International Journal of Multiphysics
Volume 19, No. 1, 2025
ISSN: 1750-9548

Providing a Model to Analyses of Customer's Behavior and A Rating Model in Banking Services Using Data Mining and Neural Network

Moloukkhatoon Bozorgzadeh[1*], Saeid Salar[2]

1. Department of Mathematics, International University of Chabahar (Corresponding author)

Email: m.bozorgzadeh83@gmail.com

2. M.Sc student, Department of Information Technology Engineering, International University of Chabahar, Iran.

Abstract

The banking industry is one of the sensitive and dependent industries. Because the main part of the working capital of banks is provided through the investment of customers. Therefore, the main goal of banks is to attract new customers and retain old customers by providing optimal services and improving customer relationship management. Therefore, in this study, using data related to private banking services in Iran, we used a combination of data mining methods and artificial neural network to evaluate the performance of each in predicting customer behavior from the point of view of churn or loyalty. So, behavior of 414 customers which have accounted in private banks in Iran was evaluated based on 17 services indicators between 2014 and 2019 using SVM, ANN, SVM & ANN, SVM & FPA and ANN & FPA algorithms. The results show that the combined ANN & FPA algorithm with 94.79% accuracy has the highest performance in predicting customer's behavior.

Keywords: Customer's behavior, Banking services, Data mining, Artificial neural network (ANN)

1- Introduction

Due to the growth of knowledge and the level of customer demand, it seems that banks have not developed a codified approach to recognize and understand the behavior of their customers and better meet the needs and prevent customers from turning to other banks. Also, the low costs of changing banks for customers and joining rival banks due to the lack of bank loyalty programs to create lasting value for customers. In fact, banks can increase customer transfer costs by providing customer loyalty factors. In today's business market, much attention has been paid to the concept of communication between service organizations and their customers; But there is no clear definition of the concept of communication and, consequently, the formation of an effective commitment between the two, both from an operational and a theoretical perspective; Thus, when establishing a relationship between the customer and the organization, this relationship must be mutually understood by the parties and determined by a specific situation [1]. Having such a feature can provide an opportunity for the organization to enjoy its competitive advantages as well as its significant results. On the other hand, the existence of this type of relationship for an organization in terms of increasing sales and market share can be of particular importance. This is because organizations need to be able to satisfy their customers in any interaction, and this is possible by establishing a long-term relationship with customers and preventing them from losing contact with the organization [2]. Therefore, the formation of an effective and efficient relationship with the customer with an emphasis on relationship marketing has been considered by many researchers and given that customer satisfaction, their referrals, the formation of trust and word of mouth leads to a committed relationship with a purpose of long-term relationships and mutual communication benefits [3].

With the importance of finding customers in the turbulent field of competition between businesses, topics such as customer knowledge management and customer relationship management have been the subject of many studies

and researches. The issue of customer retention needs assessment and customer downfall forecasting has been considered by researchers and scholars for many years. Therefore, the issue of needs assessment and obtaining useful information about estimating the behavior and retaining customers in banks is very popular and profitable. Because we all know that the most important asset is not just a bank but all production and customer service organizations, and this is to the extent that it has become a global issue, and all the investments that are made to diverse products and services or Even improving the quality of service is for customer satisfaction and to prevent their dissatisfaction and loss. But one question is how important is this for banks? It should be said that having a customer is much more important for banks and such institutions than other organizations because only the main capital and financial resources of banks are provided through these customers and they are the buyers of most or even all banking services. The survival of the banks is crucial. On the other hand, banks' services, due to their special, invisible and tangible nature, which is present in the production and provision of services, are more sensitive and delicate, and ignoring them can have serious consequences for banks[4].

New technology, regulation and change in demand has caused a rise of fintech companies challenging banks' dominant position in society [5]. In times of intensified competition, customer behavior can pose a real threat for existing companies [6]. Customer turnover, also referred to as customer churn, is a common behavior when a customer leaves or ends an engagement with a company during a given time period [7]. As a result of increasing competition, it is important for banks to maintain existing customers, as this is more cost-effective than acquiring new ones, in order to ensure their position in society. In addition, new technology has increased banks' access to data, and thus data driven customer churn analysis is feasible. Taken together, there is a growing demand for customer churn analysis which studies a set of characteristics in order to predict customer churn [8]. This has intensified the demand for predictive modelling built on for example statistical learning methods. Since, if a bank can predict customer churn, targeted marketing campaigns can be used to persuade customers to keep their engagement (Ganesh, et al., 2000). However, this raises the question of which statistical learning method can predict customer churn the best?

