

Estimation of the Probability of Bank Customers by Artificial Neural Networks

Lecturer Yavuz Selim Balçoğlu*, Bülent Sezen

Faculty of Management, Gebze Technical University, Turkey

Received December 23, 2019; Revised January 30, 2020; Accepted February 7, 2020

Copyright©2020 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract One of the most important issues in today's banking sector is that they want to add new customers to their bodies and want to keep their current customers. For this, banks spend a lot of money. Because of the methods they use, they either have to keep information flow to all their customers or they have to focus on their customers, whose traditional methods of probability leave. In this article, the probability of bank customers left by artificial neural networks is estimated. Presently, with the improvement of technology, a growing number of banks, holding existing clients for banks and combining new clients into their systems have earned significance. As the Bank's efficiency, it is essential to define the clients with the contingency of dropping within existing clients. The client pool generated by the classical methods utilized leads to the introduction of activities on a major number of groups for the bank. This outcomes in higher expenses for banks. The main purpose of this paper is to lay the foundation for further research for precision of who will the bank as a customer. The findings of our study with artificial neural networks have described a minimal and more compressed group as clients who are likely to leave. In this way, it is foreseen that the likely costs of banks will minimize. As a result of this study, the most accurate estimation was obtained by educating artificial neural networks with the most accurate values.

Keywords Artificial Neural Networks, Prediction Models, Bank Customers

1. Introduction

The banking sector is an important factor in economic development. In today's economic structure, banks develop a wide range of products and services for both institutions and individuals. Banks are constantly working to serve customers in their portfolios and to add new customers.

However, in every sector, as in the company, the main problem here is to keep your customers and not to lose them. When we consider the size of banks today, it becomes difficult to predict which customers will leave the institution. Banks use different methods for these. The most widely used method is to keep all customers constantly active, and keep them informed of innovations and developments taking place within the company at regular intervals. However, when we look at the size of the dimensions, this creates a serious cost. To minimize this cost, this effort should be made to the right customer cluster.

"Acquiring customers" and "keeping customers" are two very critical factors for companies ... To achieve this, trillions are spent and special strategies are developed. However, the cost of the new customer is not the same for every sector ... Some sectors are easier to do ... For example, while a high cost is spent for a new customer in the retail sector, the automobile is much higher than that. Retaining the existing one usually requires half the budget for the "new".

Costs cause the new customer to stay in the background. At least 2-3 times the cost of maintaining the existing customer is spent for each new customer. Therefore, this cost makes many banks think. Banks, who are afraid of cost, expect their satisfied customers to positively affect the people around them in order to increase their market share. Only banks that adhere to this method cannot contribute to the growth of the market.

However, besides protecting the existing customer, banks trying to expand their customer portfolio are extremely successful in increasing their market share. Various methods are used to acquire new customers. Intense advertising and promotional campaigns are considered the most effective way. Apart from this, targeting the masses that will start consumption enriches the customer portfolio. Another method is to draw the attention of the rival bank customer with different products and services.

In today's job market, customer satisfaction is one of the top priorities of the companies. Interaction equipment with

the best technology is used more and more in the banking sector to control and analyze customer expectations and experiences. A concept established to create this interaction: Customer Communication Management (CRM).

As in many other sectors, it is important for banks to keep their communications with their customers up-to-date, in order for companies to maintain their assets successfully and profitably. In order to get in touch with customers, banks obtain information sources from both traditional and digital channels such as telephone, information form, customer request form, etc.

Keeping customer always opens new potential. The more customers you have in the market you are in, the more innovation and diversity you need to work, you need to work customer-oriented and bring different features to your products. As long as you keep your customers in their system, it is necessary to follow the technology, to be proactive in the market, to produce new products in line with the needs. This means continuous growth and innovation.

In the literature, there are not many studies about the separation of bank customers from the bank. Two articles have been written on this subject. The first of these, "an exploration of customer exit in retail banking," was written by Kate Stewart. According to Stewart, by approaching the customer through the right communication channels, he revealed that the problems that arise can be solved and that the customer's feelings and thoughts should be taken into account [1]. The second study, "a model of customer loyalty in the retail banking market," was written by Asuncion Beerli, Josefa D. Martin and Agustin Quintana. In this study, Beerli and his colleagues developed a model on the loyalty of bank customers. They examined the effects of this model on customer loyalty. They have shown that the loyalty of bank customers is less costly than leaving the bank and moving to another bank, affecting the customer's decision to leave and making this action difficult [2].

