Variable selection

Before training a model using the common machine-learning algorithms, variable selection can be one of the most important factors affecting the performance of prediction models based on prediction rates (high TP and low FP) (Crone *et al.*, 2006; Huang *et al.*, 2012; Kim *et al.*, 2014). If a powerful set of variables (more related variables) can be selected in this stage, the TP prediction rate can be significantly increased and the FP prediction rate can be significantly decreased, leading to a significant improvement (Huang *et al.*, 2010,2012). As there are various methods for identifying effective variables in predicting the customer churning, and there is no one perfect method (Kim and Yoon, 2004), in this study, we use discriminant analysis for selecting effective variables. Discriminant analysis is a method for predicting the group membership based on independent variables, whose results are used for calculating the detection function. Its main goal is to identify the important and effective

Table IV. Definition and calculation of derived variables

Derived (New) variables	Calculation formula	Description
Total number of Items (I)	$\mathrm{I} = \sum_{i=1}^{i=n}\sum_{j=1}^{j=m} TN_i$	The total number of items purchased by the customer in the selected period
Total number of returned items (T)	$\mathrm{T} = \sum_{i=1}^{i=n}\sum_{j=1}^{j=m} RN_i$	The total number of items the customer retuned to the store after the purchase which were accepted by the store
Total discount amount (S)	$\mathrm{S} = \sum_{i=1}^{i=n}\sum_{j=1}^{j=m} DA_i$	The total amount of discounts received by the customer in the selected period
Total number of prizes (P)	$\mathrm{p} = \sum_{i=1}^{i=n}\sum_{j=1}^{j=m} PN_i$	Total prizes received by the customer in the selected period
Average delay of distribution (D)	$\mathrm{D} = \frac{\sum_{i=1}^{i=n}\sum_{j=1}^{j=m}(DD - OD)_i}{n^*\mathrm{m}}$	The time difference between the order date and the distribution date of each purchase was calculated and the average of this time difference is considered
Recency (R)	R = (LOD) – (CD)	The time difference between the date of the last purchase of the customer and the current date
Frequency (F)	F = Distinct count of (OI)	The total number of times the customer visited the store for a purchase
Monetary (M)	$\mathrm{M} = \sum_{i=1}^{i=n}\sum_{j=1}^{j=m} TA_i$	The total amount of money the customer has paid in the selected time period for purchasing items from the store

variables in accurate prediction of group membership such as churners or non-churners. In other words, the discriminant analysis allows the identification of the best combination of predicting variables for predicting group membership.

Based on Table VI, the *p*-value of RFMITSDP variables is lower than 0.05, indicating that these variables have the capability of predicting the churning status of the customers. In line with this result in Table VII, the eigenvalues, Wilks' lambda, canonical correlation and chi-square indicate the suitability of selected variables for predicting the churning of customers. Based on Table VIII, the ability to correctly predict the customer groups by RFMITSDP is 92.5 per cent. Accordingly, besides RFM variables, five other variables, which play a role in predicting the churning status of customers, are selected for the analysis and other variables are eliminated.

Table V. Chi-squared test for assessing the ratio of selected random samples to the main data

Dataset	Proportion (%)	Training/Testing	No. of samples	Chi-squared test	*p*-value
1	40	Testing	417	0.095	0.785
	60	Training	633	0.099	0.753
2	30	Testing	304	0.099	0.753
	70	Training	746	0.001	0.999
3	20	Testing	229	0.212	0.645
	80	Training	821	0.025	0.867
4	15	Testing	159	0.024	0.877
	85	Training	891	0.001	0.999

Table VI. Discriminant analysis for identifying effective variables for churning prediction

Variable	Wilks' lambda	*F*	df1	df2	*p*-value
Total number of Items (I)	0.851	183.35	1	1,046	0.000
Total discount of amount (S)	0.867	160.13	1	1,046	0.000
Total number of returned items (T)	0.837	203.97	1	1,046	0.000
Total number of prizes (P)	0.783	289.54	1	1,046	0.000
Average delay of distribution (D)	0.873	152.43	1	1,046	0.000
Recency (R)	0.404	1544.43	1	1,046	0.000
Frequency (F)	0.845	191.32	1	1,046	0.000
Monetary (M)	0.86	43.37	1	1,046	0.000
Group	1	0.034	1	1,046	0.853
Distance	1	0.009	1	1,046	0.925
Debt	1	0.173	1	1,046	0.678
Item type	0.999	0.932	1	1,046	0.334
Gender	1	0.005	1	1,046	0.945
program	1	0.361	1	1,046	0.548

Table VII. Eigenvalue and chi-square statistics for assessing the suitability of the selected detection function

Function	Eigenvalue	Canonical correlation	Wilks' lambda	Chi-square	df	Significance
1	1.612[a]	0.786	0.383	997.752	14	0.000

6.6 Prediction models construction

The main task in this stage is to use suitable modeling techniques for the accurate prediction of churning (Huang *et al.*, 2010). There have been many techniques proposed for churning prediction; in this study, state-of-the-art supervised machine-learning methods have been selected for evaluating the selected variables and predicting customer churning, based on which we create a number of churning prediction models and compare their ability to identify real churners.

