The CSI Journal on
Computer Science and Engineering
Vol. 18, No. 2, 2021
Pages 1-15
Regular Research Paper

A Targeted Hybrid Model to Customer Churn Prediction in the Insurance Industry (A Case Study)

Ali Baghaei Monireh Hosseini

IT Department, Industrial Engineering Faculty of K. N. Toosi University of Technology, Tehran, Iran.

Abstract

This study first applies a new customer value model (LRFMPG) in an insurance company that examines the impact of length of the relationship, recency, purchase frequency, monetary value, profit and groups of purchased insurance policies on the valuation of customers. This study employed the analytic hierarchy process to determine the weight of each LRFMPG variable. After the identification and selection of valued customers, development of the churn prediction model is performed with extraction of effective variables and factors on the policyholders' churn, and then the importance of these factors on the churn of valuable customer is investigated. For this purpose, predictive models are built using different methods (neural networks, decision trees, logistic regression and support vector machine) and the accuracy of the built models has been evaluated, and eventually, a hybrid model has been suggested. The results show that the proposed hybrid model has a higher accuracy than single models in churn prediction.

Keywords: customer churn prediction, customer valuation, customer relationship management, insurance industry, hybrid model.

1. Introduction

Today, competition can be achieved by increasing customers relationship, in addition to reducing price and product diversity. Customers have become transient, and switching between competitors has become more accessible and less costly for them due to the improvement of accessing information [1]. Increasing competition and decreasing customer loyalty rates in recent decades have led to the emergence of a paradigm in the area of marketing management that encourages firms to transition from a product-centric approach and adopt a customer-centric approach, or to be more precise, the definition of marketing strategies based on the look of the outside of the organization within the organization. All approaches that have been raised around this paradigm are based on focusing on understanding customer needs, not product characteristics [2].

Organizations realize that competing with cheaper, better, or different products is not sufficient, and competitive advantage cannot be achieved only by purely differentiating products, but through enhanced customer relationships [3]. Therefore, customer relationship management (CRM) always concentrates on confirmed customers that are the most fertile data source for decision making. This data reflects customers' actual individual behaviour [4]. The behavioural data can be used to evaluate the customers' potential value [5], assess the risk of stopping paying their bills, and anticipate their future needs [6].

Relationship with customers is considered an asset, but not as tangible asset. They are not owned in the same way. They provide the potential to compete with competitors. This has led to an interest in valuing and managing customer relationships [7]. CRM systems have replaced traditional mass marketing strategies and provide personalized or selectively marketing [8]. On the other hand, the 80/20 rule suggests that 80% of revenue is provided by 20% of customers [9]. So, identification of this group of customers is essential to study their needs and preferences to develop appropriate strategies to retain, satisfaction, and loyalty [10]. Therefore, focusing on retaining high-value customers is a reasonable strategy [11].

Studying the profile of lost customers and predicting customer churn is crucial for survival in highly competitive industries [11]. Customer churn prediction has been applied in numerous fields, particularly in the telecommunication and financial industries [12, 13, 14, 15, 4]. Researchers recommend a defensive marketing strategy [16, 17] that prevents customers from switching service providers.

Today in industrial countries, the number of service firms is increasing, and Iran is not an exception. The growth rate of the number of insurance companies is impressive in Iran [18]. The penetration rate of insurance in Iran is 1.5% of GDP, and if social insurance premiums, pension funds, and supportive insurance are counted in this ratio, it is 4.6% of GDP [18]. So, the importance of the insurance industry cannot be denied in this country.

Insurance companies, as well as other industries in the services markets which are faced with low switching costs, are confronted with problems such as retaining current customers, and insurance companies can no longer rely on a steady customer base [19]. Long-term relationship with customers is crucial in the insurance industry. Retaining current customers has great importance and needs comprehensive research for insurance companies.

To manage customer churn for companies effectively, it is essential to build a more effective and accurate customer churn prediction model. Statistical and data mining techniques have been used to create prediction models [4]. In literature, hybrid data mining models by combining clustering and classification data mining techniques can improve the performance of the single clustering or classification techniques individually [4]. In particular, they are composed of two learning stages, in which the first one is used for pre-processing the data and the second one for the final prediction output [20].

The contribution of this paper is twofold. First, we expanded the body of knowledge in customer value analysis, precisely length of the relationship, recency, purchase frequency, monetary value, profit, and groups of purchased products (LRFMPG), to study the effect of such a model in the insurance industry. The results indicated that this model is generalizable in this industry. Second, a combination method to create the hybrid model is examined in terms of customer churn prediction, based on combining two classification techniques, i.e., boosted C5.0 decision tree and multilayer perceptron neural networks (MLP-NN).

The remainder of this paper is organized as follows. In section 2, we review the literature related to customer churn. section 3 describes the research methodology, and then in section 4, modeling in real data is performed. Section 5 presents the analysis of the results and discussion. Finally, the conclusion is provided in section 6.

2. Literature Review

In this section, we first define the basic concepts in the field of research and then have a review of the literature in this area.

2.1. Customer Churn

Retaining customers for long periods is one of the essential ways to enhance customer value. In a sense, retained customers have more revenue and profit margins than new customers [21]. Retaining existing customers is based on rapid and timely detection of customers who want to leave the company or so-called “churn.”

Churn has several definitions. The exact definition of this concept is essential. In previous studies, observed that the definition of churn is an agreement, and there is no standard and identical definition applicable to all industries. It’s defined by experts of a particular industry or organization according to the characteristics of the industry environment and the organization [22]. In general, customer churn is his inherent tendency to cancel continuing business relationships with a company in a period [23]. According to this definition, the churned customer is the one who cut off all his activities with the company (for example, accounts) [23] or minimum purchase repeat is less than the average [24] and his lifetime value is gradually falling [25].

Research has shown that retaining existing customers is less costly for organizations to attract new customers. The loss of a customer could reduce sales revenue and increase acquisition costs [26]. Attracting new customers requires costs such as the cost of search, advertisement, and marketing [4]. These costs are several times the company's efforts to retain existing customers [27]. Hence it is effective that organizations focus more on the existing customers [3].

There are two basic approaches in managing customer churn: untargeted approaches and targeted approaches [8]. Untargeted approaches rely on superior products and mass advertising to increase brand loyalty and retain customers. Although the implementation of this method is easy, there are risks related to waste resources while providing these incentives to customers who want to stay in the company. Targeted approaches rely on identifying customers who are likely to churn then either providing them with a direct incentive or customizing a service plan to stay [8]. The adoption of targeted strategies enables companies to use customer transaction data in the execution of analyzing and predicting future customer behavior [28].

2.2. Customer value analysis

Since it cannot be said that all customers have an equal share in the company's success, the satisfaction of crucial customers is more sensitive. As well as companies believe they should not pay for attracting customers at each level of profitability for the company. Customer value analysis involves identifying patterns and group associations by using a high number of customer data [11]. By conducting customer value analysis, businesses can identify valued customers who have contributed to business revenue the most for a considerable period [29].

Among the models in determining customer value, the RFM model, which can identify customers and adopt appropriate strategies for different groups of customers, is used widely in customer behavior analysis and has received considerable attention in recent literature [11, 29, 30, 31, 32, 33, 34, 35].

The RFM model considers the customer ‘how many times’, ‘when’, and ‘how much’ has been purchased [36]. According to [37] past customer behaviors can be a measure of his loyalty. This analysis calculates a score for each customer and considers customers with a high rating as profitable and valuable customers. Three parameters considered in this analysis are recency, frequency, and monetary value of their purchases [35]. The importance of these three indicators varies among industries. Identifying the relative weight of these indicators for the specific domain of business is advantageous [11].

This model has been expanded by including additional variables. [38] suggested RFM-Customer Patterns in retail industries, which combines RFM with favorite customer purchase patterns. [39] proposed LRFMC to increase the quality of segmentation in the insurance industry by including the length of relationship and customer costs. Wei et al. [40] proposed the LRFM model, including the length of the relationship, the CRFM model in which the RFM model applied to various product categories, and the CLVRFM model based on traditional RFM analysis. [37] added product category groups to develop the GRFM model. LRFMP suggested by [11] adds the length of relationship and profit to traditional RFM. [35] developed RFMTC model by including the first purchase time (T) and the customer churn probability (C).

2.3. Data Mining

In the customer churn phenomenon, data mining techniques discover and analyze large amounts of available data to help to select customers who are more likely to leave the company [2]. The customer's past behavior is an essential predictor of his future behavior [41], and the purpose of predictive modeling is to predict what is likely to happen in the future based upon what happened in the past [42].

Building an effective customer churn prediction model has become an important topic for business and academics in recent years. To establish an efficient and accurate prediction model, many methods have been proposed recently [2]. Data mining tools can help organizations to discover hidden knowledge and extract helpful information from a large amount of data. Data mining is the process of discovering various models, summaries, and derived values from a given collection of data [43].