There are several studies on banking transactions and artificial neural networks. If we examine them in order;

Gan, Christopher; Limsombunchao, V.; Clemes, Michael D.; Weng, Yong Y. A. 2005 Consumer choice prediction: artificial neural networks versus logistic models "was written in 2005. This article examines the choice of customers for electronic or traditional banking transactions on the New Zealand people [3].

Melek Acar Boyacıoğlu, Yakup Kara wrote in 2007 "Comparison of the performance of artificial neural networks and multivariate statistical analysis techniques in estimating financial strength ratings in Turkish banking sector". In this paper, a model has been developed to predict the financial strength of Turkish banks by using ANN and

multi-parameter statistical analysis methods [4].

N. M. Abu Bakar and I. M. Tahir wrote M. "Applying Multiple Linear Regression and Neural Network to Predict Bank Performance" in 2009. In this article, a bank performance estimation study is presented using linear regression and neural network methods [5].

In 2010, Melike Kurtaran Çelik wrote "The Prediction of Financial Failures of Banks by Traditional and New Methods". This article in particular application models for which the most important factors of the financial structure of banks in Turkey aimed at building the prediction of financial unsuccessfulness. In this study, the prediction power of the equations analysis and Artificial Neural Network models, which are widely used in the literature, are compared with various prediction studies that have been formed from past to present [6].

Utku Altınöz wrote in 2013 The Predictability of Banks' Financial Failures in the Framework of Artificial Neural Networks. In this article, artificial neural network model, which is one of the multiple statistical theories used in predetermining bank failures, has been tested in parallel with the literature studies. [7].

Mohsen Nazari and Mojtaba Alidadi wrote Measuring Credit Risk of Bank Customers Using Artificial Neural Network 2013 in 2013. In this article, credit risk ratio of bank customers is measured by using artificial neural networks [8].

1.1. Artificial Neural Network

The first studies on artificial neural networks remained mostly theoretical before 1940. The first artificial neural network study was developed by W. McCulloch and W. Pitts in 1943 [9].

With the discovery of the "single layer neural network model by F. Rosenblatt in 1957, the studies in the ANN area gained momentum. The concept of a monolayer neural network model is an artificial neural network which can be grown on a single layer and has a single outlet. In 1958, Rosenblatt conducted studies on the concept of creating feedback consistent with random maturing networks [10].

In 1987, the Department of Electrical and Electronics Engineering convened a congress based on the first ANNs with nearly 2000 people. Broomhead and Lowe developed the "center-base functions" model in 1988 [11].

This network model has been developed as an alternative to multi-layer sensors. However, it has achieved much more effective success in filtering operations. From this point of view, Specht developed the more advanced models of these network models called "Multiple Variable Networks" [12] and "Randomness Networks" [13].