Customer behavior is one of the new and important topics in customer relationship management (CRM). The first books on the subject were written in the 1960s, although the history goes back a long way. An example is the 1950s, when Freud's ideas were used by marketers. Customer behavior is a controversial and challenging topic that involves people and what they buy, why and how they buy, marketing and the mix of marketing and market. In the past, customer segmentation was based more on customer needs, while in recent years, with the shift in organizations from focusing on the product as a value generator to focusing on the customer as a value-generating asset, customers are based on Their value is segmented; On the other hand, with the increasing number of customer information, we are faced with a large amount of information that requires careful evaluation. Data mining, which is data analysis, is known as the interface between parts of the data and can be a valuable resource. Data mining is a complex data retrieval capability that uses sophisticated algorithms to discover patterns and correlations between data, based on which it finds and extracts data and knowledge buried in data warehouses.

In other words, data mining is one of the fundamental tools in revealing the demographics of customers, whose techniques can be used to achieve a wide range of industry goals. The term data mining is synonymous with one of the terms knowledge extraction, information retrieval, data verification, and even data dredging, which actually describes the discovery of knowledge in a database.

In this paper, we intend to use the data mining method and using hybrid methods based on neural network, to examine the behavior of customers towards the services of private banks and whether or not they fall. Therefore, in Section 2, we introduce the database. Modeling will be presented in Section 3. In section 4, we will present the results and the conclusion is presented in section 5.

2- Database

The study population in this research includes the private banks of Iran. The sample is made up of 414 costumers which receive banking servicing between 2014 to 2019. Among these customers, 143 customers were churn and the rest were not churn. This information is collected and sorted in an Excel file and analyzed using the MATLAB version 2019b software.

3- Modeling

3-1 Preprocessing data

One of the most important steps in a data mining technique is to pre-process the data of that research. In fact, preprocessing determines what leads to what results and its importance is so much that it can lead to the best result or the weakest result. Therefore, in this research, completely pre-processing is done according to the papers and in principle, which includes:

- Deleting Noise and Flare Data: It may be unnecessary for collecting data, some empty columns or data. At this stage, it is necessary to identify these data.
- Data sorting: Data must be arranged in order to be readable and readable for MATLAB software. In this research, the number of rows indicating the number of customers and the number of columns indicates the number of characteristics of the services of private banks which leads to churn or not.
- Data Labeling: Also, since mathematical software solves numerically, it is necessary that all variables be numerically labeled.
- Data Normalization Now that we have all converted them into numeric variables, we can easily load the data in MATLAB. Another point is also necessary: since the scale of data is different, we need to convert them into standard form. The standard form is to apply all the data using the following formula in interval d1 to d2:

$$\overline{x} = \frac{(x - xmin)(d2 - d1)}{xmax - xmin} + d1$$

According to data d2=+1 and d1=0 are selected. It can be seen that all data in this standard range or so-called normalized.

- Data segmentation: In this study, 70% of the data for training is used and tested using the remaining 30% of the model. These divisions are completely random so that all the data in both works can be used. Matlab software as an automatic Randperm function can generate random indices and data related to indexes created in their corresponding matrices.
- Given that the number of final data is 4284 members, 70 % of them are 2999, so 2999 data are used to create a trained model, and out of 1285 data that is not for training, it is used to test model. It should be noted that since the number of integers must be correct, the resulting number (2998.8) has to be converted to an integer of 2999.

3-2 Classification

Classification makes a model and uses it to predict the names of the classes of unknown objects to distinguish between objects belonging to different classes. The purpose of this paper is to identify customer's behavior about banking services, so the "target" column has two classes of churn and loyalty. The classifier should be able to predict for any customer's behavior that has suffered churn in accordance to services. Finally, using the evaluation criteria, that classifier is compared to other categories. In other words, it can be said that if the final answer in target is equal to 1, that is, the costumer is churned and if it is zero, then costumer will be loyal. In the following, we will introduce the methods which we will use.

- Support Vector Mechanic (SVM): Support vector machine is one of the powerful data mining tools for the purpose of categorizing two classes or multiple classes. This method can divide data into two or more categories by identifying near-page vectors. The efficiency of this method depends on the determination of the parameters and the type of kernel function.
- Artificial Neural Networks (ANN): Artificial Neural Network, which is called ANN, has a toolbox in MATLAB software. First, it must enter the input and output data of the training and testing, using the same data to create a network with the number of arbitrary neurons and number of layers specified. In this work, 20 neurons and 11 hidden layers are used and the MLP models selected for our purpose.
- Flower Pollination Algorithm (FPA): The Flower Pollination algorithm, termed FPA, is a wildly inspired algorithm inspired by bee pollinator behavior. This algorithm is presented by Yang in 2012. Pollination of flowers is an attractive process in nature. Evolutionary features can be seen in the behavior that can

be used to design a new optimization algorithm. It is estimated that more than 250,000 different species of flowering plants exist in nature, of which about 80% of the total plants are flowering plants. The main goal of a flower is ultimately reproduction through pollination, which is accompanied by the transfer of pollen, and this transfer is often associated with pollinators such as insects, birds, bats and other animals. In fact, plants and insects have evolved jointly together and They are specialized in pollination [9].