The process of creating the prediction model involves training and testing the model. The general framework of the study is depicted in Figure 1. These prediction models are created based on SVM, ANN and DT methods. In ANN, the feed-forward neural networks, MLP and RBF, are used due to their reliability and practicality in churning prediction problems. In the SVM method, the polynomial and RBF cores are selected due to their high-quality results in related works (Huang *et al.*, 2010; Huang *et al.*, 2012; Tsai and Chen, 2010; Vafeiadis *et al.*, 2015). Moreover, to create ANN models, some core-related parameters including the number of hidden layer units, must be set to prevent overfit. Similarly, in SVM, some parameters related to the cores including the stopping criteria and gamma must be set based on the problem; their default values are considered to prevent the over-fitting problem (the default software settings were considered for all parameters and no change was made by the researchers). In DT, the C5.0 algorithm is again selected due to its common use in related works (Devi Prasad and Madhavi, 2012; Hung and Wang, 2004; Tsai and Chen, 2010). In this study, ensemble methods (boosting and bagging) are used for improving the performance and stability of the prediction models.

7. Experimental results

7.1 Evaluating the accuracy of recency, frequency and monetary model

In this section, the average accuracy of the RFM model (over four testing data sets) in predicting the churning of customers is discussed using DT (C5.0), ANN (RBF, MLP), SVM (RBF, Polynomial), and their boosting, bagging and simple versions based on criteria including precision, recall, accuracy and *F*-measure. Considering the significance and

			Predicted group membership		
		Status	Churner	non-Churner	Total
Original	Count	Churner	231	56	287
		Non-churner	23	738	761
	%	Churner	80.5	19.5	100.0
		Non-churner	3.0	97.0	100.0

Note: a. 92.5% of original grouped cases correctly classified

Table VIII. Ratio of correct prediction of customer status in the selected detection function

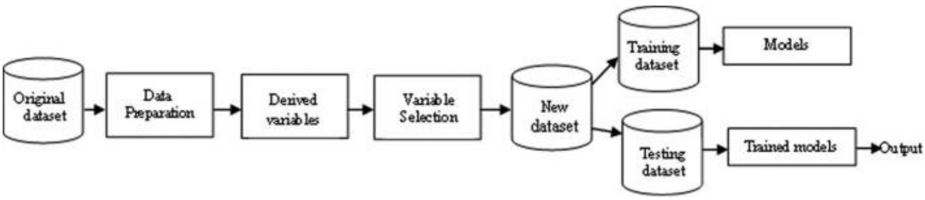

Figure 1. The general framework of the study

importance of the value of the *F*-measure in showing the real accuracy level, the analysis of variance (ANOVA) test is calculated for this criterion; the *F* values and the *p*-value are presented in Table IX. Based on the results, the maximum prediction accuracy of the RFM model based on the *F*-measure is 78.28. Considering the values of *F* (larger than the f critical that is 4.26) and *p*-value (less than 0.05), we can say that the accuracy of different methods had no significant differences, and boosting was of the highest and simple of the lowest accuracy in all versions.

In Table X, the mean and standard deviation of the *F*-measure for each different version of various machine-learning methods based on the RFM model are presented. Based on this table, the ANN-RBF method with the mean of 75.72 and the standard deviation of 2.67 has the highest *F*-measure and the DT-C5.0 method with a mean of 68.90 and the standard deviation of 2.69 has the lowest *F*-measure value. To make sure of the real difference of *F*-measure among these methods, the ANOVA tests is used.

In Table XI, considering the *p*-value (lower than 0.05), the average *F*-measure in various versions of machine learning methods was not identical, and there was a significant difference between them. In Table XII, to compare the mean difference among the methods, Tukey post hoc test was used. The results show that the accuracy difference between ANN-MLP and SVM-Poly and DT-C5.0; between ANN-RBF and SVM-Poly; SVM-RBF and DT-C5.0; and between ANN-RBF and DT-C5.0 is significant ($p < 0.05$). However, there is no significant difference between other methods. Therefore, the accuracies of ANN-MLP and

Method	Version	Precision	Recall	Accuracy	*F*-measure	*F*-ANOVA	*P*-value
ANN-MLP	Boosting	79.31	74.19	68.08	76.67	30.33	0.001
	Bagging	77.83	72.81	66.12	75.24		
	Simple	76.62	70.97	64.17	73.68		
ANN-RBF	Boosting	81.19	75.56	80.36	78.28	13.68	0.002
	Bagging	77.78	74.19	67.77	75.94		
	Simple	76.65	69.58	63.52	72.95		
SVM-Poly	Boosting	77.4	69.58	63.84	73.12	16.21	0.001
	Bagging	74.24	67.74	80.59	70.84		
	Simple	73.37	67.28	59.61	70.19		
SVM-RBF	Boosting	79.48	71.43	67.77	75.24	12.29	0.003
	Bagging	75.38	69.12	62.21	72.11		
	Simple	69.77	69.12	57	69.44		
DT-C5.0	Boosting	75.65	67. 28	61.56	71.30	11.99	0.003
	Bagging	73.33	65.90	58.96	69.42		
	Simple	70.31	62.21	54.02	66		

Table IX.
Average accuracy of RFM in boosting, bagging and simple versions of machine-learning methods

				95% confidence interval for mean			
Method	*N*	Mean	SD	Lower bound	Upper bound	Minimum	Maximum
ANN-MLP	12	74.86	2.53	73.25	76.47	70.00	78.34
ANN-RBF	12	75.72	2.67	74.05	77.39	71.00	80.00
SVM-Poly	12	71.38	1.48	70.44	72.32	69.76	73.76
SVM-RBF	12	72.26	2.89	70.42	74.10	69.00	78.00
DT-C5.0	12	68.90	2.69	67.20	70.61	64.00	73.00
Total	60	72.63	3.45	71.73	73.52	64.00	80.00

Table X.
Mean and standard deviation of f-measure criterion for RFM model in various versions of machine learning methods

ANN-RBF are about the same, which was the highest, DT-C5.0 has the lowest accuracy, and the accuracy of SVM-RBF and SVM-Poly had no significant difference and were almost at the same level.