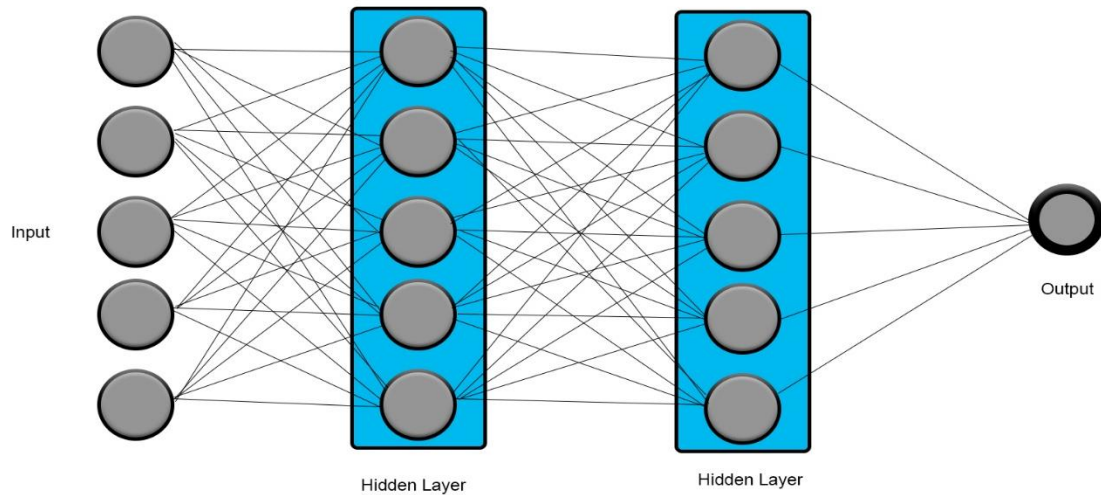


Figure 1. Working structure of artificial neural networks

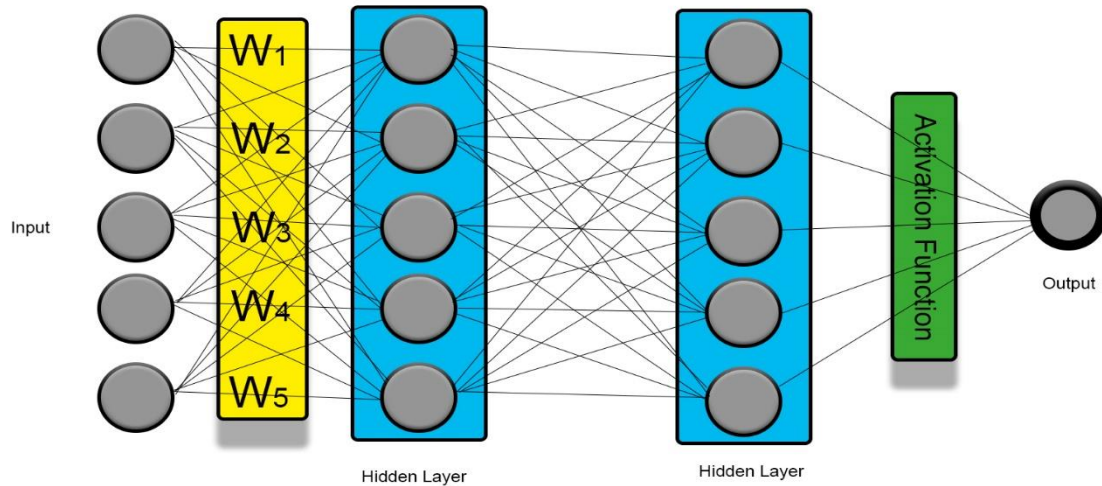


Figure 2. Internal structure of artificial neural networks.

Artificial neural networks have the ability to produce new results from current observations at the stage of learning [14].

Artificial neural networks which emerge with the formation of the nervous system show the characteristic structure of biological nervous system in terms of parallel operation and self-learning ability. Among other features, because the mutual co-operation of information more quickly penetrates and hardware can be realized easily in a way that makes artificial neural networks more practical than other methods .

Although there are many types of networks developed according to the type of problems in the studies, Multilayer Sensors are widely used in classification and prediction problems. This network is also known as a backpropagation network. Multi-based sensors were developed in 1986 by Rumelhart et al. [15]. Multilayer sensor networks operate according to self-learning purpose. In broader terms; these networks show both input data and the output data expected

to be generated in response to those inputs during training. The purpose of the network is to generate the output corresponding to that input for each input and transmit it to the next neuron [16].

In multilayer networks, during the training phase, the weight values are constantly changed to minimize the error between the actual result and the output value generated by the network. As soon as the error is minimized, the learning process is realized and the learning with the best results is obtained [17].

It is possible to examine a prediction system established in order to predict the future in three steps. The first step is model building, the second is the prediction stage, and the third is the comparison of the actual results with the results obtained. Since the predictions will be realized according to the established model, the model building stage is very important. Because the use of the correct data set becomes important at this stage. The correct structure established will enable the production of more realistic predictions [18].

Python is one of the most common methods for scientific computing as a programming language. It is the right choice for algorithmic studies and data analysis with its highly developed code structure and huge libraries [19]. However, it is growing and used not only in academic settings but also in industry. This software meets the growing need for statistical data analysis by experts in the web industry, as well as experts in areas other than computer science, such as biology or physics. The best source for Python with the available books is the Python website [20].

