Çevik, M., & Tabaru-Örnek, G. (2020). Comparison of MATLAB and SPSS software in the prediction of academic achievement with artificial neural networks: Modeling for elementary school students. *International Online Journal of Education and Teaching (IOJET), 7*(4). 1689-1707.

http://iojet.org/index.php/IOJET/article/view/999

Received: 21.07.2020
Received in revised form: 25.08.2020
Accepted: 10.09.2020

COMPARISON OF MATLAB AND SPSS SOFTWARE IN THE PREDICTION OF ACADEMIC ACHIEVEMENT WITH ARTIFICIAL NEURAL NETWORKS: MODELING FOR ELEMENTARY SCHOOL STUDENTS

Research Article

Mustafa Çevik

Karamanoğlu Mehmetbey University

mustafacevik@kmu.edu.tr

Gizem Tabaru-Örnek

Karamanoğlu Mehmetbey University

gizemtabaru@kmu.edu.tr

Mustafa Çevik is an Assoc. Prof. Dr. in the Faculty of Education in Karamanoğlu Mehmetbey Univesity. He broadly conducts research in STEM Education. Specifically, he has worked on Vocational High School Students in STEM Education.

Gizem Tabaru-Örnek is currently continuing her doctorate education at the Gazi University Educational Sciences Institute. She is working as a research assistant at the Education Faculty of Karamanoglu Mehmetbey University

COMPARISON OF MATLAB AND SPSS SOFTWARE IN THE PREDICTION OF ACADEMIC ACHIEVEMENT WITH ARTIFICIAL NEURAL NETWORKS: MODELING FOR ELEMENTARY SCHOOL STUDENTS

Mustafa Çevik
mustafacevik@mail.com

Gizem Tabaru-Örnek
gizemtabaru@mail.com

Abstract

In this study, it was aimed to compare the predictions of the academic achievement of the artificial neural networks (ANN) run in MATLAB and SPSS software and to determine the factors related to their academic achievement. Sample consisted of 465 students who were studying at Grade 4 in primary schools in the Central Anatolian Region of Turkey in 2017. A 12-questions questionnaire was used as the collection tool. For the content validity of the questionnaire, expert opinions were received. The KR20 reliability coefficient was calculated as .60. An exploratory factor analysis was run for the construct validity. In the ANN model, the items related to the academic achievement in the questionnaire were considered as independent variables / inputs, and the academic achievements of the previous year as the dependent variables / outputs. The predictions of the academic achievement of the ANN models were analyzed in MATLAB R2013a and SPSS 24.0 software and the regression coefficients of the independent variables were examined. It was found that MATLAB software had a higher rate of the correct prediction compared to SPSS. In the regression coefficients of the independent variables, some differences and similarities between the results of MATLAB and SPSS were found.

Keywords: Artificial neural network, academic achievement prediction, MATLAB, SPSS

1. Introduction

Education includes all of the social, cultural and personal processes that individuals go through to acquire desired talents, skills and other types of behavior that are valuable in the society. Through education, people's ideas, knowledge, goals, perspectives, attitudes, and moral criteria can continuously change. Thus, education is a lifelong process that starts in the family and continues throughout the life (Demirel & Kaya, 2015).

Along with the developments of technology and globalization, big and inexorable changes have occurred in society and in other fields of life. Therefore, educational standards should be increased to keep pace with these innovations and changes brought with the 21st century. In the 21st century, scientific predictions on students' academic achievement have been given a great importance as they have a significant impact on educational policies. The importance of scientific predictions on students' academic achievement is a worldwide accepted fact in order

to be able to look at the future more firmly, to realize the risks that life brings to us and to deal with the problems encountered. Individuals are expected to be able to adapt to the rapid developments in the world and make important decisions about the future. While planning the future, it is vital to make predictions for academic achievement as well as providing professional consultant services to students about their future career involving the school-family-state union. In this view, it becomes obvious that scientific predictions on students' academic achievement would be beneficial in individuals' career development (Bahadır, 2013).

Many different variables have been used to predict students' academic achievement by researchers (Bekele & McPherson, 2011; Fenollar, Roman & Cuestas, 2007; Kuncel, Hezlett & Ones, 2004; Kyndt, Musso, Cascallar & Dochy, 2015; Mukta & Usha, 2009; Turner, Chandler & Heffer, 2009). In addition, many national and international tests (e.g. Programme for International Student Assessment) have been conducted to evaluate students' academic achievement. Based on results on these tests countries are compared to each other to monitor their educational success. In these tests, students are usually asked to answer questions measuring their academic achievement and others variables related to their academic achievement. However, results of these international tests have showed that the student academic achievement is not at the satisfactory level. One reason could be that all variables affecting student academic achievement are not correctly included. If all the variables influencing student academic achievement can be revealed, students' academic acvhievement may reach at the satisfactory level (Tepehan, 2011). Moreover, a better prediction of academic achievement may have potential to contribute positively to the individual's learning process. These predictions are vital because it enable educators to early diagnose the low academic achievement and take required steps to turn a failure into success.

Many studies have focused on predicting students' academic achievement (Colom, Escerial, Chun Shih & Privado, 2007; Hailikari, Nevgi & Komulainen, 2008; Krumm, Ziegler & Buehner, 2008). In these studies, different methods have been used by researchers. Methods used in education are generally the statistical techniques including regression, discriminant analysis and structural equation modeling. Because these statistical methods have their own limitations in the prediction of academic achievement, in recent years another novel method, called Artificial Neural Networks (ANN), have emerged. The ANN has already started to be used in some other fields including medicine, engineering, meteorology and economics. Research has reported that the ANN is a strong tool to make accurate predictions (Musso, Kyndt, Cascallar & Dochy, 2013).

One advantage of the ANN model is that it allows research make complex and nonlinear model. In educational research, it is a strong tool as the relationships between the variables regarding academic achievement are rather complex and has a nonlinear structure. In recent years, the ANN has become preferred in many areas because it is an effective tool to examine the relationships between variables, classifying and predicting the results with a high accuracy. This method also allows researchers to evaluate multiple variables simultaneously. Moreover, the ANN reduce the negative results of the failure by predicting the students' possible low academic achievement in the future and increase the chances being successful. Thus, this is an important tool for students' future academic achievement predictions. Similarly, examining variables related to the academic achievement among high achiever students may provide an understanding of factors causing to positive results (Musso & Cascallar, 2009; Musso, Kyndt, Cascallar & Dochy, 2013).