7.2 Evaluating the accuracy of RFMITDSP model

Based on Table XIII, the maximum accuracy of the RFMITDSP based on the *F*-measure is 97.92 and considering the *p*-value (lower than 0.05), the accuracies of the methods differ. In RFMITDSP similar to RFM, the boosting version has the highest accuracy, the simple version has the lowest accuracy, and the bagging version is somewhere in between.

In Table XIV, the ANN-MLP method with the mean value of 95.59, and the standard deviation of 2.20 has the highest *F*-measure and the DT-C5.0 with the mean value of 89.63 and the standard deviation of 3.20 has the lowest *F*-measure. Considering the results in Table XV, the average *F*-measure differs for various versions of machine-learning methods, and there is a significant difference between them ($p \leq 0.05$). Based on the results in Table XVI, the accuracy difference between DT-C5.0 and ANN-MLP, ANN-RBF, SVM-Poly and SVM-RBF is significant ($p\text{-value} << 0.05$). However, the difference between other methods is not significant ($p > 0.05$).

7.3 Comparing the accuracy of recency, frequency and monetary and RFMITDSP models

RFM and RFMITDSP variables based on machine-learning methods are compared to independent samples test, and based on the *p*-values ($p < 0.05$) in Table XVII, there is a significant difference between their prediction accuracy in various versions of machine-learning methods. Based on the obtained mean, the prediction accuracy of the RFMITDSP model in all the methods is higher than the RFM model.

	Sum of squares	df	Mean square	*F*	*p*-value
Between groups	361.300	4	90.325	14.514	0.000
Within groups	342.275	55	6.223		
Total	703.575	59			

Table XI. ANOVA Statistic for comparing f-measure value of machine learning methods in RFM model

(I) Method	(J) Method	Mean difference (I-J)	*p*-value	95% confidence interval Lower bound	Upper bound
ANN-MLP	ANN-RBF	−0.86	0.915	−3.73	2.01
	SVM-Poly	3.48*	0.010	0.61	6.35
	SVM-RBF	2.6	0.094	−0.27	5.47
	DT-C5.0	5.6*	0.000	3.08	8.83
ANN-RBF	SVM-Poly	4.34	0.001	1.47	7.21
	RBF	3.46	0.011	0.59	6.33
	DT-C5.0	6.82*	0.000	3.94	9.69
SVM-Poly	SVM-RBF	−0.88	0.909	−3.75	1.99
	DT-C5.0	2.48	0.122	−0.39	5.35
SVM-RBF	DT-C5.0	3.36*	0.014	0.48	6.23

Note: *The mean difference is significant at the 0.05 level

Table XII. Tukey post hoc test for assessing the difference between machine learning methods in RFM model

Table XIII. Average accuracy of RFMITDSP in boosting, bagging and simple versions of machine-learning methods

Method	Version	Precision	Recall	Accuracy	*F*-measure	*F*-ANOVA	*p*-value
ANN-MLP	Boosting	98.15	97.69	97.07	97.92	20.36	0.001
	Bagging	95.83	95.39	93.81	95.61		
	Simple	94.34	92.16	90.55	93.24		
ANN-RBF	Boosting	96.73	95.39	94.46	95.05	5.28	0.03
	Bagging	95.35	94.47	92.38	94.91		
	Simple	92.92	90.78	88.6	91.84		
SVM-Poly	Boosting	95.81	94.93	93.48	95.37	49.81	0.001
	Bagging	95.39	95.39	93.48	95.39		
	Simple	89.61	95.39	88.92	92.41		
SVM-RBF	Boosting	95.77	94.01	92.83	94.88	56.84	0.001
	Bagging	94.42	93.55	91.53	93.98		
	Simple	88.44	91.7	85.67	90.04		
DT-C5.0	Boosting	92.66	93.09	89.9	92.87	91.12	0.001
	Bagging	92.31	88.47	86.64	90.36		
	Simple	87.5	83.87	80.13	85.65		

Table XIV. Mean and standard deviation of the *F*-measure criterion for RFMITDSP model in various versions of machine-learning methods

				95% confidence interval for mean			
Method	*N*	Mean	SD	Lower bound	Upper bound	Minimum	Maximum
ANN-MLP	12	95.59	2.20	94.19	96.99	92.96	100
ANN-RBF	12	93.93	2.10	92.59	95.27	91	98
SVM-Poly	12	94.39	1.52	93.42	95.36	92	96
SVM-RBF	12	92.97	2.28	91.52	94.41	89	95
DT-C5.0	12	89.63	3.20	87.59	91.66	85	94
Total	60	93.30	3.03	92.52	94.08	85	100

Table XV. ANOVA Statistic for comparing *F*-measure in various versions of machine learning methods in RFMITDSP model

	Sum of squares	df	Mean square	*F*	*p*-value
Between groups	245.253	4	61.313	11.325	0.000
Within groups	297.775	55	5.414		
Total	543.029	59			

8. Conclusions

The findings of the study involve four parts: the first part is related to selecting important and effective variables in customer churning; the second and the third parts are related to comparing the capability and the prediction accuracy of supervised machine-learning methods with their different versions based on RFM and augmented RFM (RFMITSDP), respectively; and the fourth part involves the comparison between RFM and augmented RFM models.