3-3 evaluation criteria

In a matter of decision making, the sample types are labeled positive or negative. In order to evaluate the performance of the proposed algorithm and to evaluate its accuracy, we use the evaluation criteria as presented below [10].
TP: The number of correct attributes correctly recognized.
TN: The number of incorrect attributes correctly recognized.
FP: The number of correct attributes that are mistakenly identified.
FN: The number of incorrect attributes correctly identified.

3-3-1 Accuracy

The ratio of the number of correct examples sorted to all samples, calculated from the following equation:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

3-3-2 Precision

The ratio is the ratio of right positive samples sorted to the total positive samples.

$$Precision = \frac{TP}{TP + FP}$$

4- Results

After implementing the preprocessing stage, the data is stored in a file called data.mat, and then logged in as the input for the next step for each type of classification and forecast. By considering all the parameters, we present the evaluation results for both the SVM algorithm and artificial neural network (ANN).

Table 1: The results of evaluating the attribute extraction resulting in prediction by SVM algorithm

Parameter	Quantity at the testing stage	Quantity at the training stage
TP	750	1225
TN	310	950
FP	101	365
FN	124	459
Precision(%)	88.13	77.04
Accuracy(%)	82.49	72.52

Table 2: The results of the evaluation of the extraction of the attribute resulting from the prediction with the ANN algorithm

Parameter	Quantity at the testing stage	Quantity at the training stage
TP	885	1534
TN	295	1069
FP	71	163
FN	34	249
Precision(%)	92.57	90.42
Accuracy(%)	91.83	86.35

In order to provide a neural network algorithm, we choose the MLP-like neural network algorithm, which uses the 17-neuron and "trainbr" function to train 17 attributes. The results of this method and its combination with the SVM algorithm are presented in Tables 1 and 2, respectively.

In the next step, the algorithm combines the neural network with the SVM algorithm and presents its evaluation results in Table 3.

Table 3: The results of the attribute extraction estimation predicted by the combined SVM & ANN algorithm

Parameter	Quantity at the testing stage	Quantity at the training stage
TP	814	1365
TN	296	921
FP	61	314
FN	114	399
Precision(%)	88.13	77.04
Accuracy(%)	82.49	72.52

In the next step, we combine the spinning algorithm with each of the above methods to examine the effect of this algorithm on each of these methods.

As shown in the tables above, the neural network algorithm method is more efficient than the SVM algorithm. On the other hand, the FPA algorithm has been shown to play a significant role in increasing the accuracy of the detection of the characteristics resulting to customer's behavior prediction. So, with a brief comparison with the analysis done so far, it can be admitted that the combination of the neural network algorithm and FPA algorithm has the highest accuracy in the training and testing phase.

Table 4: The results of evaluating the extracted property of the resulting prediction by the combined FPA & SVM algorithm

Parameter	Quantity at the testing stage	Quantity at the training stage
TP	802	1387
TN	273	903
FP	54	205
FN	96	379
Precision(%)	93.69	87.12

Accuracy(%)	87.76	79.68

Table 5: The results of estimating the extracted property of the resulting prediction by the combined FPA & ANN algorithm

Parameter	Quantity at the testing stage	Quantity at the training stage
TP	907	1596
TN	311	1102
FP	49	118
FN	18	102
Precision(%)	94.87	93.12
Accuracy(%)	94.79	92.46

Table 7 presents a comparison between the results of this study.