1.1. Artificial Neural Networks

The ANN model has emerged as a product of human's endeavor to discover the Nature. The ANN is parallel the information processing structures that are inspired by the human brain in which each consists of neurons with its own memory. In other words, it is a method designed to simulate the biological nervous system. A nerve cell contains neurons linked to each other in various ways in a form of network. These networks are used to reveal the relationship between the variables (Bahadır, 2013; Budak & Erpolat, 2012).

Learning in biological systems occurs by adjusting synaptic connections between neurons. In other words, since their birth, humans have learnt thorugh a process called “learning by living” (Figure 1).

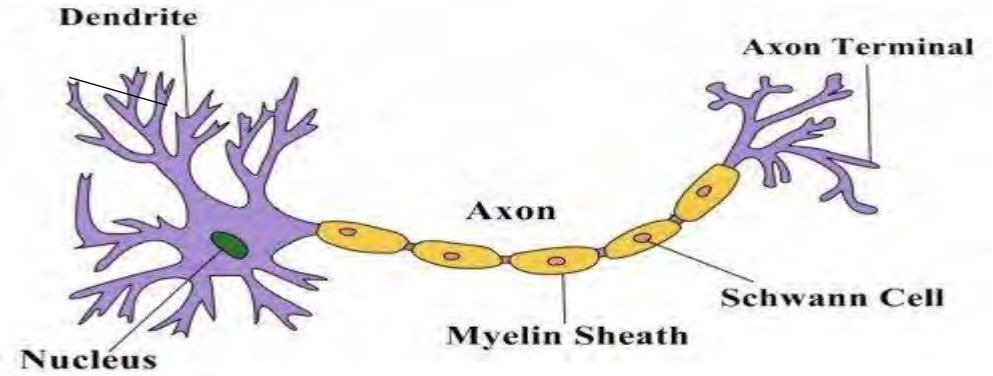

Figure 1. Structure of a nerve cell (Towle and Bradley, 2006)

The general structure of a neuron is as shown in Figure 1. The parts and functions of the neuron are as follows (Öztemel, 2006; Yurtoğlu, 2005):

Dendrites; Thse parts take electrochemical signals from other nerve cells and deliver them to the nucleus. Hundreds of dendrites can come out of a nerve cell. Nucleus: The nucleus is the body of the nerve cell. It is a microprocessor unit where the information transmiited through dendrites are combined and made sense, briefly processed and the outputs are transferred to the axon. Axon; This part sends electrochemical signals and outputs produced by the nerve cell to other nerve cells. There are many synapses at the tip of an axon. Synapses; They are the junction points that link the axon to the dendrites of other nerve cells Just as biological neural Networks are made up of nerve cells, the ANN is made up of artificial neural cells. Artificial nerve cells, which are the basic processing units of an ANN, simulate the four basic elements of the biological nerve cells described above. The general representation of an artificial nerve cell is displayed in Figure 2.

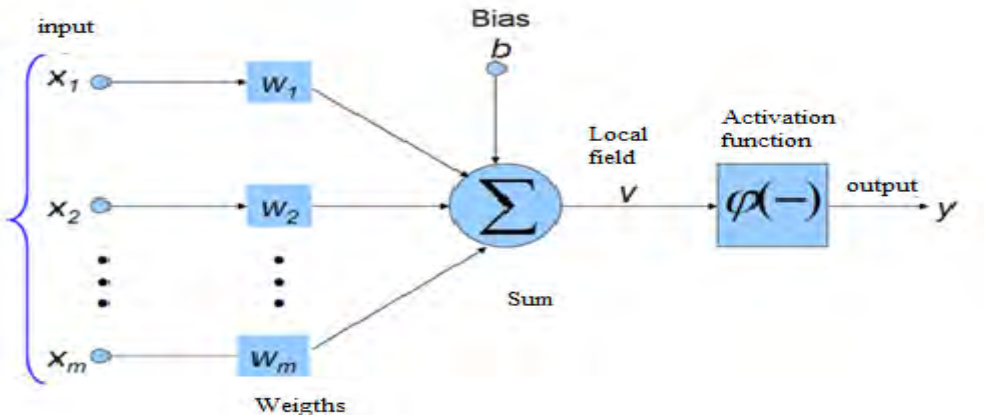

Figure 2. Model of Artificial nerve cells (Boukadıda, Hassen, Gafsı and Besbes, 2011)

In the learning process, the brain shows continuous development and change. As indivuals live and experience, synaptic connections are created and new connections are established. That is how learning occurs. The same is also true for the ANN. Like learning happens through experience, in the ANN process, with processing input / output data to make the training algorithm repeatedly until a link is established between the link weights using this data (Bahadır, 2013). As a statistical methodology the ANN has been actively utilized in many fields including business, biology, medicine, environmental research and terrorist attack predictions. In recent years, it has been used as a statistical method in social sciences to classify and define the components of data. However, although studies utilizing this statistical method have gained attention in social science, there is a limited study that use the ANN in measurement and evaluation in educational studies. These studies have used the ANN to compare different ways of prediction student achievements (Bahadır, 2016; Çırak & Çokluk, 2013; Gorr, Nagin & Szczypula 1994; Özçınar, 2006; Toprak, 2017), to predict student failures in advance (Güneri & Apaydın, 2004), to regress and classify student academic achievement (Campbell Hunt, 2000; Özdemir, 2015; Tepehan, 2011) and to determine the variables that affect achivement (Oladokun, Sc, Adebanjo & Charles-Owaba, 2008; Özdemir, 2015; Turhan, Kurt & Engin, 2013; Tezbaşaran, 2016; Aydoğan, 2018).

In addition, in literature studies exist that the ANN is compared with other statistical methods. Some studies compared the ANN with data mining (Ayık, Özdemir & Yavuz 2007; Üçgün, 2009; Şengür, 2013), linear regression (Sittirug, 1997), progressive regression (Gorr, Nagin & Szczypula, 1994), logistic regression (Bahadır, 2013; Campbell Hunt, 2000; Güneri & Apaydın, 2004) and decision tree (Altaş & Gülpınar, 2012; Ersöz, Özseven & Ersöz, 2017). In literature, it has suggested that statistical methods can be compared with the ANN as well the ANN estimates obtained in different software. In this study the same data were analyzed by utilizing the ANN approach in difference statistical softwares as MATLAB and SPSS. Then, results of the predictions about academic achievement were evaluated. Lastly, the strength and direction of relationship among variable were compared with each other.

As highlighted in the literature some studies have utilized the ANN as research methods in educational research; however, studies examining the variables related to students' academic achievement at the primary school level are quite limited. In addition, studies have mostly compared academic achievement prediction by utilizing Logistic Regression Analysis (LRA) or decision tree methods with the ANN. Comparing the prediction of the ANN on academic achievement run in MATLAB and SPSS could be beneficial to determine which one could make stronger predictions. In this view, our study addresses the gap in literature comparing two different powerful analysis softwares by uisng the same method.