In this research, a pattern was set up over 10 independent argument data sets of a European bank with 10000 clients by using artificial neural networks in calculating the probability of dropping bank clients.

2. Data and method

In the model developed with ANN, the data set of a European bank with 10,000 customers was used. Anaconda software was used in the development of ANN model. Python was used as the coding language. Stochastic Gradient Descent was applied as the method. Table 1 shows the dependent and independent variables. It is divided into 75% learning and 25% test set over 10,000 data. In the model, between 12 and 30 hidden layers are used.

Table 1. Variables used in the study

Dependent Variables	Independent variables
Possibility of Customer Leaving	Credit Score
	Geography
	Gender
	Age
	Client for year
	Money in client's account
	Number of services used by the client
	Clients owner credit card
	Active Client
	Client Salary figure

3. Findings

In the established model, the best results were obtained in 30 hidden layered structures. The estimated success rate is 86.47%. Neuron was obtained on 947. According to the results, 1400 out of a total of 10,000 bank customers were highly likely to leave the bank. Dependent variable, the probability of separation is between 0 and 1, values with probability of separation above 0.50 are considered to be 1 and values below are considered as 0.

Listed below are some sample results of artificial neural networks over the model established as a single client.

Example 1: The ANN forecast for the data set “0.5, 0,

600, 1, 40, 3, 1000, 2, 1, 1, 50000 was found False”. This client does not intend to leave the bank.

Example 2: ANN forecast for dataset “0.0, 1, 600, 1, 20, 3, 1000, 0, 1, 1, 1000” was found “True”. This client is considering leaving the bank.

Table 2. Percentages estimated using artificial neural networks

Number of Trials	Forecast Success
1	0,7963
100	0,7963
200	0,7943
300	0,8269
400	0,8553
500	0,8605
600	0,8632
700	0,8633
800	0,8619
947	0,8647

When the ANN model of the 3 variables selected among our independent variables is established, the probability of leaving customers account is obtained from the following graph. In the graphical representations of these data, Python data visualization was applied.

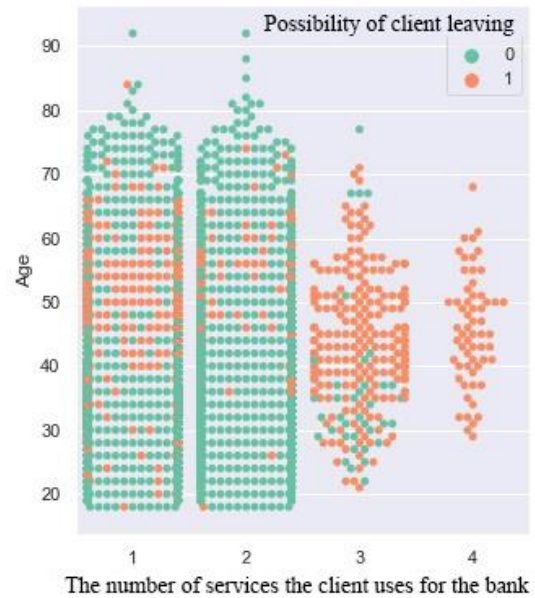


Figure 3. Results of the client's probability of leaving by the number of services used by the bank

As can be seen from this graph, as understood here, clients do not plan to leave the bank as they benefit from the bank's services. The region with the highest probability of leaving is seen among clients using 3 of the bank's services. Graphical analysis also shows that the region where clients are most likely to leave is concentrated between the ages of 40 and 60. However, as the number of services they receive from the bank increases, this age rate decreases.