Two primary objectives of data mining in practice are "describe" (i, e. discover exciting patterns or relationships in the data), and "predict" (i, e. predict or classify the behavior of the model based on available data) [44]. In fact, the description focuses on finding human-interpretable patterns describing the data, and prediction involves using some variables or fields in the dataset to predict unknown or future values of other variables of interest. The goals of description and prediction can be achieved using a variety of particular data mining methods including classification, clustering, regression, and so on [4]. Customer churn prediction can be regarded as a classification problem, in which each customer is classified into one of two classes, churn or non-churn [45].

To develop an effective customer retention program, used models should be as accurate as possible to recognize the churning of customers correctly [42, 46]. Otherwise, when incentive money is paid for customers who will not churn , these systems will be very wasteful [4].

Customer churn problems are mainly related to the selection of churn predictive techniques. Models are required to identify explanatory variables which are related to the tendency to churn. Furthermore, it includes the definition of the causality/link between these variables and the churn [47, 48]. There are several well-known methods for churn analyzing and forecasting. For this purpose, data mining techniques and statistical analysis are used. As one of the crucial phases in the churn prediction process, classification can directly affect the prediction results [49].

To understand how related work constructs their prediction models, some of the current related studies are shown in Table 1. This table shows the classification of articles based on the technique used, environment/ dataset, and the publication year in customer churn issue by examining 47 articles in this area between 2008 and 2015. Some articles used only one method for modeling and prediction of churn. Some articles used more than one technique for modeling and forecasting, and after evaluating each technique, they have chosen a technique that has the highest accuracy and precision as a good technique. Several other papers have used a combination of two or more techniques in churn modeling and forecasting. For example, one method for preprocessing and another technique for prediction. Also, some papers have improved and developed previous techniques and have invented a new technique. There are some studies that show that one technique is better than another one, and vice versa [47, 48].

The decision tree is the most commonly used technique among the 34 methods and techniques used in customer churn. This technique has been used in 24 articles from 47 articles reviewed in total (51%). This algorithm has been used either in hybrid approaches in combination with other methods and as an independent method to predict customer defections in the literature. After that, regression technique by frequency of 17 (36%), neural networks by 14 (30%), support vector machine by 12 (26%), and random forest with 15% are the most used techniques in the field of customer churn.

Also, about the distribution of articles based on environment/dataset, several articles used from real datasets to analyse and predict customer churn, and some articles have used prepared datasets, such as the UCI machine learning repository dataset. The telecom and mobile operators are the most environments where customer churn is the subject of consideration, and 22 articles investigated it, and it has the highest share in this field. Newspaper subscription services and retail with six studies are in the next orders. After them, the bank by five studies is in the following

There is little literature in the field of customer churn in the insurance industry. Authors of article [50] have studied life insurance in one of the insurance companies in Switzerland. They used the TF/IDF representation from information retrieval for compiling time-related features of the dataset. Experimental results show that these new features lead to superior results in terms of accuracy, precision, and recall. A heuristic is given, which calculates how much the feature space is enlarged by the transformation to TF/IDF.

Table 1. Related literature about customer churn

Author, Year	Method / Technique	Environment/ Dataset
M. Ballings and D. Van den Poel, 2012 [51]	Decision tree, Logistic regression	Newspaper subscription
I. Bose and X. Chen, 2009 [52]	Decision tree, Boosting, K-medoid, K-Means, SOM, FCM, BIRCH	Teradata Data Center at Duke University
J. Burez and D. Van den Poel, 2008 [53]	Random forest, Survival analysis	Pay-Tv company
J. Burez and D. Van den Poel, 2009 [54]	Logistic regression, Random forest, Boosting	Europe churn dataset (Bank, Telecommunication, Pay-Tv, Super Market, Newspaper)
K. Chen et al., 2014 [11]	Decision tree, MLP neural network, Support vector machine, Logistic regression	Logistics industry
Z-Y. Chen et al., 2012 [55]	Support vector machine	Telecommunication, Grocery store, Adventure game
K. Coussement and W. De Bock, 2013 [42]	Decision tree, Generalized additive model (GAM), Random forest	Online gaming company
K. Coussement and D.Van den Poel, 2008 [56]	Support vector machine, Logistic regression, Decision tree	Newspaper subscription
K. Coussement and D. Van den Poel, 2008 [57]	Logistic regression	Newspaper company
K. Coussement and D. Van den Poel 2009 [46]	Random forest, Logistic regression, Support vector machine	Newspaper subscription
K. Coussement et al., 2010 [58]	Generalized additive model	Newspaper company
W. De Bock and D. Van den Poel, 2011 [59]	Rotation forest, RotBoost	Suppliers, bank, Telecommunication, Mail-order garments
W. De Bock and D. Van den Poel, 2012 [60]	Generalized additive model, GAMensPlus	Super Market, bank, Telecommunication, Mail-order garments
N. Glady et al., 2009 [25]	Neural network (MLP), Decision tree, Logistic regression, AdaCost	Financial Services
W. Hengliang and Z. Weiwei, 2012 [43]	Decision tree, Neural network, K-means	E-commerce company (Online shopping)
B. Huang et al., 2010 [61]	Decision tree, Neural network, Support vector machine	Telecommunication
B. Huang et al., 2010 [62]	NSGA-II, Decision tree	Telecommunication
B. Huang et al., 2012 [12]	Decision tree, Neural network, Support vector machine, Logistic regression, Naive Bayes, Evolutionary data mining algorithm	Telecommunication
A. T. Jahromi et al., 2014 [63]	Decision tree, Logistic regression, Boosting	Retail
A. T. Jahromi et al., 2010 [21]	Two-step clustering, Decision tree	Telecommunication
C. Kirui et al., 2013 [64]	Decision tree, Bayesian belief network, Naive Bayes	Cellular phone
P. Kisioglu and Y. I. Topcu, 2011 [14]	Bayesian belief network	Telecommunication
Y. Lai and J. Zeng, 2014 [65]	Survival analysis	Digital library
S. Lessmann and S. Voß, 2009 [66]	Support vector machine, Decision tree, Logistic regression	UCI machine learning repository
C. S. Lin et al., 2011 [67]	Rough set theory, Flow network graph	Bank credit cards
D. S. Liu and C. H. Ju (2009) [68]	Support vector machine, Principal component analysis (PCA)	Mechanical engineering products Vendor
N. Lu et al., 2014 [69]	Boosting, Logistic regression	Telecommunication
VL. Migue´is et al., 2012 [47]	Logistic regression, Sequence mining	Grocery retail
VL. Miguéis et al., 2012 [48]	Markov chains, Logistic regression, Random forest	Grocery retail
S. Nabavi and S. Jafari, 2013 [70]	Random forest, Boosted trees	Food industry
G. Nie et al., 2011 [15]	Logistic regression, Decision tree	Bank
M. Owczarczuk, 2010 [71]	Logistic regression, Linear regression, Fisher linear discriminant analysis, Decision tree	Mobile telephony services
P. C. Pendharkar, 2009 [72]	Neural network, Genetic Algorithm	Mobile telephony services
J. Qi et al., 2009 [73]	Decision tree, Logistic regression	Mobile telephony services
A. Sharma and P. Kumar Panigrahi, 2011 [27]	Neural network	Mobile telephony services
C. F. Tsai and M. Y. Chen, 2010 [74]	Decision tree, Neural network, Association rules	Telecommunication
C. F. Tsai and Y. H. Lu, 2009 [4]	Hybrid neural network, Self-organizing map	Telecommunication
T. Vafeiadis et al., 2015 [75]	Perceptron neural network, Support vector machine (SMV-RBF, SMV-POLY), Decision tree, Naive Bayes, Logistic regression, Boosting	UCI machine learning repository
W. Verbeke et al., 2012 [13]	Decision tree, Neural network, Support vector machine	Telecommunication
W. Verbeke et al., 2014 [76]	Relational and non-relational classifiers	Telephone operator
W. Verbeke et al., 2011 [77]	Ant Colony (AntMiner+), ALBA	KDD Library
T. Verbraken et al., 2014 [78]	Bayesian network classifier	Telecommunication, CRM center web at Duke University, UCI
Y. F. Wang et al., 2009 [10]	Decision tree	Wireless network company
Y. Xie et al., 2009 [79]	Random forest	Bank
G. E. Xia and W.D. Jin, 2008 [80]	Support vector machine	UCI, mobile Telecommunication
X. Yu et al., 2011 [81]	Support vector machine, Neural network, Decision tree	E-commerce company
X. Zhang et al., 2012 [45]	Neural network, Decision tree, Logistic regression, Propagation model	Mobile phone service provider

Article [18] has investigated insurance customer churn prediction considering the CLV (car insurance field). In this study first, they identified important factors causing customer churn in the insurance industry to have a specific behaviour by using a k-means clustering algorithm. Then they tried to predict the future behaviour of them by logistic regression.

The authors of the article [19] presented a dynamic modeling approach for predicting individual customers' risk of leaving an insurance company. A logistic longitudinal regression model that incorporates time-dynamic explanatory variables and interactions is fitted to the data. As an intermediate step in the modeling procedure, they applied a generalized additive model (GAM) to identify nonlinear relationships between the logit and the explanatory variables. Both out-of-sample and out-of-time predictions indicate that the model performs well in identifying customers likely to leave the company each month.