1.2. Need for Study

In the literature many methods and techniques were used to predict students' academic achievement enrolled in different educational institutions across the world. However, we have not come across any study that compares the academic achievement predictions of primary school students with the utilization of two different programs by using ANN analysis, which is the most reliable prediction mechanism of the information age and that works like the human brain. Using the ANN analysis, the main objective of this research was to compare parameter of variables affecting academic achievement in MATLAB and SPSS.

1.3. Hypothesizes

Hypothesis 1. The data collection tool developed by researchers are valid and reliable.

In literature, "father's education status", "mother's education status" and "father's employment status" (Anıl, 2009; Cameron & Heckman, 2001; Levpušček, Zupancic & Socan, 2012; Ural & Çınar, 2014), "having the internet connection", "having a computer", "having a study room", "having a study desk", "having a library" (TIMMS, 2011) were found to be significantly related to student academic achievement. The first hypothesis of the study is test that data collection tool that included the aforementioned variable are valid and reliable.

Hypothesis 2. The model using the ANN in MATLAB software has classified students' achievements to a high degree.

In the literature, studies utilizing the ANN models in MATLAB software have reported to classify students' academic achievement to a high degree (Atasayar, 2019; Bahadır, 2013; Özkan, 2019). In this view, it is expected that the second hypothesis of this research is to test whether the model classifies student achievement to a high degree.

Hypothesis 3. The model constructed by ANN in SPSS classifies students' achievements to a high degree.

Studies have reported that the ANN model run in the SPSS software highly accurately classifies the academic achievement prediction (Lye et al. 2010; Luft, Gomes, Priori & Takase, 2013; Musso, Kyndt, Cascallar & Dochy, 2013; Tepehan, 2011; Turhan, Kurt & Engin, 2013; Wongkhamdi & Serensangtakul, 2010). The third hypothesis is to test the ANN model run in SPSS classifies students' achievements to a high degree.

Hypothesis 4. The magnitudes of the parameters obtained in MATLAB and SPSS are close to each other.

In the literature we haven't located any study comparing the parameters obtained in the MATLAB and SPSS softwares. However, some studies used these two softwares simultaneously (Bahadır, 2016; Demir, 2015; Yorgancı and Işık, 2019) and others compared the parameters obtained in SPSS and WECA softwares (Altaş and Gülpınar, 2012; Cortez & Silva, 2008; Özbay, 2015). The same token, it is hypothesized that the magnitudes of the parameters obtained in MATLAB and SPSS softwares are close to each other.

2. Methodology

2.1. Research model and participants

In this study a quantitative survey and correlational research model was used. Correlational research models are research approaches to examine the relation between variables whereas in survey models, the event, individual or object that is the subject of the research is tried to be defined within its own conditions and as it is (Karasar, 2010). A total of 465 fourth grade students who enrolled in primary schools located in Central Anatolia in Turkey participated in

this study in 2017. The sample in this location reflects the cultural and racial diversity of Turkey. The demographic information of the participants was displayed in Table 1.

Table 1. *Demographic information about participants*

Independent variables	Group	f (frequency)	% (percent)
Gender	Female	248	%53.3
	Male	217	% 46.7
Academic achievement status at end of third grade	Successful	459	% 98.7
	Unsuccessful	6	% 1.3

As seen in Table 1, approximately 53.3% of the participants in the study were girls and 46.7% were boys. Additionally, nearly 98.7% were successful at the end of third grade and 1.3% were unsuccessful. At the third grade students in Turkey were graded into three types of achievement status as “needs improvement", “good" and "very good". Within the scope of the research, those who were graded good and very good were put into the successful group and those who were evaluated as needs improvement were into the unsuccessful group.

2.3. Instrument

An achievement test developed by researchers were used as the data collection tool. In preparing the achievement test, first, a question pool was created with 35 questions created by the researchers. The number of questions was reduced to 23 items after consulting experts in a science, two mathematics, a measurement and evaluation, and two primary teaching fields. These 12 questions were removed either because they did not meet the language and measurement criteria. By doing so, the content validity of the questionnaire was met. An item factor analysis was performed for the construct validity of the 23 items. In order to analyze the factor, firstly KMO value was calculated and Barlett test was performed. The KMO value of the study was .62, Bartlett's test value was .00. Additionally, the KR20 value was .60.

2.4. Data analysis

After the validity and reliability of the data collection tool were established, data were analyzed with in MATLAB R2016a and SPSS 24.0 software. In the study, the third grade end-of-year achievement status of 4th grade students were determined as the output variable for ANN. A two-category discontinuous variable was created as "unsuccessful” (0) and "successful" (1). The aforementioned variables were considered as input. The ANN models tested in MATLAB R2016a and SPSS 24.0 software were performed to predict the students' academic achievement. Parameters found in analyses of the ANN in MATLAB and SPSS software were compared based on their regression coefficient.

3. Findings

In this part of the study, findings related to the hypotheses were presented.

3.1. Findings related to the Hypothesis 1.

To validate the instrument, an exploratory factor analysis was run. In order to run factor analysis of the instrument consisting of 23 items, it is suggested that KMO value should be higher than .60 cutoff value (Pallant, 2007). Addition to this, cutoff value of factor loading should be higher than .30. Therefore, 11 items not meeting this criterion were deleted. Varimax rotation was utilized. A new exploratory factor analysis resulted with three factors of twelve items. The factor distributions and loadings were presented in Table 2.

Table 2. *Factor loadings of the items based varimax rotation*

Items	Factor loadings		
	F1	F2	F3
Q3. Having a study room	.359		
Q4. Having a study desk	.630		
Q5. Having a personal computer	.727		
Q6. Having the Internet connection at home	.639		
Q7. Having a library	.669		
Q8. Developing self-study habit			.432
S12 In-class activities and experimenting			.739
S15.Teachers teach with different methods in class			.686
S16. Using technological tools in the classroom			.539
S17.Mother educational status		.869	
S18.Father educational status		.840	
S20. Father working status		.455	

The factor loading distributions of items ranged between .359 and .869. In the literature it is suggested that the factor loadings value should be over .30 and the difference between the two high loading values should be at least .10 (Çokluk, Şekercioğlu & Büyüköztürk, 2010).