The first part of the findings shows that based on discriminant analysis (which is one of the innovations related to the selection of important variables for churning prediction) among the initial 14 factors, 8 factors played a role in customer churning; three of these factors were the famous RFM, and the other five factors were the number of prizes, the acceptance of returned items, discount, distribution date of the item after being ordered by the customer and the number of purchased items. These eight factors together can correctly predict the churning status of the customers in 92.5 per cent of the cases. However, the other six factors, distance between the customer and the store, customer group (consumer, retailer, wholesaler), the type of purchased item, debt status, gender and education level play no role in customer churning. It is therefore concluded that financial issues and extra services have a more significant impact on the churning of customers than demographic features.

In the second part of the findings, based on Table IX, among the five machine-learning methods, the ANN-RBF method provides the highest prediction accuracy and the maximum accuracy of RFM is 78.28 per cent. Based on the obtained results, the accuracy level in five machine-learning methods varies based on the version. Accordingly, the mean values show that for all the methods, their boosting version (compared to the bagging and simple versions) provides a better performance. In other words, when various samples are selected without substitution and the samples selected in the first time are not considered the second time, the prediction will be more accurate. Therefore, the best method for predicting customer churning based on the RFM variables is the boosting version of the algorithms.

The third part of the results of the current study is related to the RFMITSDP model, and its highest prediction accuracy is 97.92 per cent. Based on Table XIII, the ANN-MLP method with the boosting version provides the highest prediction accuracy. Based on the ANOVA results, the accuracies of the five machine learning methods vary based on their versions, and considering the mean values in all the methods, the boosting version provides a better performance, while their simple version has a lower prediction power for customer churning. Therefore, the boosting version improves the prediction accuracy. The results in Tables XIV, XV and XVI show that the ANN-MLP method has the highest accuracy, while the DT-C5.0 has the lowest accuracy; the accuracies of these five methods are not on the same level. Therefore, in the RFMITSDP model, the highest accuracy is provided by ANN, and the lowest accuracy is provided by DT.

In the fourth part of the research findings, we compared the accuracy of RFM and RFMITSDP models. Based on the results in Table XVII, there is a significant difference between the accuracy level of RFM and RFMITSDP model in all the methods; in all the methods, the RFMITSDP model has more power and higher prediction accuracy compared to RFM and the highest difference between these two models is 23.90 per cent for the

(I) Method	(J) Method	Mean difference (I-J)	*P*-value	95% confidence interval Lower bound	Upper bound
ANN-MLP	ANN-RBF	1.65	0.416	−1.02	4.33
	SVM-POLY	1.2	0.714	−1.48	3.88
	SVM-RBF	2.62	0.058	−0.05	5.308
	C5.0	5.96*	0.000	3.28	8.64
ANN-RBF	SVM-POLY	−0.46	0.989	−3.13	2.22
	SVM-RBF	0.97	0.846	−1.71	3.64
	C5.0	4.31*	0.000	1.63	6.98
SVM-Poly	SVM-RBF	1.42	0.568	−1.25	4.10
	C5.0	4.76*	0.000	2.08	7.42
SVM-RBF	C5.0	3.34*	0.008	0.66	6.02

Note: *The mean difference is significant at the 0.05 level

Table XVI. Tukey post hoc test for assessing the difference between machine-learning methods in RFMITDSP model

Table XVII.
Independent t-test statistic for comparing RFM and RFMITDSP models

								95% confidence interval of the difference	
Method	Model	N	Mean	SD	t	p-value	Mean difference	Lower	Lower
ANN-MLP	RFM	12	71.69	6.06	12.84	0.001	−23.90	−27.75	−20.04
	RFMITSDP	12	95.59	2.21					
ANN-RBF	RFM	12	75.72	2.63	18.74	0.001	−18.21	−20.22	−16.19
	RFMITSDP	12	93.93	2.10					
SVM-Poly	RFM	12	71.38	1.48	37.44	0.001	−23.007	−24.28	−21.73
	RFMITSDP	12	94.39	1.57					
SVM-RBF	RFM	12	72.26	2.89	19.47	0.001	−20.70	−22.91	−18.50
	RFMITSDP	12	92.97	2.28					
DT-C5.0	RFM	12	68.91	2.69	17.17	.001	−20.72	−23.22	−18.22
	RFMITSDP	12	89.63	3.21					

ANN-MLP method. Accordingly, expanding the RFM model can increase the prediction accuracy of customer churning about 24 per cent.