Article [82] focused on a real case of the motor insurance sector and proposed four different methods to predict lapsing from a portfolio of policies. Their comparison analyses the outcomes of logistic regression, a conditional tree, a neural network, and a support vector machine. This paper shows depending on the type of analysis and the objective of the researcher, the optimal prediction method may differ.

According to the literature, despite the varied academic research in the analysis and prediction of customer churn in various industries, few academic research has addressed customer churn in insurance which is one of the most important service industries. Also, the existing studies in this field didn't perform any customer value analysis to identify the valuable insurance customers. Therefore, this study first applies a new customer value model (LRFMPG) in the insurance company to identify valuable customers and then, after the identification and selection of valued customers, with extracting the essential variables on the policyholders' churn, development of the churn predictive model is done using different predictive methods (neural networks, decision trees, logistic regression and support vector machine), and finally, a hybrid model is proposed.

3. Construction of model

The modeling process is shown in Figure 1. This research case study is Saman insurance company (an Iranians NGO insurance company) which sells various types of insurance policies in different categories. In this research, initial data of the company have been used.

To collect the required data, the most reasonable and appropriate data fields for modeling concerning the purposes of the study have been derived according to the literature, available databases, and the comments of experts. The period time for data extraction is from March 2010 until March 2015 (5 years). 198146 data records which correspond to 105644 customers have been given to researchers. In the pre-processing phase, integration of customers, removing duplicate data and data consensus were conducted to prepare the data and provide the data for the subsequent phases. Also, it was observed that a significant number of rows in the dataset (34303 customers) are new policyholders. These customers are new in the company and have been paid premiums for one (or more) categories last year. Since these customers' data cannot be valuable data for customer valuation, and also not possible to predict their churn (since data is not available for next purchase(s) and they are new in the company), therefore excluded from the dataset. After excluding new customers, data pre-processing, and excluding outlier data, a total of 71337 unique customers remained in the dataset. Also, several fields have been excluded because of ineffective.

To label customers in the case study, according to the opinions of experts, customers who have purchased insurance policies only last year and did not purchase any insurance policies before this year have been considered as new customers. On the other hand, all customers who have purchased insurance policies in the last years of the dataset and also are not new customers are considered as active customers because they were active in the last year.

For definition of turned customers, the purchase history of the last two years has been investigated. If the customer did not purchase any insurance policies in the last two years, he has been labeled a churner, but if he did not purchase only in the last year and has bought a year before the last year, it has been considered as active. The reason that only last year of customer's purchase history has not been considered as an indicator to define the churn is that there are many cases in the dataset that the customer didn't purchase for one year, but he is back again now. This group of customers churned temporarily and may have been incidental churn.

As well as using existing data, new variables were defined according to the research objectives. These variables are essential for determining the customer value and also churn prediction. After consulting with experts and according to the literature and also taking into account the existing databases, concluded that a total of 17 variables in two general categories (customer profile and transaction behavior) are essential to determine the customer value and analysis and prediction of churn which are shown in Table 2.

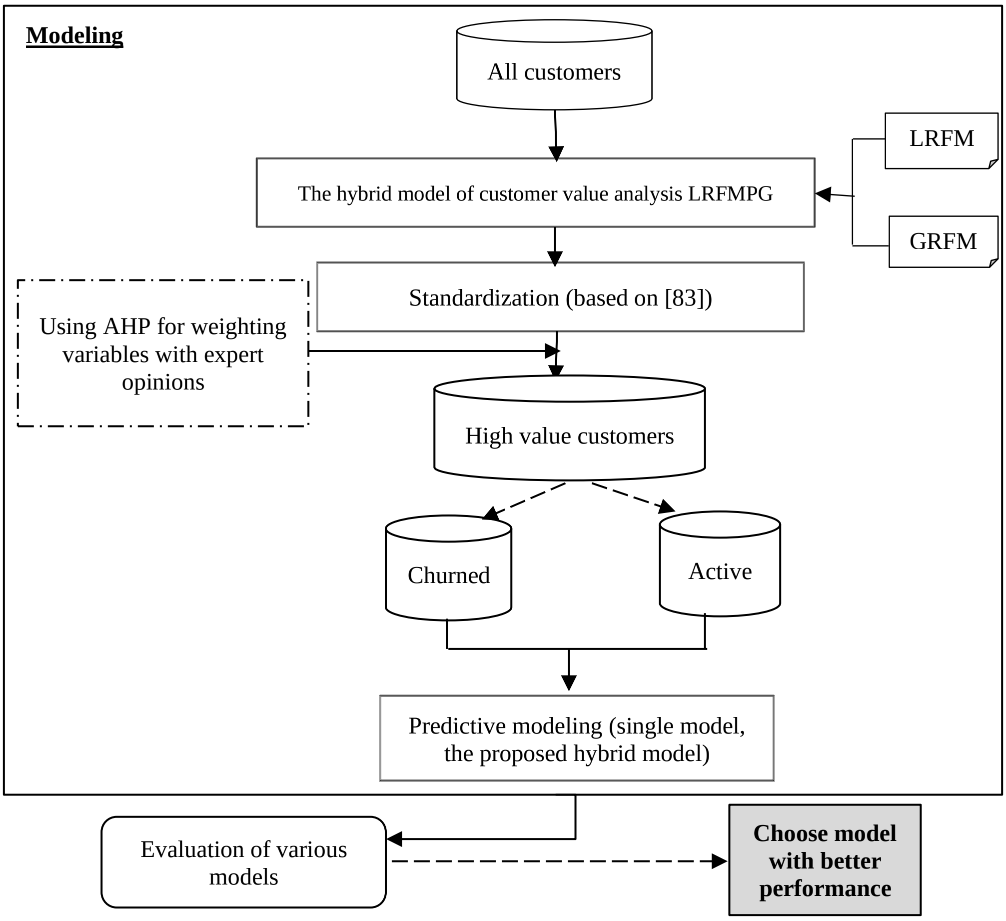

Fig. 1 Modeling processes

Table 2. Considered variables

Group	Name	Variable description
Customer profile	State	Geographical area of the customer
	Gender	Customer gender (0 women and 1 man)
	Age	Customer age
	Martial	Marital status (married and single)
Transactional behavior	G	Purchased insurance policies groups
	L	The number of days between the first and last purchase of the customer (length of relationship)
	R	The number of days between the last purchase and date of 2015/3/21 (recently)
	F	The total number of customer purchases (Frequency)
	M	Average of customer payments (premiums) per year
	P	The customer profitability
	F/L	The relative frequency of purchases that is the frequency division number of years.
	ClaimCount	Total claims by the customer
	ClaimAve	The average amount of claims by customers each year
	Variation	The average difference paid in two consecutive years
	MIN_TIME	Minimum interval between two consecutive purchases
	MAX_TIME	The maximum interval between two consecutive purchases
	AVG_TIME	The average interval between two consecutive purchases
Active/ Churn	Label	Customer Labels (Active/Churn)

4. Modeling

The modeling is described in this section. This section consists of two parts. First, the modeling for customer value analysis and customer rating is performed, and then valued customers are determined. In the second part, predictive modeling is done.

4.1. Customers Valuation

Since the aim of this study is targeted modeling, and targeted modeling focuses on high-value customers to properly allocate resources, a new model for the valuation of customers is proposed. The proposed model in this study is the combination of 2 models. First, LRFMP model [11], which added L (long-term customer relationship with the company) and P (customer profitability) variables to RFM, and second, GRFM model by including customer purchased product groups (G) [37], and creates a new model named LRFMPG to the valuation of customers (taking into account the opinions of the experts) that six variables to determine the value of customers include: length of the relationship (L), recency of purchase (R), purchase frequency (F), customer monetary value (M), customer profitability (P) and groups of purchased insurance policies (G).

Then calculated LRFMPG values mapped to the standardized numbers in the interval (1, 5). The standardization process has been done by the method described by Miglautsch in 2000 [83]. So first, all of the customers in the dataset were sorted in ascending order according to the variable recency (R), and in ascending order according to the other five variables. For R, M and, P variables, the customers were partitioned into equal quintiles. Then, based on the quintiles, these variables have been mapped to numbers in the range of one to five so that these quintiles were assigned numbers from 5 (highest customer value) to 1 (lowest customer value).

In the case of variable F, since there are a lot of customers who have F=1 (over 20 percent), we cannot use the above method. Therefore, when scoring this variable, buyers with one purchase, receive a score of 1. Then the average frequency of the remaining customers with F>1 is calculated. If the total number of purchases of the customer is less than the calculated average, will receive a score of 2. This process continues twice until scores are calculated for all customers. The above method is similar for the variable L because a large number of customers in the dataset have L= 0, meaning that these customers bought only once. Hence, the length of their relationship with the company is zero.

Now we're going to determine the importance (weight) of each variable of LRFMPG. To determine the weight of these variables, this study has used the analytic hierarchy process (AHP). In the first step, the impact between the features with the help of experts has been determined. So that, from the factors extracted (LRFMPG) in determining the value of customers, one survey matrix was prepared. These factors form the rows and columns of the matrix. To study the relative importance of these variables were asked from nine Saman insurance sale experts of different branches to show the importance of each row factor on the column factors by pair comparing each factor located in each row of matrix with each factor located in the columns of the matrix numerically. An example of a matrix of pairwise comparison of analytical hierarchy process for LRFMPG model formed by one of the experts is given in Table 3.