The first factor in exploratory factor analysis included a total of 5 items and its factor loading values varied between .35 and .72. This factor explained 18.987% of the total variance. The second factor emerged from items consisted of 4 items and the factor loading values varied between .43 and .73. The second factor explained 13.562% of the total variance. The third factor included 3 items and factor loading values varied between .45 and .86. This factor explained 11.768% of the total variance. The three-factor structure explained 44.317% of the total variance. Considering items and experts' views, the factors were named as "student features", "teacher-school features" and "parental features", respectively. The KR20 reliability of the questionnaire, which consists of 3 sub-dimensions and 12 items, was recomputed and found to be .89. These findings indicated that the data collection tool was valid and reliable.

3.2. Findings related to the second hypothesis

In the analysis in MATLAB software, Levenberg-Marquardt (trainLM) algorithm was used in the training of feed forward-back propagation network. A three-layer feed forward-network as the input, hidden and output layers was used. In the hidden layer the sigmoid and in the output layer the linear activation function was used. In the academic achievement prediction model, there were 12 inputs in the input layer, 15 neurons and 1 output in the hidden layer (Figure 1).

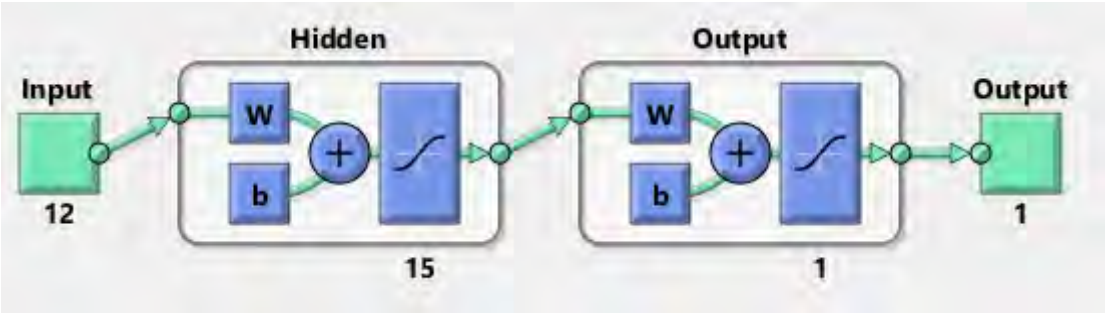

Figure 3. The layer, input and output layers of the model in MATLAB

Of the data set, 70% (325) in the training, 15% (70) in the verification and 15% (70) in the test were used. The ratios of the training, the verification and the test for modeling were showed in Table 3.

Table 3. *The classification based on results of the ANN model in MATLAB*

ANN steps	Observed	Predicted		Correct prediction percentage (%)
		Successful	Unsuccessful	
Training	Successful	274	43	86.4
	Unsuccessful	3	5	62.5
	Total	%98.9	%10.4	85.8
Verification	Successful	57	13	81.4
	Unsuccessful	0	0	0
	Total	%100	%0	98.6
Test	Successful	56	12	82.4
	Unsuccessful	1	1	50
	Total	%98.2	%7.7	81.4

As seen in Table 3, 317 out of 325 individuals were classified as successful and 8 as unsuccessful, and the correct classification rate was 86.4%. Of the 8 students who failed, 5 were correctly classified and 3 were incorrectly, and the correct classification rate was 62.5%. Of the 70 individuals who participated in the verification analysis, 57 successful students were correctly classified and 13 students were incorrectly. The correct classification rate of the verification analysis was 81.4%. Of the 70 individuals in the test step, 56 of 68 successful students were correctly classified and 12 were incorrectly. The correct classification rate was 82.4%. Of the 2 students who failed, 1 was incorrectly classified and 1 was correctly, and the correct classification rate was 50%. The total correct classification rate was computed as 81.4%.

Table 4. *The overall results of ANN analysis in MATLAB*

Observed	Predicted	Correct prediction percentage (%)
Successful	393	85.1
Unsuccessful	72	60
Total	465	84.5

As seen in Table 4, the total correct estimation percentage for the overall study was 84.5%.

3.3. Findings related to the third hypothesis

In analysis of the ANN model in the SPSS 24.0, Multilayer Sensor (MLR- Multiplayer Perceptron) was used. Whereas the "Hyperbolic Tangent Function" was used as the activation function of the artificial nerve cells in the input layer, the "Softmax Function" was used in the output layer. All variables were equalized to make sure that the model constructed was the same as MATLAB. There were 12 inputs in the input layer of the model, 15 neurons and 1 output in the hidden layer.

70% (325) of the data set were used in the training, 15% (70) in the verification and 15% (70) in the test. The training, verification and testing rates for modeling were displayed in Table 5.

Table 5. *The classification based on results of the ANN model in SPSS*

ANN steps	Observed	Predicted		Correct prediction percentage (%)
		Successful	Unsuccessful	
Training	Successful	265	56	82.5
	Unsuccessful	4	0	0
	Total	%98.5	%0	84.9
Verification	Successful	53	2	96.3
	Unsuccessful	15	0	0.0
	Total	%77.9	%0	75.7
Test	Successful	61	4	93.8
	Unsuccessful	5	0	0
	Total	%92.4	%0	87.1

As seen in Table 5, in the academic achievement prediction, there were a total of 321 students of 265 students who were successful. The correct classification rate was 82.5%. ,And 56 were incorrectly. Of the 4 students who were unsuccessful, all of them were misclassified. The correct classification rate was 0%. The total correct classification rate was 84.9%. There were a total of 70 students of 53 students who were successful. Of the 70 individuals included in the test, 53 of 55 successful students were correctly classified and 2 were incorrectly. The correct classification rate was 96.3%. Of the 15 students who were unsuccessful, all of them were misclassified. 15 students who were unsuccessful were incorrectly classified and the correct classification rate was 0%. The total correct classification rate was 98.6%. Of the 70 individuals included in the test, 61 of 6 successful students were correctly classified and 4 were incorrectly. The correct classification rate was 93.8%. Five students who were unsuccessful were classified incorrectly and the correct classification rate was 0%. Total correct classification rate was 87.1%.

Table 6. *The overall ANN analysis results in SPSS*

Observed	Predicted	Correct prediction percentage (%)
Successful	379	85.9
Unsuccessful	86	27.9
Total	465	81.5

As seen in Table 6, the overall correct prediction percentage of the ANN model was 81.5%.

3. 4. Findings related to the fourth hypothesis

The regression coefficient of a predictor or input variable shows the contribution of each of the variables on the prediction of the dependent variable. In order to quantify the relative importance of independent variables in predicting the output variable for the neural network, the proposed method for the MATLAB are a brief description of the connection weights algorithm (Olden & Jackson, 2002). Connection weights algorithm was used for relative

importance of independent variables. The relative importance of a given input variable can be defined as:

$RI_x = \sum_{y=1}^{m} Wxy\ Wyz$

RI_x is the relative importance of input neuron x and $\sum_{y=1}^{m} Wxy\ Wyz$ is the sum of product of final weights of the connection from input neuron to hidden neurons with connection from hidden neurons to output neuron. y is the total number of hidden neurons, and z is output neurons. This approach is based on estimates of network final weights obtained by training the network (See Table 7).