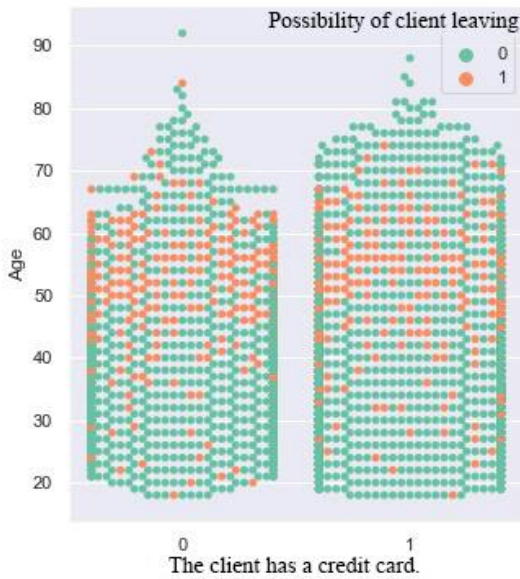


Figure 4. The results of the clients's probability of leaving according to the client's credit card holder

As can be seen in this graph, client with credit cards issued by the bank are less likely to leave. Clients who use credit cards are more numerous than clients who do not. However, when the possibilities of leaving are analyzed, the thoughts of leaving, whether they have a credit card or not, are less than the general distribution.

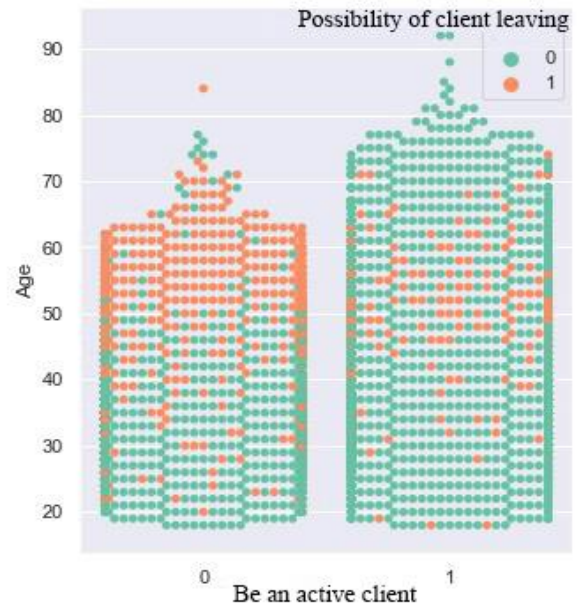


Figure 5. The probability that the client will leave, depending on whether the bank is an active client

In this graph, clients who are not active clients of the bank reach high level in middle age and above, although they are more likely to leave the bank. On the other hand, it is seen that the probability of active clients leaving the bank is very low.