After collecting the questionnaires, the weight of each variable of LRFMPG in determining the value of customers (based on their importance) was obtained by AHP assessment. This LRFMPG relative weight of each variable is shown in Figure 2. This figure shows that the purchases frequency of customers- according to experts- has the highest weight and importance. Then the length of the relationship and the number of purchased insurance policy groups had the most significant weight. Profitability, monetary, and recency are also next ranked respectively.

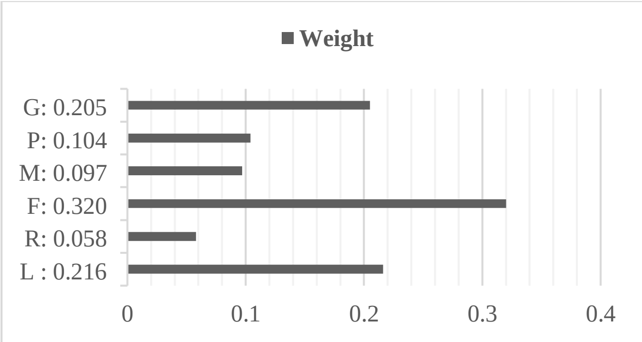

Fig. 2 Calculated weight (importance) for each LRFMPG variable using AHP

When the standardized LRFMPG score was collected for each customer, and the impact factor was determined for each L, R, F, M, P variable using the AHP, the total score for each customer was calculated as follows.

Assume that:

$C = c_1 \cup c_2 \cup c_3 \cup \ldots \cup c_n$

(Where C is customers in insurance company)

LRFMPG standardized scores for customer C_i ($C_i \in C$) are: $SL_i, SR_i, SF_i, SM_i, SP_i, SG_i$

Also assume that total customer score C_i is shown as $ST(C_i)$. This value is calculated from the following equation:

$$ST(C_i) = (0.216 * SL_i) + (0.058 * SR_i) + (0.320 * SF_i) + (0.097 * SM_i) + (0.104 * SP_i) + (0.205 * SG_i).$$

Table 3. An example of a matrix of pairwise comparison AHP- complemented by one of the sales experts

Attribute	**Compare in terms of importance**									**Attribute**
	1:9	1:7	1:5	1:3	1:1	3:1	5:1	7:1	9:1	
Recency						✓				**Length**
Frequency			✓							
Monetary					✓					
Profitability						✓				
Groups of purchased insurance policies				✓						
Frequency				✓						**Recency**
Monetary				✓						
Profitability	✓									
Groups of purchased insurance policies						✓				
Monetary							✓			**Frequency**
Profitability					✓					
Groups of purchased insurance policies						✓				
Profitability			✓							**Monetary**
Groups of purchased insurance policies				✓						
Groups of purchased insurance policies								✓		**Profitability**

The calculated weight by the AHP method is the coefficient values for each variable. Accordingly, the highest total score for a customer is five, and the least total score is one. After determinate a score for each customer, a split sample analysis is done to show the difference between high-value and less-value customers, so that a customer who has more score (or less) than the average population (fifty percent of customers with high scores) is labeled as a valuable customer (or less valuable) [84].
Accordingly, almost half of the customers (35617 persons) that had lower total scores were omitted from the overall dataset. Totally, 35720 customers are considered as valued customers. General information about the classification of low-value and high-value customers (active and churn) is shown in Figure 3. Also, among valued customers, 39% (13940 of 35720 valuable customers) have turned away and left the company, according to the definition of the churn.

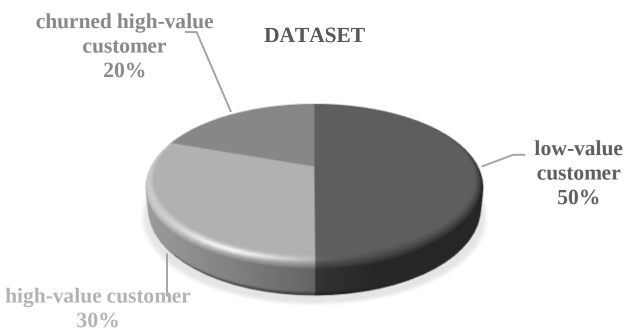

Fig. 3 Classification of low-value and high-value customers

4.2. Prediction modeling and evaluation

In this section, prediction modeling will do. The first step is dividing the data into training and testing partitions. To construct prediction models, we should split data into training (learning) and testing parts; The training part is used to learn the basic model, and then the test part is used to test and validate models [85].
To divide the data into training and testing, churn and active customers were separated first, and then to create the training part, randomly seventy percent of each of these two datasets were selected and combined. Then, the remaining thirty percent of the active and churn dataset combined for the testing part [86].
Also, to achieve more accurate results, 10-fold cross validation is used. In 10-fold cross validation, the data divide into ten sub-set. Each time from the ten sub-set, one is used for the validation, and nine others are used for the training. This process repeats ten times; All the data is used once for training and once for validation. Finally, the average result of the 10-fold validation is selected as the final estimate [10].
In the second step, making the model is performed. To build a predictive model of valued customers of Saman insurance, various supervised learning techniques that have been used widely in the literature of churn are used. For this purpose, the modeling has been done with different algorithms of artificial neural networks (MLP and RBF), different decision tree algorithms (QUEST, CHIA, C&R and, C5.0), support vector machine (with different kernels), and logistic regression. Also, a new hybrid model that combines decision trees and neural networks has been proposed.

4.2.1. Modeling and evaluation- neural network

At first, various models of the neural network were built with different parameter values of Alpha (momentum factor) and ETA (learning rate). Based on the accuracy of the built models, the Alpha and Eta optimal values were considered as 0.9 and 0.3 respectively. Also, several different neural network models with different numbers of hidden layers (1 to 3 hidden layers) and neurons (2 to 5 neurons in each layer) were created. The obtained values of the evaluation criteria for the accuracy, precision, recall, and F1 of some built models with the neural network are given in Table 4.
As the Table 4 indicates, MLP neural network with one hidden layer and three neurons has the best prediction performance in created models with the neural network.

4.2.2. Modeling and evaluation- neural network

As mentioned, to achieve an effective prediction model, several widely used techniques in the literature of churn have been used to build the predictive model. In this study, in addition to the artificial neural network performed in the previous section, modeling is conducted with the decision tree, SVM, and logistic regression. Also, to increase the accuracy of the modeling, CHAID, and C5.0 decision tree algorithms have been used as boosted too. The performance of the models is given in Table 5 by different evaluation criteria.

Table 4. Evaluation of built models with neural network

Technique	Hidden layer	Precision	Recall	Accuracy	F_1
MLP N. N (Back Propagation)	1 hidden layer and 2 neurons	81.60%	74.81%	74.35%	78.05%
	1 hidden layer and 3 neurons	**78.21%**	**81.28%**	**74.77%**	**79.81%**
	1 hidden layer and 4 neurons	73.56%	82.49%	71.25%	77.76%
	2 hidden layers and 4 neurons	76.59%	83.08%	74.20%	79.70%
	3 hidden layers and 9 neurons	84.35%	51.62%	64.66%	64.04%
RBF N. N	---	76.95%	74.05%	70.65%	75.47%

Table 5. Evaluation of models

Technique	Algorithm	Precision	Recall	Accuracy	F_1
Decision Tree	QUEST	80.43	84.33	77.94	82.33
	CHAID	85.17	85.21	81.94	85.18
	Boosted CHIAD	88.38	88.59	85.94	88.48
	C5.0	86.44	86.63	83.56	86.53
	Boosted C5.0	**87.38**	**90.17**	**86.06**	**88.75**
	C&R	80.50	82.65	77.22	81.65
SVM	RBF Kernel	76.73	82.18	73.94	79.41
	Poly Kernel	81.69	75.93	74.90	78.65
	Linear Kernel	76.25	83.20	73.95	79.57
Logistic regression	---	78.20	77.13	72.94	77.66

Results show that the C5.0 decision tree has better performance to churn prediction than other methods and generally, decision tree algorithms considerably compared to the other methods are superior. Lift and ROC curves for these algorithms are given in Figure 4 and Figure 5.