Table 7. *Final connection weights*

Inputs	hidden1	hidden2	hidden3	hidden4	hidden5	hidden6	hidden7	hidden8	hidden9	hidden10	hidden11	hidden12	hidden13	hidden14	hidden15
W.room	0.26	0.36	0.27	0.11	0.45	0.11	0.71	0.76	0.84	0.12	1.05	-0.27	-0.07	-0.61	0.01
W.table	0.98	0.75	1.42	1.25	0.21	0.35	0.88	1.06	0.27	0.29	0.86	-0.66	-0.34	-1.31	-0.82
Comptr	0.27	0.62	0.90	0.45	0.15	0.46	0.07	0.88	0.36	0.14	0.03	-1.24	-0.60	-0.30	-1.13
Internet	0.75	0.13	0.50	1.49	0.51	0.16	1.50	0.20	1.22	0.99	0.59	-0.46	-1.51	-0.64	-0.63
Bookcse	0.20	0.05	0.36	0.33	0.17	0.50	0.50	0.39	0.48	0.81	0.13	-0.51	-0.27	-0.68	-0.78
R.work	0.55	0.51	0.49	1.50	0.61	1.06	1.27	0.08	0.97	0.35	0.48	-0.83	-0.01	-1.0	-0.36
Cl.activt	0.93	0.38	0.33	0.57	0.50	0.59	1.27	0.82	1.04	1.15	0.10	-0.77	-0.48	-1.10	-0.11
Df.meth	1.10	0.71	1.06	0.86	0.28	0.92	0.31	0.39	0.02	0.43	0.43	-0.44	-0.06	-1.03	-0.14
Technol	0.28	0.13	0.55	0.74	1.22	1.32	0.36	0.18	0.83	0.94	0.02	-1.04	-0.16	-0.11	-0.14
Moth.lit	1.22	1.23	0.49	0.55	0.09	1.19	0.27	0.86	0.15	0.84	1.10	-0.76	-0.16	-0.35	-0.18
Fath.litr	0.04	0.03	0.79	1.08	1.20	0.90	0.18	0.14	0.66	-0.54	0.36	-0.01	-0.43	-0.61	-0.42
Fa.work	0.80	0.82	0.35	0.05	0.70	1.26	2.25	0.46	1.37	-0.61	1.12	0.013	-0.23	-0.42	-0.39
GY	1.08	0.72	0.46	0.94	0.84	0.69	0.95	0.63	0.86	0.87	0.05	-0.86	-1.10	-1.26	-1.32

Table 8. *Connection weights products, relative importance and rank of inputs.*

Inputs	hidden 1	hidden 2	hidden 3	hidden 4	hidden 5	hidden 6	hidden 7	hidden 8	hidden 9	hidden 10	hidden 11	hidden 12	hidden 13	hidden 14	hidden 15	Sum	R.Imp %	Rank
W.room	0.28	0.26	0.12	0.10	0.38	0.07	0.68	0.48	0.73	0.11	0.0	0.24	0.07	0.78	0.02	4.44	4.89	12
W.table	1.06	0.54	0.66	1.18	0.17	0.24	0.84	0.67	0.24	0.25	0.05	0.57	0.38	1.66	1.08	9.67	10.64	2
Comptr	0.29	0.45	0.42	0.43	0.12	0.32	0.06	0.56	0.31	0.13	0.00	1.07	0.66	0.38	1.49	6.76	7.44	8
Internet	0.82	0.09	0.23	1.40	0.43	0.11	1.44	0.13	1.05	0.87	0.03	0.39	1.67	0.81	0.83	10.37	11.41	1
Bookcse	0.21	0.040	0.17	0.31	0.15	0.35	0.47	0.24	0.42	0.70	0.00	0.44	0.30	0.86	1.03	5.76	6.34	11
R.work	0.59	0.37	0.23	1.41	0.51	0.74	1.21	0.05	0.84	0.31	0.0	0.72	0.01	1.28	0.48	8.82	9.71	4
Cl.activt	1.01	0.27	0.15	0.53	0.43	0.41	1.22	0.51	0.90	1.00	0.00	0.67	0.53	1.39	0.14	9.24	10.27	3
Df.meth	1.19	0.51	0.49	0.81	0.24	0.64	0.29	0.25	0.02	0.37	0.02	0.33	0.07	1.30	0.18	6.83	7.52	7
Technol	0.30	0.09	0.26	0.69	1.03	0.92	0.34	0.11	0.72	0.82	0.00	0.90	0.18	0.15	0.18	6.76	7.44	9
Moth.lit	1.32	0.89	0.22	0.52	0.07	0.83	0.2	0.54	0.13	0.73	0.06	0.66	0.18	0.44	0.24	7.17	7.89	6
Fath.litr	0.04	0.02	0.37	1.01	1.01	0.63	0.17	0.09	0.57	0.47	0.02	0.01	0.48	0.77	0.56	6.28	6.91	10
Fa.work	0.87	0.59	0.1	0.05	0.59	0.87	2.15	0.29	1.19	0.53	0.06	0.01	0.26	0.54	0.52	8.74	9.63	5
Sum	8.02	4.19	3.54	8.50	5.20	6.18	9.21	3.96	7.15	6.34	0.37	6.09	4.85	10.4	6.81	90.90		

As seen in Table 8, the order of the predictors influencing academic achievement to the most important to the least important; having internet connection at home (11.41%), having a study desk (10.64%), in-class activities and experimenting (10.27%), having self-study habit (9.71%), father's working status (9.63%), mother's education (7.89%) , Teachers teach with different methods in class (7.52%), having a personal computer (7.44%), the teacher teaches in the classroom with technological tools (7.44%), father's educational status (6.91%), having a library (6.34%) and having a study room (4.89%).

Since the SPSS automatically put the predictors in order of their importance, no formula was used. The result of the analysis in SPSS was displayed in Table 9.