9. Discussion

Based on the obtained results, along with the RFM variables which have been considered extensively in previous studies, five other factors including the number of prizes, the level of accepting returned items, discount, the distribution date of the item and the number of purchased items can play a role in the customer behavior, particularly being a churner or a non-churner; these findings are in line with the findings of Noyan and Simsek (2014) who believe discount and service quality play a significant role in customer's loyalty, and they are also in line with the findings of Zakaria *et al.* (2014) who argue that prizes and discounts increase the loyalty of customers. Accordingly, when the store considers a prize for buying a certain number of items to encourage customers to purchase more of that item and the prize is significant, and just from the customer's point of view, the customer is less likely to churn. On the other hand, if a store accepts returned items without charging a fee, the customer will understand that if they buy an item and after a while they want to return it due to reasons such as not needing it, not using it, or because of its fault and the store will accept it, they will be very unlikely to churn from such a store. Another factor which can reduce the churning of customers is the discount considered for the customers based on a certain amount of purchase. If the store gives a discount for a certain amount of purchase and when the amount of purchase is increased, the discount is also increased, the selling as well as the customer loyalty will increase and the probability of churning will decrease. Based on the results, another factor affecting the churning of customers is the distribution date of the item after the customer orders it. If the store delivers the item immediately after the order with a free delivery, the customers will be encouraged to purchase more from the store and be loyal to it. Therefore, the proper management of product distribution, particularly for those customers who are sellers themselves and order a lot of items, is an integral issue for retaining such customers. The last factor affecting the status of customer churning is the number of items purchased by the customer. The higher the number of items purchased by the customer, the less likely the customer will be to churn. The reason for this may be that by increasing the number of purchased items, the number of prizes is also increased, the customer can benefit from more discount and also the probability of quicker delivery for a large number of items will be higher because the store considers him or her a very important customer and knows that the person who orders a large number of items is probably a retailer or wholesaler and the quick delivery of the ordered items is very important; quicker services will increase the loyalty and decrease the probability of churning.

In comparing various machine-learning methods with their boosting, bagging and simple versions, the boosting version provides higher accuracy and prediction power, which is probably because the sampling is without substitution while in the bagging method, the sampling is done with substitution. In this method, any sample which is chosen is returned to the data set, and it is possible that the sample is chosen again; this causes some items to be selected in each iteration of sampling, while some other items are never chosen, which leads to losing real information and repetitions. The results of the previous studies show that boosting is a suitable and powerful method for increasing the accuracy of predicting customer behavior, particularly customer churning (Tamaddoni Jahromi *et al.*, 2014; Lu *et al.*, 2014; Vafeiadis *et al.*, 2015; Coussement and De Bock, 2013). However, it is worth mentioning that in the reviewed literature, the bagging method is not considered, and it is not compared to other methods. However, in the current study, these three methods are studied and compared simultaneously, which is one of the innovations of the current study.

Based on the results of the study, compared to SVM and particularly DT, the ANN method is more power and provides a higher accuracy; this becomes more apparent by adding new and effective variables, indicating the higher power of this method compared to other methods in learning patterns which helps it to predict with a higher accuracy. The precious studies confirm these results (Huang *et al.*, 2010; Hung and Wang, 2004; Keramati *et al.*, 2014; Runge *et al.*, 2014).

Based on the other results of the study, the prediction accuracy of customer churning based on the RFMITSDP model compared to the RFM model is 23.90 per cent higher, indicating the important role played by variables other than RFM in customer churning. By considering the three variables of RFM, the average prediction probability of customers' churning behavior based on the best machine learning method is 71.69; however, by considering the five new variables, prizes, discount, the number of purchased items, accepting returned items and the distribution date of the items, along with the RFM variables, the prediction accuracy can be increased to 97.92, indicating the fact that if salespeople and business owners decide to make decisions about customer behavior solely based on RFM variables, they will encounter a 28 per cent error. However, by considering the other five variables mentioned earlier, the error rate will drop down to less than 3.5 per cent, which is very significant and can be very useful in decision making about the customers' behaviors. The previous studies confirm that expanding the RFM model can increase the prediction accuracy of customer behavior (Buckinx and Van den Poel, 2005; Coussement and Poel, 2009; Hosseini *et al.*, 2010; Li *et al.*, 2011; Murakani and Natori, 2013; Yeh *et al.*, 2009). Accordingly, adding important and basic variables can be very effective in analyzing customer behavior.

10. Implications for practice

- Business owners must try to enforce a clear rule to provide a prize for a certain number of purchased items. Of course, the prize can be something other than the purchased item.
- Business owners must accept the items returned by the customers for any reasons and the conditions for accepting returned items and the deadline for accepting the returned items must be clearly communicated to the customers.
- Store owners must consider a discount for a certain amount of purchase from the store. They have to use an exponential rule to increase the discount when the amount of purchase is increased in order to encourage customers for more purchase.
- The managers of large stores must try to quickly deliver the ordered items; they should use equipped and new transporting vehicles and skilled and friendly workforce for delivering the items.
- It is recommended that the types of services, the rules for prizes, the discount, the rules for accepting the returned items and the method of distributing the items be prepared and shown in the store for all the customers to see.
- The special services and reward rules of the store must be communicated to the customers using new media such as social networks.
- To predict the customers' behaviors based on the data, the future researchers should use the boosting method because it increases efficiency and accuracy of prediction.
- It is recommended that for predicting the customers' behaviors, particularly their churning status, the ANN method be used.

- To extract and select the important and effective variables influencing customers' behaviors, the discriminant analysis method can be used which is a very accurate and powerful method for predicting the classes of the customers.

11. Future research

- In future research based on RFM model for analyzing the customers' behaviors, other variables should also be considered and analyzed to gain a better understanding of the customer behavior.
- It is recommended that future research works use data with long-term periods of three or more years.
- The pair-wise comparison of the accuracy of learning machines.