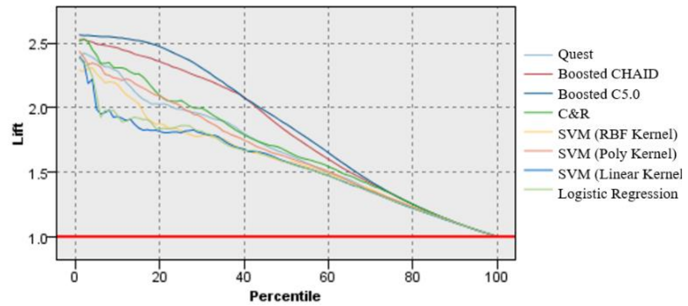

Fig. 4 Lift curve

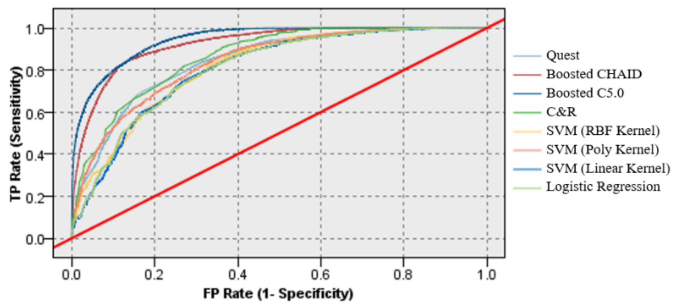

Fig. 5 ROC curve

Also, the importance of variables that have the most significant impact on the customer who is churned or active has been assessed. According to the survey conducted with the C5.0 decision tree, the length of customer relationship is the most influential variable. After long-term relationship (length), customer profitability, relative purchases frequency, the maximum interval between two consecutive purchases, the average difference paid in two consecutive years, the average of customer payments (premiums) per year, and the number of purchased insurance policies groups are essential variables in predicting customer churn, respectively. In Figure 6, the priority variables in predicting using the C5.0 are shown.

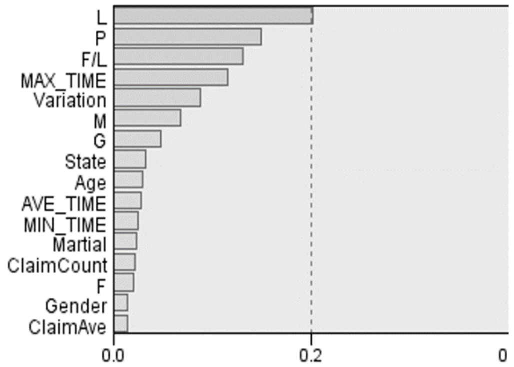

Fig. 6 The importance of variables in predicting using C5.0

4.2.3. Modeling and evaluation- proposed hybrid model

A hybrid model is proposed to reduce the miss alarm rate (type I error) and increase system efficiency and accuracy of its predictions. The proposed hybrid model includes two classification models. The first technique is boosted C5.0 decision tree (that has the best performance among models), and the second is the neural network. In this model, the decision tree has the task of data reduction, and the second one does the prediction. That is, the main primary dataset is used to create and test the built-model with the C5.0 tree. Because, as can be seen, there is no 100 percent accuracy, there are several samples that have been correctly predicted, and some of them are incorrect. As a result, data that are predicted incorrectly can be considered as the outlier, because the model is unable to predict them correctly. Then correct predicted data from the first model is used to train the neural network as a predictive model. In Figure 7, the process of combining these two models to build the hybrid model is given.

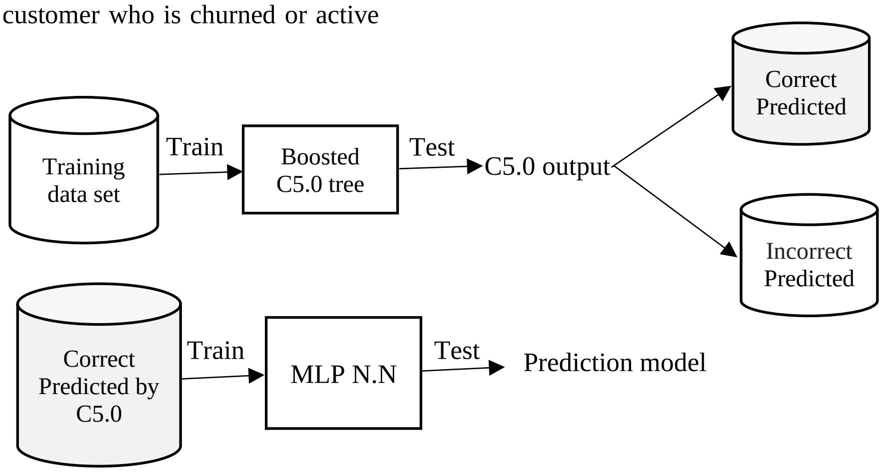

Fig. 7 Process of building the proposed hybrid model

In this hybrid method using boosted C5.0 decision tree, 4976 samples (13.93%) of 35720 samples in the initial dataset of valued customers were considered as the outlier and were removed from the dataset. So, 19640 active customers (64 percent) and 11104 churned customers (36%) remained in the dataset. The MLP neural network is used with different numbers of hidden layers and neurons in each layer to build the prediction model, and the model that has the best performance among the models is considered as the final prediction model. In Table 6, information of different built predictive models and the number of neurons in each layer are provided.

Table 6. Built predictive models on hybrid models

Row	The number of neurons in hidden layer 1	The number of neurons in hidden layer 2	Model Name	Row	The number of neurons in hidden layer 1	The number of neurons in hidden layer 2	Model Name
1	1	---	*X1*	**25**	3	1	*X25*
2	2	---	*X2*	**26**	3	2	*X26*
3	3	---	*X3*	**27**	3	3	*X27*
4	4	---	*X4*	**28**	3	6	*X28*
5	5	---	*X5*	**29**	3	9	*X29*
6	6	---	*X6*	**30**	3	12	*X30*
7	7	---	*X7*	**31**	6	1	*X31*
8	8	---	*X8*	**32**	6	2	*X32*
9	9	---	*X9*	**33**	6	3	*X33*
10	12	---	*X10*	**34**	6	6	*X34*
11	11	---	*X11*	**35**	6	9	*X35*
12	12	---	*X12*	**36**	6	12	*X36*
13	1	1	*X13*	**37**	9	1	*X37*
14	1	2	*X14*	**38**	9	2	*X38*
15	1	3	*X15*	**39**	9	3	*X39*
16	1	6	*X16*	**40**	9	6	*X40*
17	1	9	*X17*	**41**	9	9	*X41*
18	1	12	*X18*	**42**	9	12	*X42*
19	2	1	*X19*	**43**	12	1	*X43*
20	2	2	*X20*	**44**	12	2	*X44*
21	2	3	*X21*	**45**	12	3	*X45*
22	2	6	*X22*	**46**	12	6	*X46*
23	2	9	*X23*	**47**	12	9	*X47*
24	2	12	*X24*	**48**	12	12	*X48*

To build the hybrid model, after using the C5.0 decision tree to reduce and remove outlier data, 48 different neural network models were created to predict and determine which model has the best performance. In Table 7, all models and evaluation criteria are shown.

As it is clear from Table 7, the best performance of the hybrid models is X47 which has 16 inputs, two hidden layers with 12 neurons in the first layer, and 9 neurons in the second layer.

Table 7. Evaluation of built hybrid predictive models

Algorithm	Precision	Recall	Accuracy	F_1	Algorithm	Precision	Recall	Accuracy	F_1
X1	90.24	88.29	86.42	89.25	X25	95.23	89.66	90.53	92.36
X2	91.11	88.09	86.90	89.58	X26	92.26	89.34	88.40	90.78
X3	90.77	92.25	89.05	91.50	X27	90.99	92.01	89.08	91.50
X4	91.82	94.69	91.22	93.23	X28	90.50	90.32	87.76	90.41
X5	91.73	94.07	90.79	92.88	X29	91.63	92.92	90.06	92.27
X6	88.62	91.49	87.06	90.03	X30	93.03	89.74	89.15	91.35
X7	91.71	94.51	91.03	93.09	X31	94.12	94.39	92.65	94.26
X8	92.31	95.17	91.85	93.72	X32	92.69	94.06	91.47	93.37
X9	92.15	95.06	91.67	93.58	X33	91.23	92.31	89.41	91.76
X10	92.21	94.84	91.59	93.51	X34	93.59	92.61	91.23	93.10
X11	93.18	93.99	91.77	93.58	X35	94.02	94.42	92.60	94.22
X12	89.43	90.95	87.35	90.18	X36	92.83	93.72	91.37	93.28
X13	88.80	86.24	84.26	87.50	X37	93.84	93.74	92.07	93.79
X14	90.34	87.55	86.07	88.92	X38	92.49	93.68	91.11	93.08
X15	90.12	88.38	86.39	89.24	X39	94.69	93.40	92.44	94.04
X16	90.17	86.73	85.48	88.42	X40	93.06	93.40	91.34	93.23
X17	91.00	88.06	86.81	89.50	X41	92.66	94.98	91.99	93.81
X18	89.96	84.64	84.15	87.22	X42	93.49	94.10	92.05	93.79
X19	94.67	88.48	89.46	91.47	X43	93.85	93.09	91.70	93.47
X20	93.46	89.02	89.01	91.19	X44	92.90	89.26	88.79	91.05
X21	94.67	87.92	89.12	91.17	X45	93.66	94.48	92.39	94.07
X22	90.69	91.90	88.81	91.30	X46	92.86	92.15	90.46	92.50
X23	94.36	88.55	89.37	91.41	**X47**	**94.35**	**94.37**	**92.79**	**94.36**
X24	90.86	87.34	86.30	89.07	X48	93.94	94.44	92.56	94.19

5. Analysis of the results and discussion

Results show that the proposed hybrid model has a better performance compared to single models. Table 8 is a comparison between the proposed model and other models.