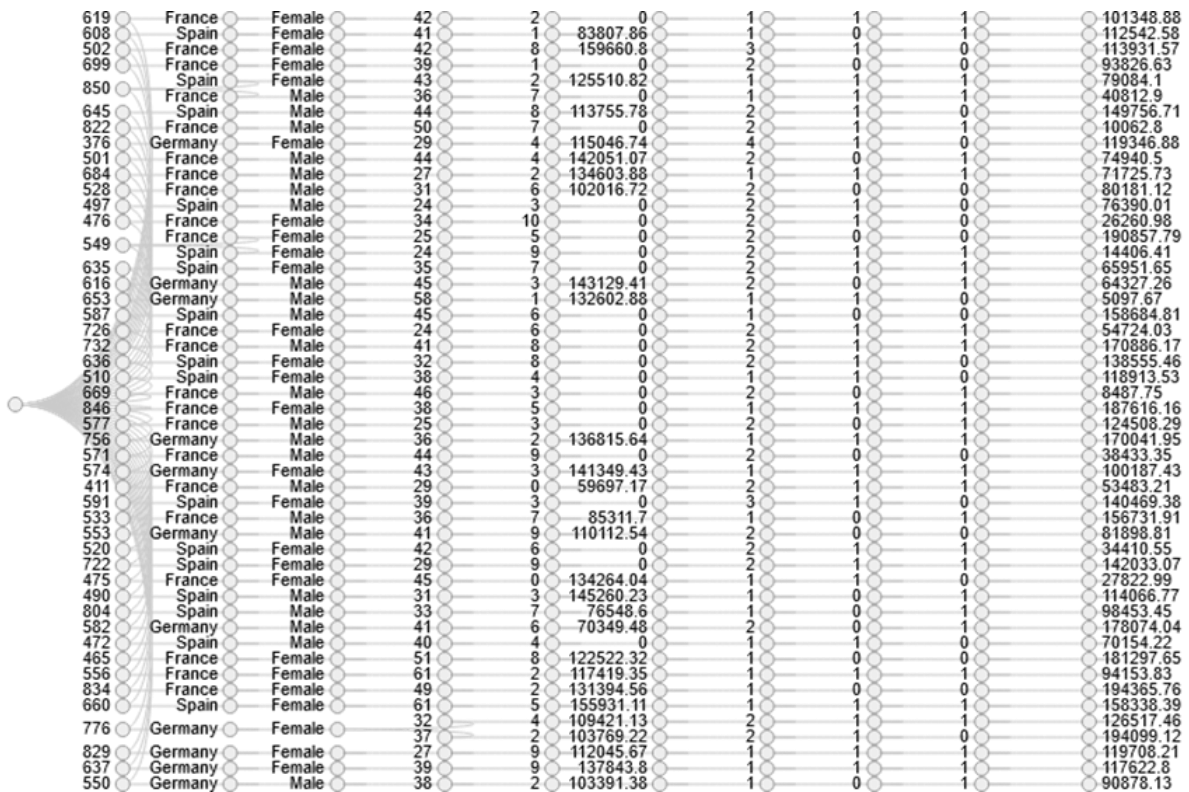


Figure 6. Sequence of the first 50 data obtained with Artificial Neural Networks

Figure 6 shows the sequence and order of entry through the artificial neural networks.

4. Conclusions

In this research, artificial neural networks are used to estimate the possibility of bank clients dropping. Artificial neural networks grant best outcomes than many other methods. One of the most significant causes for this is artificial neural networks, which involves more data and the higher the number of data, the better the find out and practice phases. In this research, credit score, geography, gender, age, client for years, money in client's account, number of services used by the client, client's owner credit card, active client, client salary figure, independent variables were used in forecasting the probability of dropping bank clients. While these arguments' variables were the inputs of the artificial neural network model, possibilities of dropping clients as the dependent variable were used. The results gained with artificial neural networks comprise a forecast model for banks. While it is a cost-reduction factor for banks in the suitable use, it will activate banks to use their energy properly. The results obtained in this study show us that the results of the test and training values of the dataset were handled precisely. The computer provided over 10000 models over 30 hidden layers with a total of 947 neuron. This study shows that choosing the sensitive values here will ensure the success of the analysis. However, the point to be remembered here is the high estimation rate of the study. As a result of this study, the most accurate estimation was obtained by educating artificial neural networks with the most accurate values.

Among the variables used, geography, gender and age variables only have an effect on dividing the analysis into demographic variables. The reason they are included in the analysis is that this data set will show demographic variability in future studies. Our independent variables, other than these variables, directly affect the likelihood of bank customers leaving apart from demographic factors.

In the following studies, it is possible to compare the statistics of the turnover rates of the customers belonging to previously obtained banks with ANN model.