12. Research limitations

- The period of the available data was limited to two years.
- The research data being limited to only one grocery store whereby it may not be applicable to other industries, so generalizing the results to other business centers should be used with caution.

References

Agarwal, A., Harding, D.P. and Schumacher, J.R. (2004), "Organizing for CRM", *McKinsey Quarterly*, Vol. 3, pp. 80-91.

Aggelis, V. and Christodoulakis, D. (2005), "Customer clustering using RFM analysis", *9th WSEAS International Conference on Computers, Special Session Data Mining, Techniques and Application*, *Wisconsin*.

Ali, O.G. and Arıtürk, U. (2014), "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.

Bock, K.W.D. and Van, D.P.D. (2011), "An empirical evaluation of rotation-based ensemble classifiers for customer Churn prediction", *Expert Systems with Applications*, Vol. 38 No. 10, pp. 12293-12301.

Braun, M. and Schweidel, D.A. (2011), "Modeling customer lifetimes with multiple causes of churn", *Marketing Science*, Vol. 30 No. 5, pp. 881-902.

Buckinx, W. and Van den Poel, D. (2005), "Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting", *European Journal of Operational Research*, Vol. 164 No. 1, pp. 252-268.

Chan, C.C.H. (2008), "Intelligent value-based customer segmentation method for campaign management: a case study of automobile retailer", *Elsevier Conference on Expert Systems with Applications*, Vol. 34 No. 4, pp. 2754-2762.

Chang, H.H. and Tsay, S.F. (2004), "Integrating of SOM and K-Meanin data mining clustering: an empirical study of CRM and profitability evaluation", *Journal of Information Management*, Vol. 11 No. 4, pp. 161-203.

Chen, S.H. (2016), "The gamma CUSUM chart method for online customer churn prediction", *Electronic Commerce Research and Applications*, Vol. 17 No. 1, pp. 99-111.

Chiang, D.A., Wang, Y.F., Lee, S.L. and Lin, C.J. (2003), "Goal-oriented sequential pattern for network banking churn analysis", *Expert Systems with Applications*, Vol. 25 No. 3, pp. 293-302.

Coussement, K. and De Bock, K.W. (2013), "Customer churn prediction in the online gambling industry: the beneficial effect of ensemble learning", *Journal of Business Research*, Vol. 66 No. 9, pp. 1629-1636.

Coussement, K. and Poel, D.V. (2009), "Improving customer attrition prediction by integrating emotions from client/company interaction emails and evaluating multiple classifiers", *Expert Systems with Applications*, Vol. 36 No. 1, pp. 6127-6134.

Crone, S.F., Lessmann, S. and Stahlbock, R. (2006), "The impact of preprocessing on data mining: an evaluation of classifier sensitivity in direct marketing", *European Journal of Operational Research*, Vol. 173 No. 3, pp. 781-800.

Devi Prasad, U. and Madhavi, S. (2012), "Prediction of churn behavior of bank customers using data mining tools", *Business Intelligence Journal*, Vol. 5 No. 1.

Dick, A.S. and Basu, k. (2003), "Customer loyalty: toward an integrated conceptual framework", *Journal of the Academy of Marketing Science*, Vol. 22 No. 2, pp. 99-113.

Dursun, A. and Caber, M. (2016), "Using data mining techniques for profiling profitable hotel customers: an application of RFM analysis", *Tourism Management Perspectives*, Vol. 18 No. 1, pp. 153-160.

Freund, Y. and Schapire, R.E. (1996), "Experiments with a new boosting algorithm", *13th International Conference on Machine Learning*, *Murray Hill, NJ*, pp. 148-156.

Glady, N., Baesens, B. and Croux, C. (2009), "Modeling churn using customer lifetime value", *European Journal of Operational Research*, Vol. 197 No. 1, pp. 402-411.

Guo-en, X. and Wei-dong, J. (2008), "Model of customer churn prediction on support vector machine", *Systems Engineering–Theory and Practice*, Vol. 28 No. 1, pp. 71-77, doi: 10.1016/S1874-8651(1009)60003-X.

Hadden, J., Tiwari, A., Roy, R. and Ruta, D. (2005), "Computer assisted customer churn management: state-of-the-art and future trends", *Computers and Operations Research*, Vol. 34 No. 10, pp. 2902-2917.

Haywood, K.M. (1988), "Repeat patronage: cultivate ingalliances with customers", *International Journal of Hospitality Management*, Vol. 7 No. 3, pp. 225-237.

Heykin, S. (1999), *Neural Networks a Comprehensive Foundation*, 2nd ed., Prentice Hall, Upper Saddle River, NJ, pp. 14-29.

Hosseini, S.M.S., Maleki, A. and Gholarmian, M.R. (2010), "Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty", *Expert Systems with Applications*, Vol. 37 No. 7, pp. 5259-5264.

Huang, B.Q., Kechadi, T.M., Buckley, B., Kiernan, G., Keogh, E. and Rashid, T. (2010), "A new feature set with new window techniques for customer churn prediction in land-line telecommunications", *Expert Systems with Applications*, Vol. 37 No. 5, pp. 3657-3665.

Huang, B., Kechadi, M.T. and Buckley, B. (2012), "Customer churn prediction in telecommunications", *Expert Systems with Applications*, Vol. 39 No. 1, pp. 1414-1425.