Table 8. Evaluating the hybrid predictive model and other models

Technique	Algorithm	Precision	Recall	Accuracy	F_1
SVM	RBF Kernel	76.73	82.18	73.94	79.41
	Poly Kernel	81.69	75.83	74.90	78.65
	Linear Kernel	76.25	83.20	73.95	79.57
LGR	---	78.20	77.13	72.94	77.66
D.T	QUEST	80.43	84.33	77.94	82.33
	Boosted CHIAD	88.38	88.59	85.94	88.48
	Boosted C5.0	87.38	90.17	86.06	88.75
	C & R	80.50	82.65	77.22	81.56
MLP N. N	One hidden layer, and three neurons	78.21	81.28	74.77	79.71
Proposed hybrid model	*Boosted C5.0 and N. N (two hidden layers and 21 neurons)*	*94.35*	*94.37*	*92.79*	*94.36*

As Table 8 indicates, the hybrid model has a high performance better than the single neural network model, 18.02% in terms of accuracy, and 14.65% in terms of the F1 measure. Also, in comparison to the single C5.0 tree, this model has outperformed 6.73% in terms of accuracy, and 5.61% in terms of F1 measure in predicting churn. The confusion matrix of the proposed hybrid model and also the single neural network is given in Table 9. In terms of qualitative, the proposed model misdiagnosed 1231 cases from 19640 active customers in the dataset (only 6 percent) in misclassification of really active customers as churned customers. Also, this model misclassified 1340 cases from 11104 churned customers in the dataset (12 percent), which performs much better compared to the single neural network.

Table 9. Confusion matrix of the hybrid model and single N.N

N. N (one hidden layer and three neurons)	Predicted active	Predicted churn
Active	17703	4077
Churn	4932	9008
Proposed hybrid model	**Predicted active**	**Predicted churn**
Active	18409	1231
Churn	1340	9764

Sixteen critical variables in predicting with the hybrid model are shown in Figure 8.

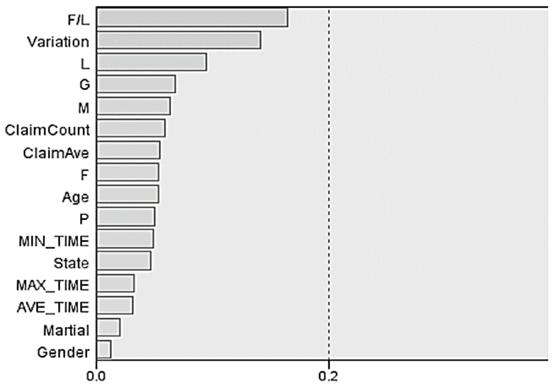

Fig. 8 Importance of the variables in predicting with the hybrid model

By a comparing the hybrid model and the C5.0 decision tree in Table 10, it is concluded that from the seven essential input variables in both methods, five variables have been considered essential in common. Length of the customer relationship, purchase frequency, the average difference paid in two consecutive years, average of customer payments (premiums) per year, and the number of purchased insurance policy groups have been identified as important variables in both models.

Table 10. Essential variables in the decision tree and hybrid model

7 important variables in hybrid model (respectively)	7 important variables in the C5.0 tree (respectively)
1.Relative frequency of purchase	1.Length of customer relationship
2.Average difference paid in two consecutive years	2.Customer profitability
3.Length of customer relationship	3.Relative frequency of purchase
4.Number of purchased insurance policies groups	4.Maximum interval between two consecutive purchases
5.Average of customer payments (premiums) per year	5.Average difference paid in two consecutive years
6.Number of claims by the customer	6.Average of customer payments (premiums) per year
7.The average amount of claims by customers each year	7.Number of purchased insurance policies groups

The survey of the importance of predictor variables in the proposed hybrid model indicates that the relative purchase frequency has the most significant impact on being churned/active customers. After this variable, the average difference paid in two consecutive years is the most crucial influential variable for prediction. Length of the customer relationship variable, which denotes the period that the customer is with the company and uses the company benefits and services, is the third important variable. Types of purchased insurance policies (groups), and the average of customer payments (premiums) per year are other important variables in predicting customer churn, respectively.

Accordingly, those customers that have a higher relative frequency of the purchase, average difference paid in two consecutive years is not declining, have more length of relationship with the company. They use the company services for a long time, have purchased different insurance policies in different fields, and they have higher average payments per year, more likely to remain active. Contrary to what was thought, the variables related to customer profile have minimal impact on churn. Except for the age of the customer (which based on the model is the ninth influential variable), all other profile variables are among the low impact variables. According to what was said, the insurance company should meet valuable and profitable customers' satisfaction. Appropriate strategies in this area include retention, development, and reducing costs. Organizations need to implement resources to increase customer satisfaction and build long-term relationships with profitable customers. Also, resources should be allocated to encourage customers with profitability potential. To have loyal valuable customers, organizations should try to take into account the essential variables. That is, they should try to increase the length of the valuable customer relationship by increasing customer satisfaction with the service provided to them. This way can increase their relative purchase amount. In addition, by offering appropriate suggestions and introducing other insurance policies to them, purchase diversity of different insurance policies will be increased. As well as providing incentives for customers who use several different services leads to increased customer satisfaction and loyalty.

6. Conclusion

The churn prediction model can be used as an early warning tool for businesses and the extractor of critical factors related to customer churn and provides the possibility of providing additional helpful knowledge for decision support. This study, on the one hand, examined the factors involved in the loss of high-value customers at an insurance company in Iran, and on the other hand, performed predicting of valuable customer churn of the company. At first, essential and influential variables in the valuation and prediction of customer churn were identified. According to the purpose of this paper, which is targeted prediction modeling, and considering that targeted modeling focuses on high-value customers, in the first phase of modeling, a new customer value model is suggested .

The proposed model in this study is the combination of the LRFMP model which the L (long-term customer relationship with the company) and P (customer profitability) variables was added to the RFM and GRFM model by including product groups that the customer has purchased (G), and create a new model named LRFMPG to the valuation of customers that six variables are considered in determining the value of customers. Also, since these six variables do not have the same value in determining the value of customers, this study has used AHP with the survey of experts to determine the weight of these variables. Then valuable customers were identified based on composite scores that were calculated using weighted LRFMPG variables. In the customers' valuation based on experts' opinions, the customer purchases frequency has the highest weight and importance. After that, the length of relationship and purchased insurance policy groups have the most significant weight among LRFMPG variables. Customer profitability, the average monetary, and recency are also ranked next.

In the second phase, modeling was done initially using four different methods of prediction (logistic regression, neural networks, decision trees, and support vector machine). By comparing the performance of different prediction models built using different algorithms, found that the boosted C5.0 decision tree algorithm has a better performance compared to other models in predicting customer churn. Five important, and influential variables on prediction with this model are the length of the customer relationship, customer profitability, frequency of purchases, the maximum distance between two consecutive purchases, and mean difference of paid in two consecutive years. Then, a hybrid model was proposed. The hybrid model includes two classification techniques that the first is boosted C5.0 decision tree model (which has better performance among models), and the second is MLP neural network. In this model, the decision tree technique is used to reduce data (removing the outlier), and the second is the prediction model. The results show that compared with the single models, the proposed hybrid model has better performance. The hybrid model has high performances better than the single neural network model, 18.02% in terms of accuracy, and 14.65% in terms of the F1 measure. Also, in comparison to the single C5.0 tree, this model has outperformed, 6.73% in terms of accuracy, and 5.61% in terms of the F1 measure in predicting churn. From the importance of predictor variables in the proposed hybrid model, the relative frequency of purchases has the highest impact on being churned/active customer. After this variable, the average difference paid in two consecutive years is the most essential influential variable for prediction. The length of the customer relationship is the third important variable. The purchased insurance policies (groups), and the average of customer payments (premiums) per year are other essential variables in predicting customer churn, respectively. The variables related to customer profile have minimal impact on churn.

Several aspects can be considered in future research. First, the data used in this study are provided by a single insurance company. Data can be collected from multiple companies and compared to raise the generalization of the LRFMPG model. However, since the number of variables considered in the proposed model has little importance for the customer value analysis and prediction, therefore, future studies can apply other models of customer value analysis. Also, meta-heuristic algorithms can be used for churn prediction to raise the accuracy of prediction models.

References

[1] F. Wiersema, "The B2B Agenda: The current state of B2B marketing and a look ahead", *Industrial Marketing Management*, 4(42), pp. 470-488, 2013.

[2] S. Y. Hung, D. Yen, and H. Y. Wang, "Applying data mining to telecom churn management", *Expert System with Applications*, 31(3), pp. 515-524, 2006.

[3] T. S. H. Teo, P. Devadoss, and S. L. Pan, "Towards a holistic perspective of customer relationship management (CRM) implementation: A case study of the Housing and Development Board", Singapore, *Decision Support Systems*, 42(3), pp. 1613–1627, 2006.