REFERENCES

- [1] Stewart, K. (1998), "An exploration of customer exit in retail banking", *International Journal of Bank Marketing*, Vol. 16 No. 1, pp. 6-14. <https://doi.org/10.1108/02652329810197735>
- [2] Beerli, A., Martín, J. and Quintana, A. (2004), "A model of customer loyalty in the retail banking market", *European Journal of Marketing*, Vol. 38 No. 1/2, pp. 253-275. <https://doi.org/10.5267/j.ac.2016.8.002>
- [3] Gan, C., Limsombunchai, V., Cledes, M., Weng, A., Division, C., Box, P. O., & Zealand, N. (2005). *Consumer Choice Prediction: Artificial Neural Networks versus Logistic Models* Christopher Gan, Visit Limsombunchai, Mike Cledes and Amy Weng Vol. 1, Issue 4 Pages 211-219 doi : 10.3844/jssp.2005.211.219
- [4] Melek Acar Boyacıoğlu. Yakup Kara. (2007). *Türk Bankacılık Sektöründe Finansal Güç Derecelerinin Tahmininde Yapay Sinir Ağları Ve Çok Değişkenli İstatistiksel Analiz Tekniklerinin Performanslarının Karşılaştırılması*. Dokuz Eylül Üniversitesi İktisadi İdari Bilimler Fakültesi Dergisi, Vol., Issue 2 (2007)
- [5] Bakar, N. M. A., & Tahir, I. M. (2014). *Applying Multiple Linear Regression and Neural Network to Predict Bank Performance*. *International Business Research*. Vol. 2, No. 4 (2009) <https://doi.org/10.5539/ibr.v2n4p176>
- [6] Kurtaran Çelik, M. (2010). *Bankaların Finansal Başarısızlıklarının Geleneksel Ve Yeni Yöntemlerle Öngörüsü*. *Yönetim Ve Ekonomi: Celal Bayar Üniversitesi İktisadi Ve İdari Bilimler Fakültesi Dergisi*, Vol. 17, Issue: 2, (2010)
- [7] Altunöz, U. (2013). *Bankaların Finansal Başarısızlıklarının Yapay Sinir Ağları Modeli Çerçevesinde Tahmin Edilebilirliği*. *Dokuz Eylül Üniversitesi İktisadi İdari Bilimler Fakültesi Dergisi*, Vol. 28, Issue: 2, (2013)
- [8] Nazari, M., & Alidadi, M. (2013). *Measuring Credit Risk of Bank Customers Using Artificial Neural Network*. *Journal of Management Research*. Vol 5, No 2 (2013) <https://doi.org/10.5296/jmr.v5i2.2899>
- [9] Ahmed, S. (2010), "Multi-directional search to optimize neural network error function", *Kybernetes*, Vol. 39 No. 7, pp. 1145-1166
- [10] Ziols, R. (2019), "The Perceptron: A Partial History of Models and Minds in Data-Driven Educational Systems", Jules, T. and Salajan, F. (Ed.) *The Educational Intelligent Economy: Big Data, Artificial Intelligence, Machine Learning and the Internet of Things in Education* (International Perspectives on Education and Society, Vol. 38), Emerald Publishing Limited, pp. 51-67.
- [11] Wang, W., Xu, Z. and Weizhen Lu, J. (2003), "Three improved neural network models for air quality forecasting", *Engineering Computations*, Vol. 20 No. 2, pp. 192-210
- [12] Khan, W., Chung, S., Awan, M. and Wen, X. (2019), "Machine learning facilitated business intelligence (Part I): Neural networks learning algorithms and applications", *Industrial Management & Data Systems*, Vol. 120 No. 1, pp. 164-195.
- [13] Mostafa, M. (2010), "A neuro-computational intelligence analysis of the US retailers' efficiency", *International Journal of Intelligent Computing and Cybernetics*, Vol. 3 No. 1, pp. 135-162.
- [14] Delen, D., Sharda, R., & Bessonov, M. (2006). *Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks*. *Accident Analysis and Prevention*. Vol. 38, Issue 3, May 2006, Pages 434-444 <https://doi.org/10.1016/j.aap.2005.06.024>
- [15] Church, K. B., & Curram, S. P. (1996). *Forecasting consumers' expenditure: A comparison between*

econometric and neural network models. *International Journal of Forecasting*. Vol. 12, Issue 2, June 1996, Pages 255-267 [https://doi.org/10.1016/0169-2070\(95\)00631-1](https://doi.org/10.1016/0169-2070(95)00631-1)

- [16] PapatyaYayincilik, E. Ö.-, Istanbul, U., & 2003, U. (2008). Yapay sinir ağıları KURULUŞ. In papatyabilim.com.tr NNNN.
- [17] Velo, R., López, P., & Maseda, F. (2014). Wind speed estimation using multilayer perceptron. *Energy Conversion and Management*. Vol. 81, May 2014, Pages 1-9 <https://doi.org/10.1016/j.enconman.2014.02.017>
- [18] Rezaei, S., Shokouhyar, S. and Zandieh, M. (2019), "A neural network approach for retailer risk assessment in the aftermarket industry", *Benchmarking: An International Journal*, Vol. 26 No. 5, pp. 1631-1647.
- [19] Millman, K. J., & Aivazis, M. (2011). Python for scientists and engineers. *Computing in Science and Engineering*. Vol. 13, Issue: 2, March-April 2011 <https://doi.org/10.1109/MCSE.2011.36>
- [20] Python Language Reference. <https://www.python.org>