Hughes, A.M. (1996), "Boosting response with RFM", *Marketing Tools*, Vol. 3 No. 3, pp. 4-10.

Hung, S.Y. and Wang, H.Y. (2004), "Applying data mining to telecom churn management", *Pacific Asia Conference on Information Systems (PACIS)*.

Ivanovic, S., Mikinac, K. and Perman, L. (2011), "CRM development in hospitality companies for the purpose of increasing the competitiveness in the tourist market", *UTMS Journal of Economics*, Vol. 2 No. 1, pp. 59-68.

Juan, G., Luo, S., Jia, H., Zhang, T. and Han, Y. (2007), "Type 2 diabetes data processing with EM and C4.5 Algorithm", *IEEE Complex Medical Engineering Internaonal Conference*, *Beijing*, pp. 371-377.

Keramati, A., Jafari-Marandi, R., Aliannejadi, M., Ahmadian, I., Mozzafari, M. and Abbasi, U. (2014), "Improved churn prediction in telecommunication industry using data mining technique", *Applied Soft Computing*, Vol. 24, pp. 994-1012.

Kim, H.S. and Yoon, C.H. (2004), "Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market", *Telecommunications Policy*, Vol. 28 Nos 9/10, pp. 751-765.

Kim, K., Jun, C.H. and Lee, J. (2014), "Improved churn prediction in telecommunication industry by analyzing a large network", *Expert Systems with Applications*, Vol. 41 No. 15, pp. 6575-6584.

Kincaid, J. (2003), *Customer Relationship Management: Getting It Right*, Prentice-Hall PTR, New Jersey, NJ.

King, S.F. (2007), "Citizens as customers: exploring the future of CRM in UK local government", *Government Information Quarterly*, Vol. 24 No. 1, pp. 47-63.

Kotler, P. and Keller, L. (2006), *Marketing Management*, 2nd ed., Pearson Prentice Hall, New Jersey, NJ.

Lai, X. (2009), "Segmentation study on enterprise customers based on data mining technology", *IEEE First International Workshop on Database Technology and Applications*, Wuhan, Hubei, pp. 247-250.

Li, D.C., Dai, W.L. and Tseng, W.T. (2011), "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. 1, pp. 7186-7191.

Lu, N., Lin, H., Lu, J. and Zhang, G. (2014), "A customer churn prediction model in telecom industry using boosting", *IEEE Transactions on Industrial Informatics*, Vol. 10 No. 1, pp. 1659-1665.

McCarty, J.A. and Hastak, M. (2007), "Segmentation approaches in data mining: a comparision of RFM, CHAID, and logistic regression", *Journal of Business Research*, Vol. 60 No. 6, pp. 656-662.

Madani, S. (2009), "Mining changes in customer purchasing behavior -a data mining approach", Master Thesis, Lulea University, Luleå University of Technology.

Marcus, C. (1998), "A practical yet meaningful approach to customer segmentation", *Journal of Consumer Marketing*, Vol. 15 No. 5, pp. 494-504.

Miglautsch, J.R. (2000), "Thoughts on RFM scoring", *Journal of Database Marketing & Customer Strategy Management*, Vol. 8 No. 1, pp. 67-72.

Murakani, K. and Natori, S. (2013), "New customer management technique: CRM by 'RFM + I' analysis", *Nomura Research Institute*, Vol. 186.

Ngai, E.W.T., Xiu, L. and Chau, D.C.K. (2009), "Application of data mining techniques in customer relationship management: a literature review and classification", *Expert Systems with Applications*, Vol. 36 No. 2, pp. 2592-2602.

Noyan, F. and Simsek, G.G. (2014), "The antecedents of customer loyalty", *Procedia- Social and Behavioral Sciences*, Vol. 109 No. 1, pp. 1220-1224. *World Conference on Business, Economics and Management (WC-BEM 2013)*.

Parvatiyar, A. and Sheth, J. (2001), "Customer relationship management: emerging practice, process and discipline", *Journal of Economic and Social Research*, Vol. 3 No. 2, pp. 1-34.

Payne, A. and Frow, P. (2004), "The role of multichannel integration in customer relationship management", *Industrial Marketing Management*, Vol. 33 No. 6, pp. 527-538.

Qiu, J., Lin, Z. and Li, Y. (2015), "Predicting customer purchase behavior in the e-commerce context", *Electronic Commerce Research*, Vol. 15 No. 4, doi: 10.1007/s10660-015-9191-6.

Quinlan, R. (1996), "Bagging predictors", *Machine Learning*, Vol. 24 No. 2, pp. 123-140.

Qureshi, S.A., Rehman, A.S., Qamar, A.M., Kamal, A. and Rehman, A. (2013), "Telecommunication subscribers' churn prediction model using machine learning", *Eighth International Conference on Digital Information Management (ICDIM), IEEE, Islamabad*, pp. 131-136.

Reichheld, F.F. (1993), *Loyalty-Based Management*, Harvard Business Review, Brighton Watertown, MA, pp. 64-73.

Reichheld, F.F. and Sasser, E.W. (1990), "Zero defections: quality comes to services", *Harvard Business Review*, Vol. 68 No. 5, pp. 105-111.