Table 9. *Importance of predictors in the ANN analysis in SPSS*

Order of importance	Factors	Importance	Normalized importance percentage (%)
1	Having the internet connection at home	.15	100.0
2	Having a study desk	.14	94.5
3	Father educational status	.12	84.3
4	In-class activities and experimenting	.11	76.9
5	Developing self-study habit	.11	73.7
6	Using technological tools in the classroom	.09	64.9
7	Teachers teach with different methods in class	.09	61.0
8	Having a study room	.059	39.1
9	Mother educational status	.068	47.8
10	Having a personal computer	.032	21.5
11	Father working status	.030	20.0
12	Having a library	.023	15.3

The inputs / factors affecting to output/academic achievement were listed in the order of the importance in the ANN analysis in SPSS in Table 9. Results indicated that among the normalized inputs significant at the level of .01, the order from the most influential to the least was having internet connection at home (100%), having a study desk (94.5%), father's working status (84.3%), in-class activities and experimenting (76.9%), developing self-study habit (73.7), using technological tools in the classroom (64.9%), teachers teach with different methods in class (61.0%), having a study room (39.1%), mother's educational status (47.8%), having a personal computer (21.5%), father's educational status (20%) and having a library (15.3%).

The comparison of the orders of importance of the predictors found in both softwares was displayed Table 10.

Table 10. *Comparison of MATLAB and SPSS*

Importance	MATLAB ANN	SPSS ANN
1	Having the internet connection at home	Having the internet connection at home
2	Having a study desk	Having a study desk
3	In-class activities and experimenting	Father working status
4	Developing self-study habit	In-class activities and experimenting
5	Father working status	Developing self-study habit
6	Mother educational status	Using technological tools in the classroom
7	Teachers teach with different methods in class	Teachers teach with different methods in class
8	Having a personal computer	Having a study room
9	Using technological tools in the classroom	Mother educational status
10	Father educational status	Having a personal computer
11	Having a library	Father educational status
12	Having a study room	Having a library

4. Discussion

In this study, it was aimed to compare the predictions of the ANN in MATLAB and SPSS software and to determine the factors affecting academic achievement of primary school students. Participants were from schools located in Central Anatolia in Turkey. In the ANN model Multilayer Sensor Model was used in both softwares. 98% of the participants were successful at the previous academic year whereas %2 were not. The reason for the low rate of unsuccessful students is that, in the Turkish education system, the evaluation in primary school is oriented to promote to students to the upper class rather than grade repetition (Legal Gazette, 2014). However, some students with a really weak academic status can be considered unsuccessful and take a grade repetition with the consent of the parent.

In the light of the findings, the following conclusions were reached.

1. The measurement tool developed in determining the variables affecting the academic achievement set out in the first hypothesis of the research was confirmed as valid and reliable. Once the scope validity of the instrument was established by experts' opinions, the KMO value of the instruments was .62. Bartlett's test value of the instrument was .00. KR20 value was .89. 11 items were removed based on the result of factor analysis showed as their factor loadings were lower than .30 cutoff value. Remained 12 item explained 44.317% of the total variance. All these results indicated that the instrument was valid and reliable.

2. The second research hypothesis was related to whether the ANN model in MATLAB predicted the academic achievement with a high accuracy. The results showed that the model estimated the achievement accurately at high degree. The accuracy rate of the ANN model in MATLAB for the successful students was 85.1%, while the rate was 60.0 was for the

unsuccessful students. The overall accuracy rate for all students were 84.5%. The accuracy rate of the ANN model was higher for successful students than did for unsuccessful students. These findings were consistent with the previous studies that the ANN model predicted the achievement with high accuracy (Atasayar, 2019; Özkan, 2019; Yorgancı and Işık, 2019).

3. The third hypothesis was related to the ANN model tested in the SPSS with the multi-layered sensor estimated academic achievement of students with a high accuracy. It was found that the ANN model predicted the academic achievement status of 85.9% of successful students accurately. However, the model correctly predicted the academic achievement of 27.9% of unsuccessful students. Overall the correct estimating rate of ANN model estimation for academic achievement was 81.5%. Again, the rate of accurate estimation was higher for successful students than did for unsuccessful students. the findings were in parallel line with the previous studies in the literature (Çırak 2012; Luft, Gomes, Priori & Takase, 2013; Lye, et al., 2010; Musso, Kyndt, Cascallar & Dochy, 2013; Tepehan, 2011; Turhan, Kurt & Engin 2013; WongKhamdi & Serensangtakul, 2010). However, it was found that the accuracy rate of the ANN model was higher in MATLAB than did in SPSS. Especially, the correct estimation rate of SPSS for unsuccessful students was not at expected level; thus, it was more plausible to reject the third hypothesis.

4. The fourth hypothesis was found to be accepted. The results of the ANN model in MATLAB and SPSS showed that that the accuracy rate of the ANN model in MATLAB was higher than did in SPSS. Considering the findings, it can be suggested that the ANN model in MATLAB could compute a highly accurate prediction of students' academic achievement in the coming years than did in SPSS. These findings were consistent with the previous studies (Demir, 2015; Yorgancı & Işık, 2019). Therefore, because of higher accuracy rate at estimation, the ANN model in MATLAB was chosen to evaluate the importance of independent variables. However, the importance order of the variables affecting academic achievement in MATLAB and SPSS were quite close to each other. In both ANN models revealed that having internet connection at home and having a study desk were the strongest predictors of the academic achievement. These findings were consistent with the previous studies (Alamdar, 2015; Erbaş, 2005; TIMSS, 2011). The ANN model in MATLAB indicated that In-class activities and experimenting, using technological tools in the classroom, and teachers teach with different methods in class were found to be other most significant variables affecting the academic achievement. This result is consistent with the previous studies (Çalışkan, 2008; Şaşmazel, 2006). In addition, mother's educational status, father's educational status and father's employment status were significantly related to academic achievement. Some other previous studies have highlighted that these factors were significantly related to the achievement (Anıl, 2009; Ekmekyermezoğlu, 2010; Gelbal, 2008; Levpušcek, Zupancic & Socan, 2012). The high accuracy rate of the ANN model in MATLAB suggested that this analysis method and software can be an alternative method in the prediction studies in the field of education. Also, the findings suggest that the ANN model in MATLAB is a powerful tool that can be used in determining the factors affecting student academic achievement (Yagci & Çevik, 2019).

5. Implications and limitation

Although in this study the factors affecting academic achievement were measured with a valid and reliable measurement tool, these parameters are context-dependent which means that they change in the contexts of different countries, cultures, races or socioeconomic conditions. Therefore, considering the contexts of the study, reconstructing the factors affecting academic

achievement is very important. In this study, within the scope of the research, the ANN models in MATLAB and SPSS were chosen to predict academic achievement and their predictive powers were tested. In future studies, ANN methods in different statistical software may be preferred or comparisons can be made. In this study, primary school students' academic achievement was predicted. Therefore, the future research can be carried out at a wider range, at different grade levels and at different school types. In this study, it was found that factors including having internet at home and having a study desk were a powerful predictor of academic achievement; thus, it is necessary for every student or their parents to have internet at home and a study desk in order to increase their academic achievement at primary schools. Additionally, it can also be suggested that they should have a suitable study space of their own to create self-study atmosphere for their studies. Similarly, it was found that in-class activities and experimenting were found to be a predictor of academic achievement. Addition to this, self-study habit was another factor related to academic achievement. Therefore, these findings suggest that primary school teachers should help their students foster self-study habit and use different hands-on activities to improve their achievement. Results highlight importance of the usage of technological tool in classroom, that makes clear that primary teachers should promote to use technology in their classroom.