Table 6: Comparing the results of algorithms

Parameter	ANN & FPA	SVM & FPA	SVM & ANN	ANN	SVM
TP	907	802	814	885	750
TN	311	273	296	295	310
FP	49	54	61	71	101
FN	18	96	114	34	124
(%) Precision	94.87	93.69	88.13	92.57	88.13
(%) Accuracy	94.79	87.76	82.49	91.83	82.49

4-1 Sensitive Analyses

Also, to investigate the accuracy of the algorithms, we use the ROC, the first type error, and the Kappa statistics to determine the sensitivity analysis. The AUC represents the level below the ROC chart, the higher the number is for a larger batch, the better performance of the bundle is evaluated more favorable. The ROC chart is a method for checking the performance of the batches. In fact, ROC curves are two-dimensional curves in which the DR is the same as the positive detection rate on the Y-axis and the FAR, respectively, or the wrong detection rate on the X-axis:

$$AUC = \frac{TP + TN}{TP + TN + FN + TP}$$

Also, the first type error means that insolvent companies are mistakenly subjected to a model by a group of insolvent companies.

$$error\_type\_1(FAR) = \frac{FP}{TN + FP}$$

Finally, the Kappa-Cohen criterion is calculated as follows:

$$k = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)}$$

Where Pr (α) are correct observational values and Pr (*e*) are observable values. Thus, Table 8 presents the results of sensitivity analysis of algorithms.

Table 7: Sensitivity analysis of algorithms

Criterion	**The first error value**	**AUC**	**Kappa**
SVM	20.93	0.756	0.6932
ANN	2.35	0.983	0.9324
SVM & ANN	6.87	0.813	0.7154
SVM & FPA	19.05	0.830	0.7487
ANN & FPA	0.00	1.00	1.00

Finally, we will look at customer's behavior predictions from the years 2014 to 2019 (a six-year period). To do this, we use the MLP-type neural network algorithm in combination with a FPA algorithm which had the highest accuracy (based on the results of Tables 6 and 7).

5- Conclusion

In today's world economy, having accurate and timely information for owners, investors, creditors and other interest groups is very useful for making financial decisions. With the development of technology, the use of simple customer's behavior prediction models is possible for all groups specially banking industry. Churn prediction is a powerful tool for businesses to maintain long-term relationship with their customers. However, limited work has been done in capturing customer behavior about banking services on longitudinal data. This research developed a hybrid classification approach based on data mining and neural network to optimizing model specifications for predicting customer behavior decisions over time. The approach was used to identify duration of training data, length of prediction window and the number of multiple time periods to support accurate prediction.

Based on the obtained and compared results, it is shown that the artificial neural network (ANN) algorithm has better performance than the SVM. It can also be seen that the FPA algorithm has resulted in an average improvement of about 9% in increasing the efficiency of the methods. In order to increase the efficiency of this research, more features can be used in addition to the 17 features studied in this paper, which improves the modeling performance of this research. For future research in customer behavior analysis, methods that yield high interpretability and high predictability would be interesting to study. This is, develop hybrid methods that combine interpretability to ANN with predictability from both ANN and SVM. To be able to combine the knowledge of if a customer ends their engagement, as well as understanding why. Taken together, this could not only prevent customer churn but also help outline which variables affect customer churn and the bank could use this information to improve their business and customer relations. Moreover, interesting future research would be to analyze at which dataset size leave-one-out cross-validation and *k*-Fold cross-validation yields the same results.

References

[1] Jain H, Yadav G, Manoov R. Churn Prediction and Retention in Banking, Telecom and IT Sectors Using Machine Learning Techniques. InAdvances in Machine Learning and Computational Intelligence (pp. 137-156). Springer, Singapore.

[2] Satria WA, Fitri I, Ningsih S. Prediction of Customer Churn in the Banking Industry Using Artificial Neural Networks. Jurnal Mantik. 2020 May 31;4(1, May):936-43.

[3] Davagdorj K, Ryu KH. Prediction of Bank Customer Behavior using Multivariate Adaptive Regression Splines.

[4] Le-Khac NA, Markos S, Kechadi MT. Towards a new data mining-based approach for anti-money laundering in an international investment bank. InInternational Conference on Digital Forensics and Cyber Crime 2009 Sep 30 (pp. 77-84). Springer, Berlin, Heidelberg.

[5] Thirlwell M. The economist: Bond market blues. Company Director. 2019 Oct;35(9):26.

[6] De Caigny A, Coussement K, De Bock KW, Lessmann S. Incorporating textual information in customer churn prediction models based on a convolutional neural network. International Journal of Forecasting. 2019 Aug 21.

[7] Colgate M, Stewart K, Kinsella R. Customer defection: a study of the student market in Ireland. International Journal of Bank Marketing. 1996 Jun 1.

[8] De Caigny A, Coussement K, De Bock KW. A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. European Journal of Operational Research. 2018 Sep 1;269(2):760-72.

[9] Jabeur, S. B. (2017). Bankruptcy prediction using partial least squares logistic regression. *Journal of Retailing and Consumer Services*, *36*, 197-202.

[10] Mselmi, N., Lahiani, A., & Hamza, T. (2017). Financial distress prediction: The case of French small and medium-sized firms. *International Review of Financial Analysis*, *50*, 67-80.