Runge, J., Gao, P., Garcin, F. and Faltings, B. (2014), "Chum prediction for high-value players in casual social games", *Proceeding of 2014 IEEE Conference on Computational Intelligence and Games, IEEE Computer Society Press, Washington, DC.*

Rygielski, C., Wang, J. and Yen, D. (2002), "Data mining techniques for customer relationship management", *Technology in Society*, Vol. 24 No. 4, pp. 483-502.

SAS Institute (2000), "Best price in Churn prediction", A SAS Institute White Paper.

Shoemaker, S. and Bowen, J.T. (1998), "Loyalty: a strategic commitment", *Cornell Hotel and Restaurant Administration Quarterly*, Vol. 39 No. 1, pp. 12-25.

Simidjievski, N., Todorovski, L. and Džeroski, S. (2015), "Predicting long-term population dynamics with bagging and boosting of process-based models", *Expert Systems with Applications*, Vol. 42 No. 22, pp. 8484-8496.

Singh, H. and Samalia, H.V. (2014), "A business intelligence perspective for churn management", *Procedia-Social and Behavioral Sciences*, Vol. 109, pp. 51-56. *2nd World Conference on Business, Economics And Management*.

Sohrabi, B. and Khanlari, A. (2007), "Customer lifetime value (CLV) measurement based on RFM model", *Iranian Accounting and Auditing Review*, Vol. 14 No. 47, pp. 7-20.

Soudagar, R. (2012), "Customer segmentation and strategy definition in segments", Master Thesis, Luleå University of Technology.

Swift, R. (2001), *Accelerating Customer Relationships Using CRM and Relationship Technologies*, Prentice-Hall PTR, New Jersey, NJ.

Tamaddoni Jahromi, A., Stakhovych, S. and Ewing, M. (2014), "Managing B2B customer churn, retention and profitability", *Industrial Marketing Management*, Vol. 43 No. 7, pp. 1258-1268.

Tomar, D. and Agarwal, S. (2015), "A comparison on multi-class classification methods based on least squares twin support vector machine", *Knowledge-Based Systems*, Vol. 81, pp. 131-147.

Tsai, C.F. and Chen, M.Y. (2010), "Variable selection by association rules for customer churn prediction of multimedia on demand", *Expert Systems with Applications*, Vol. 37 No. 1, pp. 2006-2015.

Turban, E., Leidner, D., McLean, E. and Wetherbe, J. (2008), "Information technology for management", *Transforming Organizations in the Digital Economy*, 6th ed., John Wiley and Sons, Hoboken, NJ.

Umayaparvathi, V. and Iyakutti, K. (2012), "Applications of data mining techniques in telecom churn prediction", *International Journal of Computer Applications*, Vol. 42 No. 20, pp. 0975-8887.

Vafeiadis, T., Diamantaras, K.I., Sarigiannidis, G. and Chatzisavvas, K.C. (2015), "A comparison of machine learning techniques for customer churn prediction", *Simulation Modelling Practice and Theory*, Vol. 55, pp. 1-9.

Van den Poel, D. and Larivie're, B. (2004), "Customer attrition analysis for financial service using proportional hazard models", *European Journal of Operational Research*, Vol. 157 No. 1, pp. 196-217.

Vapnik, V.N. (1995), *The Nature of Statistical Learning Theory*, Springer Verlag, New York, NY.

Verbeke, W., Dejaeger, K., Martens, D., Hur, J. and Baesens, B. (2012), "New insights into churn prediction in the telecommunication sector: a profit driven data mining approach", *European Journal of Operational Research*, Vol. 218 No. 1, pp. 211-229, available at: http://dx.doi.org/210.1016/j.ejor.2011.1009.1031

Wang, C.H. (2010), "Apply robust segmentation to the service industry using kernel induced fuzzy clustering techniques", *Expert Systems with Applications*, Vol. 37 No. 12, pp. 8395-8400.

Wei, J.T., Lee, M.C., Chen, H.K. and Wu, H.H. (2013), "Customer relationship management in the hairdressing industry: an application of data mining techniques", *Expert Systems with Applications*, Vol. 40 No. 18, pp. 7513-7518.

Wei, J.T., Lin, S.Y. and Wu, H.H. (2010), "A review of the application of RFM model", *African Journal of Business Management*, Vol. 4 No. 19, pp. 4199-4206.

Williams, P.H., Eyles, R. and Weiller, G. (2012), "Plant microRNA prediction by supervised machine learning using C5.0 decision trees", *Journal of Nucleic Acids*, Vol. 2012 No. 2012, available at: http://dx.doi.org/10.1155/2012/652979

Yeh, I.C., Yang, K.J. and Ting, T.M. (2009), "Knowledge discovery on RFM model using Bernoulli sequence", *Expert Systems with Applications*, Vol. 36 No. 3, pp. 5866-5871.

Yu, X., Guo, S., Guo, J. and Huang, X. (2011), "An extended support vector machine forecasting framework for customer churn in e-commerce", *Expert Systems with Applications*, Vol. 38 No. 3, pp. 1425-1430.

Zakaria, I., Rahman, B.A., Othman, A.K., Yunus, N.A.M., Dzulkipli, M.R. and Osman, M.A.F. (2014), "The relationship between loyalty program, customer satisfaction and customer loyalty in retail industry: a case study", *Procedia – Social and Behavioral Sciences, 2nd International Conference on Innovation, Management and Technology Research*, Vol. 129, pp. 23-30.

Corresponding author
Samira Khodabandehlou can be contacted at: s.khodabandehlou@yahoo.com