6. Conflict of Interest

The authors declare that there is no conflict of interest.

7. Ethics Committee Approval

The authors confirm that the study does not need ethics committee approval according to the research integrity rules in their country.

Endnote

This research involves a part of the project funded by Karamanoğlu Mehmetbey University Scientific Research Projects Commission (grant number: 26-M-16) under Scientific Research Projects.

References

Alamdar, S. (2015). *Analysing the elements affecting academic success in the context of efficient school concept.* Unpublished master thesis. Ankara Üniversity, Ankara.

Altaş, D., & Gülpınar, V. (2012). A comparison of classification performances of the decision trees and the artificial neural networks: European Union. Journal of *Trakya University Social Sciences, 14*(1), 1-22.

Anıl, D. (2009). Factors effecting science achievement of science students in Programme for International Students' Achievement (PISA) in Turkey. *Education ve Science*, *34*(152), 15-34.

Atasayar, A. (2019). Predicting primary school LGS achievement with artificial neural networks. Unpublished master thesis. Kutahya Dumlupinar University, Kutahya.

Aydoğan, İ. (2018). Estimation of student successes by artificial neural networks. *YYU Journal of Education Faculty, 15*(1), 577-610.

Ayık, Y. Z., Özdemir, A., & Yavuz, U. (2007). Lise türü ve mezuniyet başarısının kazanılan fakülte ile ilişkisinin veri madenciliği tekniği ile analizi. *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi. 10*(2), 441-454.

Bahadır, E. (2013). *Prediction of student teachers' academic success with logistic regression analysis and artificial neural networks methods.* Unpublished Doktoral dissertation, Marmara University. İstanbul.

Bahadır, E. (2016). Using neural network and logistic regression analysis to predict prospective mathematics teachers' academic success upon entering graduate education. *Educational Sciences: Theory & Practice, 16*(3). 943-964.

Bekele, R., & McPherson, M. (2011). A Bayesian performance prediction model for mathematics education: A prototypical approach for effective group composition. *British Journal of Educational Technology, 42*(3), 395–416.

Boukadıda, H., Hassen N., Gafsı, Z., & Besbes, K. (2011). A highly time-efficient digital multiplier based on the A2 binary representation. *International Journal of Engineering Science and Tecnology, 3*(5), 4498-4509.

Budak, H. & Erpolat, S. (2012). Comparison of artificial neural networks and logistic regression analysis in the credit risk prediction. *Online Academic Journal of Information Technology,* 3(9), 23-30.

Büyüköztürk, Ş. (2009). *Data analysis handbook for Social Sciences.* Ankara: Pegem Akademi.

Cameron, S., & Heckman, J. (2001). The dynamics of educational attainment for black. hispanic. and white males. *Journal of Political Economy, 109*(3), 455-499.

Campbell Hunt, C. (2000). What have we learned about generıccompetıtıve strategy? A meta-analysıs. *Strategic Management Journal,* 21, 127-154.

Colom, R., Escorial, S., Chun Shih, P., & Privado, J. (2007). Fluid intelligence, memory span, and temperament difficulties predict academic performance of young adolescents. *Personality and Individual Differences, 42,* 1503-1514.

Cortez, P., & Silva, A. M. G. (2008*). Using data mining to predict secondary school student performance.* In A. Brito, & J. Teixeira (Eds.), Proceedings of 5th Annual Future Business Technology Conference, Porto, 5-12.

Çalışkan, M. (2008). *The impact of school and student related factors on scientific literacy skills in the programme for international student assessment-PISA 2006.* Unpublished Doktoral Dissertation, Middle East Technology University, Ankara.

Çırak, G. (2012). *The usage of artifical neural network and logistic regression methods in the classification of student achievement at higher education.* Unpublished master thesis, Ankara University, Ankara.

Çırak, G., & Çokluk, Ö. (2013). The usage of artifical neural network and logistic regression methods in the classification of student achievement at higher education. *Mediterranean Journal of Humanities*, *3*(2), 71-79.

Çokluk, O., Şekercioğlu, G., & Büyüköztürk, Ş. (2010). *Multivariate analysis in SPSS and Lisrel for sociel science.* Ankara: Pegem Akademi.

Demir, M. (2015). Predicting pre-service clssroom teachers 'civil servant recruinment examinations educational sciences test scores using srtifical neutral networks. *Educational Sciences: Theory and Practice*, *15*(5), 1169-1177.

Demirel, Ö. & Kaya, Z. (2015). Introduction of education. Ankara: Pegem Akademi.

Ekmekyermezoğlu, N. (2010). *Socio-cultural factors effecting academic successes of 8th grade students of primary education at English class*. Unpublished master thesis, Inonu University, Malatya.

Erbaş, Ü. (2005). Artifical neural network and implications. Unpublished master thesis, Marmara University. İstanbul.

Ersöz, T., Özseven, T., & Ersöz, F. (2017). Modelling the costumers' cell-phone preferences with decision tree. *Gümüşhane Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 8*(19), 128-136.

Fenollar, P., Roman, S., & Cuestas, P. J. (2007). University students. academic performance: An integrative conceptual framework and empirical analysis. *British Journal of Educational Psychology*, *77*, 873- 891.

Gelbal, S. (2008). The effect of socio-economic status of eighth grade students on their achievement in Turkish. *Education and Science, 33*, 150-163.

Gorr, W. L., Nagin, D., & Szczypula, J. (1994). Comparative study of Artificial Neural network and Statistical models for predicting student grade point averages. *Interntional Journal of Forecasting, 10*(2), 17-34.

Güneri, N., & Apaydın, A. (2004). Logistic regression and artificial neural network in classification of students' achievement. *Ticaret ve Turizm Eğitim Fakültesi Dergisi, 1,* 170-188.

Hailikari, T., Nevgi, & A., Komulainen, E. (2008). Academic self-beliefs and prior knowledge as predictors of student achievement in Mathematics: a structural model. *Educational Psychology, 28*(1), 59-71.

Hunt, L. D. (2000). *Comparison of neural network and logistic regression models for predicting the academic success of college freshmen.* Unpublished dissertation, Raleigh, NC: North Carolina State University.