[4] C. F. Tsai, and Y. H. Lu, "Customer churn prediction by hybrid neural networks", *Expert Systems with Applications*, 36(10), pp. 12547–12553, 2009.

[5] C. Hung, and C.F. Tsai, "Segmentation based on hierarchical self-organizing map for markets of multimedia on demand", *Expert Systems with Applications*, 34(1), pp. 780–787, 2008.

[6] M. J. A. Berry, and G. S. Linoff, "Data mining techniques: For marketing, sales, and customer support", *John Wiley & Sons*, 2003.

[7] L. Ryals, "Are your customers worth more than money?", *Journal of Retailing and Consumer Services*, 9(2002), pp. 241–251, 2002.

[8] J. Burez, and D. Van den Poel, "CRM at a Pay-TV Company: Using analytical models to reduce customer attrition by targeted marketing for subscription services", *Expert Systems with Applications*, 32(2), pp. 277–288, 2007.

[9] M. Xu, and J. Walton, "Gaining customer knowledge through analytical CRM", *Industrial management & data systems*, 105(7), pp. 955-971, 2005.

[10] W.F. Wang, D.A. Chiang, M.H. Hsu, C.J. Lin, and I.L. Lin, "A recommender system to avoid customer churn: A case study". *Expert Systems with Applications*, 36(4), pp. 8071–8075, 2009.

[11] K. Chen, Y. H. Hu, and Y. C. Hsieh, "Predicting customer churn from valuable B2B customers in the logistics industry: a case study", *Information Systems and e-Business Management*, pp. 1-20, 2014.

[12] B. Huang, M.T. Kechadi, and B. Buckley, "Customer churn prediction in telecommunications", *Expert Systems with Applications*, 39(1), pp. 1414-1425, 2012.

[13] W. Verbeke, K. Dejaeger, D. Martens, J. Hur, and B. Baesens, "New insights into churn prediction in the telecommunication sector: A profit driven data mining approach", *European Journal of Operational Research*, 218(1), pp. 211-229, 2012.

[14] P. Kisioglu, Y.I. and Topcu, "Applying Bayesian Belief Network approach to customer churn analysis: A case study on the telecom industry of Turkey". *Expert Systems with Applications*, 38(6), pp. 7151–7157, 2011.

[15] G. Nie, W. Rowe, L. Zhang, Y. Tian, and Y. Shi, "Credit card churn forecasting by logistic regression and decision tree", *Expert Systems with Applications*, 38(12), pp. 15273-15285, 2011.

[16] C. Fornell, and B. Wernerfel, "Defensive marketing strategy by customer complaint management: a theoretical analysis", *Journal of Marketing research*, pp. 337-346, 1987.

[17] J. H. Ahn, S. P. Han, and Y. S. Lee, "Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry", *Telecommunications policy*, 30(10), pp.552-568, 2006.

[18] V. D. Mohammadi, A. Albadvi, and B. Teymorpur, ‘Predicting Customer Churn Using CLV in Insurance Industry’, *Shiraz Journal of System Management*, 2(1), pp. 39-49, 2014.

[19] C. C. Günther, I. F. Tvete, K. Aas, G. I. Sandnes, and Ø. Borgan, “Modelling and predicting customer churn from an insurance company”, *Scandinavian Actuarial Journal*, 2014(1), pp. 587–598, 2014.

[20] M. J. Lenard, G. R. Madey, and P. Alam, “The design and validation of a hybrid information system for the auditor’s going concern decision”, *Journal of Management Information Systems*, 14(4), pp. 219-237, 1998.

[21] A. T. Jahromi, M. M. Sepehri, B. Teimourpour, and S. Choobdar, “Modeling customer churn in a non-contractual setting: the case of telecommunications service providers”, *Journal of Strategic Marketing*, 18(7), pp. 587–598, 2010.

[22] W. J. Reinartz, “On the profitability of long-life customers in a noncontractual setting: An empirical investigation and implications for marketing”, *Journal of Marketing*, Vol. 64, pp. 17-35, 2000.

[23] M. Chandar, A. Laha, and P. Krishna, “Modeling churn behavior of bank customers using predictive data mining techniques”, *National conference on soft computing techniques for engineering applications (SCT-2006)*, pp. 24-26, 2006.

[24] W. Buckinx, and D. Van den Poel, Customer base analysis: partial defection of behaviorally loyal clients in a non-contractual FMCG retail setting. *European Journal of Operational Research*, 164(1), pp. 252–268, 2005.

[25] N. Glady, B. Baesens, and C. Croux, “Modeling churn using customer lifetime value”, *European Journal of Operational Research*, vol. 197, pp. 402–411, 2009.

[26] H. Risselada, P. C. Verhoef, and T. H. A. Bijmolt, “Staying power of churn prediction models”, *Journal of Interactive Marketing*, 24(3), pp. 198–208, 2010.

[27] A. Sharma, and P. K. Panigrahi, “A Neural Network based Approach for Predicting Customer Churn in Cellular Network Services”, *International Journal of Computer Applications*, 27(11), pp.26-31, 2011.

[28] P.S. Fader, and B.G. S. Hardie, “Probability models for customer-base analysis”, *Journal of Interactive Marketing*, 23(1), pp. 61–69, 2009.

[29] C. H. Cheng, and Y. S. Chen, “Classifying the segmentation of customer value via RFM model and RS theory”, *Expert systems with applications*, 36(3), pp. 4176-4184, 2009.

[30] E. C. Chang, S. C. Huang, and H. H. Wu, “Using K-means method and spectral clustering technique in an outfitter’s value analysis”, *Quality & Quantity*, 44(4), pp. 807-815, 2010.

[31] M. C. Chen, A. L. Chiu, and H. H. Chang, “Mining changes in customer behavior in retail marketing”, *Expert Systems with Applications*, 28(4), pp. 773-781, 2005.

[32] S. M. S. Hosseini, A. Maleki, and M. R. Gholamian, “Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty”, *Expert Systems with Applications*, 37(7), pp. 5259-5264, 2010.

[33] Y. H. Liang, “Integration of data mining technologies to analyze customer value for the automotive maintenance industry”, *Expert systems with Applications*, 37(12), pp. 7489-7496, 2010.

[34] D. R. Liu, and Y. Y. Shih, “Hybrid approaches to product recommendation based on customer lifetime value and purchase preferences”, *Journal of Systems and Software*, 77(2), pp. 181-191, 2005.

[35] I. C. Yeh, K. J. Yang, and T. M. Ting, “Knowledge discovery on RFM model using Bernoulli sequence”, *Expert Systems with Applications*, 36(3), pp.5866-5871, 2009.

[36] R. S. Winer, “A framework for customer relationship management”, *California management review*, 43(4), pp.89-105, 2001.

[37] H. C. Chang, and H. P. Tsai, “Group RFM analysis as a novel framework to discover better customer consumption behavior”, *Expert Systems with Applications*, 38(12), pp.14499-14513, 2011.

[38] Y. H. Hu, and T. W. Yeh, “Discovering valuable frequent patterns based on RFM analysis without customer identification information”, *Knowledge-Based Systems*, 61, pp. 76-88, 2014.

[39] R. A. Soeini, and E. Fathalizade, “Customer Segmentation based on Modified RFM Model in the Insurance Industry”, *Proceeding of the 4th International Conference on Machine Learning and Computing*, pp. 101-104, 2012.

[40] J.T. Wei, S. Y. Lin, C. C. Weng, and H. H. Wu, “A case study of applying LRFM model in market segmentation of a children’s dental clinic”, *Expert Systems with Applications*, 39(5), pp.5529-5533, 2012.

[41] P. S. Fader, B. G. S. Hardie, and K. L. Lee, “RFM and CLV: Using iso-value curves for customer base analysis”, *Journal of Marketing Research*, 62, pp. 415–430, 2005.

[42] 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*, 66(9), pp. 1629–1636, 2013.

[43] W. Hengliang, and Z. Weiwei, “A Customer Churn Analysis Model in E-business Environment”, *International Journal of Digital Content Technology and its Applications (JDCTA)*, 6(9), pp. 296–302, 2012.

[44] E. W. T Ngai, L. Xiu, and D. C. K. Chau, “Application of data mining techniques in customer relationship management: A literature review and classification”, *Expert Systems with Application*, 36(2), pp. 2592–2602, 2009.

[45] X. Zhang, J. Zhu, S. Xu, and Y. Wan, “Predicting customer churn through interpersonal influence”, *Knowledge-Based Systems*, Volume 28, pp. 97–104, 2012.

[46] K. Coussement, and D. Van den Poel, “Improving customer attrition prediction by integrating emotions from client/company interaction emails and evaluating multiple classifiers*”, Expert Systems with Applications*, 36(3), pp. 6127–6134, 2009.

[47] V.L. Miguéis, D. Van den Poel, A. S. Camanho, and J. F. Cunha, “Modeling partial customer churn: On the value of first product-category purchase sequences”, *Expert Systems with Applications*, 39(12), pp. 11250–11256, 2012.

[48] V.L. Miguéis, D. Van den Poel, A. S. Camanho, and J. F. Cunha, "Predicting partial customer churn using Markov for discrimination for modeling first purchase sequences", *Advances in Data Analysis and Classification*, 6(4), pp. 337-353, 2012.