Karasar. N. (2010). *Scientific research methods.* Ankara: Nobel Akademi Yayınları.

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology, 86*(1), 148-161.

Krumm, S., Ziegler, M. & Buehner, M. (2008). Reasoning and working memory as predictors of school grades. *Learning and Individual Differences, 18*(2), 248-257.

Kyndt, E., Musso, M., Cascallar, E., & Dochy, F. (2015), Predicting academic performance: the role of cognition, motivation and learning approaches: a neural network analysis, *Methodological Challenges in Research on Student Learning, 1,* 55-76.

Levpušcek, M. P., Zupancic, M., & Socan, G. (2012). Predicting achievement in mathematics in adolescent students: The role of individual and social factors. *The Journal of Early Adolescence*, *20*(10), 1–29.

Luft, C. B., Gomes, J. S., Priori D., & Takase E. (2013). Using online cognitive tasks to predict mathematics low of school achievement. *Computers & Education, 67*, 219-228.

Lye, C. T., Ng, L. N., Hassan, M. D., Goh, W. W., Law, C. Y., & Ismail, N. (2010). Predicting pre-university students' mathematics achievement. In *Procedia - Social and Behavioral Sciences 8*, 299–306). Elsevier Ltd. https://doi.org/10.1016/j.sbspro.2010.12.041.

Mukta, P., & Usha, A. (2009). A study of academic performance of business school graduates using neural network and statistical techniques. *Expert Systems with Applications*, *36*(4), 7865-7872.

Musso, M., Kyndt, E., Cascallar, E., & Dochy, F. (2013). Predicting general academic performance and identifying the differential contribution of participating variables using artificial neural networks. *Frontline Learning Research*, *1,* 42 – 71.

Musso, M. F., & Cascallar, E. C. (2009). New approaches for improved quality in educational assessments: Using automated predictive systems in reading and mathematics. *Journal of Problems of Education in the 21st Century, 17*, 134-151.

Oladokun, V., Adebanjo, A., Sc, B., & Charles-Owaba, O. (2008). Predicting students' academic performance using artificial neural network : a case study of an engineering course. *The Pacific Journal of Science and Technology, 9*(1), 72-79.

Olden, J. D. Joy, M. K., & Death, R. G. (2004). An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data, *Ecological Modeling, 178,* 389-397.

Özbay, Ö. (2015). *Analysis of student use of learning management system through data mining methods in undergraduate level.* Unpublish master thesis. Baskent University, Ankara.

Özçınar, H. (2006). *Predicting KPSS results using data mining methods.* Unpublish master thesis. Pamukkale University, Denizli.

Özdemir, İ. (2015). *Acting on science and technology teachers' professional burnout variables forecasting with artificial neural network (the sampling of zonguldak)* Unpublished doctoral dissertation. Ahi Evran University. Kırşehir.

Özkan, İ. (2019). *Analysis of the relationship between self-efficacy perception and academic achievement in science and technology literacy with artificial neural networks.* Unpublish master thesis. Burdur Mehmet Akif Ersoy University, Burdur.

Öztemel, E. (2003). *Artifical neural networks*. İstanbul: Papatya Yayınları.

Pallant, J. (2007). *SPSS survival manual:a step-by-step guide to data analysis using SPSS for Windows*. Philadelphia. PA:Open University Press.

Legal Gazette, (2014). Ministry of National Education primary level regulations. Retrived from http://mevzuat.meb.gov.tr/dosyalar/1703.pdf.

Şaşmazel, A. G. (2006). *Factors that affecting success of scientific literacy on students in Turkey that participate programme for international student assessment (PISA)*. Unpublish master thesis, Hacettepe Univeristy, Ankara.

Sittirug, H. (1997). *The predictive value of science process skills, cognitive development, attitude toward science on academic achievement in a Thai teacher institution*. Doctora Thesis, University of Missouri, Columbia.

Şengür, D. (2013). *Prediction of student's academic achievements by using the data mining methods.* Unpublish master thesis. Fırat University, Elazığ.

Tepehan, T. (2011). *Performance comparison of artificial neural network and logistic regression model in predicting Turkish students' PISA success.* Unpublish doctoral dissertation. Hacettepe Univeristy, Ankara

Tezbaşaran, E. (2016). *An Investigation on usability of principal component analysis and artificial neural network models in the process of scale development.* Unpublish doctoral dissertation. Mersin University, Mersin.

TIMSS (2011). Uluslararası matematik ve fen eğilimleri araştırması ulusal raporu. Ankara: MEB: Eğitimi Araştırma ve Geliştirme Dairesi Başkanlığı.

Toprak, E. (2017). *Comparison of classification performances of mathematics achievement at PISA 2012 with artificial neural network, decision trees and discriminant analysis.* Unpublish doctoral dissertation. Hacettepe Univeristy, Ankara.

Towle, A., & Bradley, J. V. (2006). *Modern biology*. Toronto: Holt, Rinehart and Wilson.

Turhan, K., Kurt, B., & Engin, Y. Z. (2013). Estimation of student success with artificial neural networks. *Education and Science, 170*(38), 112-120.

Turner, E. A., Chandler, M., & Heffer, R. W. (2009). Influence of parenting styles, achievement motivation, and self-efficacy on academic performance in college students. *Journal of College Student Development*, 50(3), 337-346.

Ural, A., & Çınar, F. N. (2014). The effect of parents' level of education on students' mathematics achievement. *Mehmet Akif Ersoy University Journal of the Institute of Educational Sciences, 3*(4),42-57.

Üçgün, K. (2009). *The design of student automation for middle education schools and data mining applications on students' data,* Unpublished master thesis, Marmara University, İstanbul.

Wongkhamdi, T., & Seresangtakul, P. (2010). A comparison of classical discriminant analysis and artificial neural networks in predicting student graduation outcomes. *Proceedings of the Second International Conference on Knowledge and Smart Technologies*, (pp. 29-34).

Yagci, A & Çevik, M. (2019). Prediction of academic achievements of vocational and technical high school (VTS) students in science courses through artificial neural networks (comparison of Turkey and Malaysia). *Education and Information Technology, 24***,** 2741–2761 (2019). https://doi.org/10.1007/s10639-019-09885-4.

Yorgancı, N., & Işık, N. (2019). The use of artificial neural networks in classifying the average grade of science teachers' candidates. *Education And Society In The 21st Century, 8*(22), 146-159.

Yurtoğlu, H. (2005). A perdiction model with artificial neural network*: Sample of Turkey for some macroeconomic variables*. Unpublished doctoral dissertation. Ankara University, Ankara.