[49] Y. Huang, B. Huang, and M. T. Kechadi, "A Rule-Based Method for Customer Churn Prediction in Telecommunication Services", *Advances in Knowledge Discovery and Data Mining, 15th Pacific-Asia Conference*, PAKDD 2011, Shenzhen, China, May 24-27, Proceedings, Part I, pp. 411-422, 2011.

[50] K. Morik, and H. Köpcke, "Analysing customer churn in insurance data–a case study", *In European Conference on Principles of Data Mining and Knowledge Discovery*, pp. 325-336, 2004.

[51] M. Ballings, and D. Van den Poel, "Customer event history for churn prediction: How long is long enough?", *Expert Systems with Applications*, 39(18), pp. 13517-13522, 2012.

[52] I. Bose, and X. Chen, "Hybrid models using unsupervised clustering for prediction of customer churn", *Journal of Organizational Computing and Electronic Commerce*, 19(2), pp. 133-151, 2009.

[53] J. Burez, and D. Van den Poel, "Separating financial from commercial customer churn: A modeling step towards resolving the conflict between the sales and credit department", *Expert Systems with Applications*, 35(1), pp. 497-514, 2008.

[54] J. Burez, and D. Van den Poel, "Handling class imbalance in customer churn prediction", *Expert Systems with Applications*, 36(3), pp. 4626-4636, 2009.

[55] Z.Y Chen, Z.P. Fan, and M. Sun, "A hierarchical multiple kernel support vector machine for customer churn prediction using longitudinal behavioral data", *European Journal of Operational Research*, 223(2), pp. 461-472, 2012.

[56] K. Coussement, and P. Van den, "Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques". *Expert Systems with Applications*, 34 (1), pp. 313-327, 2008.

[57] K. Coussement, and P. Van den, "Integrating the voice of customers through call center emails into a decision support system for churn prediction", *Information & Management*, 45(3), pp. 164-174, 2008.

[58] K. Coussement, D. F. Benoit, and D. Van den Poel, "Improved marketing decision making in a customer churn prediction context using generalized additive models*", Expert Systems with Applications*, 37(3), pp. 2132-2143, 2010.

[59] K. W. De Bock, D. and Van den Poel, "An empirical evaluation of rotation-based ensemble classifiers for customer churn prediction", *Expert Systems with Applications*, 38(10), pp. 12293-12301, 2011.

[60] K. W. De Bock, D. and Van den Poel, "Reconciling performance and interpretability in customer churn prediction using ensemble learning based on generalized additive models", *Expert Systems with Applications*, 39(8), pp. 6816-6826, 2012.

[61] B. Huang, T. Kechadi, B. Buckley, G. Kiernan, E. Keogh, T. and Rashid, "A new feature set with new window techniques for customer churn prediction in land-line telecommunications", *Expert Systems with Applications*, 37(5), pp. 3657-3665, 2010.

[62] B. Huang, B. Buckley, and T. Kechadi, "Multi-objective feature selection by using NSGA-II for customer churn prediction in telecommunications", *Expert Systems with Applications*, 37(5), pp. 3638-3646, 2010.

[63] A. T. Jahromi, S. Stakhovych, and M. Ewing, "Managing B2B customer churn, retention and profitability", *Industrial Marketing Management*, 43(7), pp. 1258-1268, 2014.

[64] C. Kirui, L. Hong, W. Cheruiyot, and H. Kirui, "Predicting Customer Churn in Mobile Telephony Industry Using Probabilistic Classifiers in Data Mining", *IJCSI International Journal of Computer Science*, 10(2), pp. 165-172, 2013.

[65] Y. Lai, and J. Zeng, "Analysis of customer churn behavior in digital libraries", *Program: electronic library and information systems*, 48(4), pp. 370-382, 2014.

[66] S. Lessmann, and S. Voß, "A reference model for customer-centric data mining with support vector machines", *European Journal of Operational Research*, 199(2), pp. 520-530, 2009.

[67] C.S Lin, G. H. Tzeng, and Y. C. Chin, "Combined rough set theory and flow network graph to predict customer churn in credit card accounts", *Expert Systems with Applications*, 38(1), pp. 8-15, 2011.

[68] D. S. Liu, and C. H. Ju, "Customer Churn Analysis Model in Manufacturing Industry", *In Advanced Materials Research*, 69, pp. 675-679, 2009.

[69] N. Lu, H. Lin, J. Lu, and G. Zhang, "A customer churn prediction model in telecom industry using boosting", *Industrial Informatics, IEEE Transactions on*, 10(2), pp. 1659-1665, 2014.

[70] S. Nabavi, S. and Jafari, "Providing a Customer Churn Prediction Model Using Random Forest and Boosted TreesTechniques (Case Study: Solico Food Industries Group)", *Journal of Basic and Applied Scientific Research*, 3 (6), pp. 1018-1026, 2013.

[71] M. Owczarczuk, "Churn models for prepaid customers in the cellular telecommunication industry using large data marts", *Expert Systems with Applications*, 37(6), pp. 4710–4712, 2010.

[72] P. C. Pendharkar, "Genetic algorithm based neural network approaches for predicting churn in cellular wireless network services", *Expert Systems with Applications*, 36(3), pp. 6714- 6720, 2009.

[73] J. Qi, L. Zhang, Y. Liu, L. Li, Y. Zhou, Y. Shen, L. Liang, and H. Li, "ADTreesLogit model for customer churn prediction", *Annuls of Operations Research*, 168(1), pp. 247–265, 2009.

[74] C. F. Tsai, and M. Y. Chen, "Variable selection by association rules for customer churn prediction of multimedia on demand", *Expert Systems with Applications*, 37(3), pp. 2006-2015, 2010.

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

[76] W. Verbeke, D. Martens, B. and Baesens, "Social network analysis for customer churn prediction", *Applied Soft Computing*, 14, pp. 431-446, 2014.

[77] W. Verbeke, D. Martens, C. Mues, and B. Baesens, “Building comprehensible customer churn prediction models with advanced rule induction techniques”, *Expert Systems with Applications*, 38(3), pp. 2354-2364, 2011.
[78] T. Verbraken, W. Verbeke, and B. Baesens, “Profit optimizing customer churn prediction with Bayesian network classifiers”, *Intelligent Data Analysis*, 18(1), pp. 3-24, 2014.
[79] Y. Xie, X. Li, E. W. T. Ngai, and W. Ying, “Customer churn prediction using improved balanced random forests”, *Expert Systems with Applications*, 36(3), pp. 5445 5449, 2009.
[80] G. E. Xia, and W. D. Jin, “Model of customer churn prediction on support vector machine”, *Systems Engineering-Theory & Practice*, 28(1), pp. 71-77, 2008.
[81] X. Yu, S. Guo, J. Guo, X. Huang, “An extended support vector machine forecasting framework for customer churn in e-commerce”, *Expert Systems with Application*, 38(3), pp. 1425–1430, 2011.
[82] C. Bolancé, M. Guillen, and A. E. Padilla-Barreto, “Predicting probability of customer churn in insurance”, *International Conference on Modeling and Simulation in Engineering, Economics and Management* (pp. 82-91). Springer, Cham, 2016.
[83] J. R. Miglautsch, “Thoughts on RFM scoring”, *The Journal of Database Marketing*, 8(1), pp.67-72, 2000.
[84] J. Q. Zhang, A. Dixit, and R. Friedmann, “Customer loyalty and lifetime value: an empirical investigation of consumer-packaged goods”, *Journal of Marketing Theory and Practice*, 18(2), pp. 127-140, 2010.
[85] B. Luo, P. Shao, and J. Liu, “Customer churn prediction based on the decision tree in personal handy phone system service”, *International Conference on Service Systems and Service Management*, pp. 1–5, 2007.
[86] G. Li, and X. Deng, “Customer Churn Prediction of China Telecom Based on Cluster Analysis and Decision Tree Algorithm”, *Emerging Research in Artificial Intelligence and Computational Intelligence, International Conference*, AICI 2012, Springer Berlin Heidelberg, China, October 26-28, pp. 319-327, 2012.

Ali Baghaei received his master degree in e-commerce engineering from K. N. Toosi University of technology. His research interests include customer behavior and quantitative methods in marketing. He has skills and works in the field of market and customer analytics.
Email: a.baghaei@mail.kntu.ac.ir

Monireh Hosseini is currently an associate professor at the information technology department of industrial engineering faculty at K. N. Toosi University of technology. She has teaching experience in internet marketing, ecommerce strategies and management information systems, internet banking, research methods. She has published a number of research papers in international scientific journals (ISI) and conference proceedings and received scientific awards and honors (Outstanding Doctoral Research Award 2011 by Emerald Group Publishing and EFMD Global., and best papers in March 2008, January 2011, and March 2018).
Email: hosseini@kntu.ac.ir

Paper Handling Data:

Submitted: 07-26-2019
Received in revised form: 11-07-2021
Accepted: 11-21-2021
Corresponding author: Monireh Hosseini, K. N. Toosi University of Technology, hosseini@kntu.ac.ir