KURAM VE UYGULAMADA EĞİTİM BİLİMLERİ EDUCATIONAL SCIENCES: THEORY & PRACTICE

Received: August 3, 2015
Revision received: November 2, 2015
Accepted: February 26, 2016
OnlineFirst: April 20, 2016

www.estp.com.tr
DOI 10.12738/estp.2016.3.0214 • June 2016 • 16(3) • 943-964

Research Article

Using Neural Network and Logistic Regression Analysis to Predict Prospective Mathematics Teachers' Academic Success upon Entering Graduate Education

Elif Bahadır[1]
Yildiz Technical University

Abstract

The ability to predict the success of students when they enter a graduate program is critical for educational institutions because it allows them to develop strategic programs that will help improve students' performances during their stay at an institution. In this study, we present the results of an experimental comparison study of Logistic Regression Analysis (LRA) and Artificial Neural Network (ANN) for predicting prospective mathematics teachers' academic success when they enter graduate education. A sample of 372 student profiles was used to train and test our model. The strength of the model can be measured through Logistic Regression Analysis (LRA). The average correct success rate of students for ANN was higher than LRA. The successful prediction rate of the back-propagation neural network (BPNN, or a common type of ANN was 93.02%, while the success of prediction of LRA was 90.75%.

Keywords

Back-propagation neural network • Logistic regression analysis • Academic success • Graduate education

1 **Correspondence to:** Elif Bahadır, Department of Primary Education, Faculty of Education, Yildiz Technical University, Istanbul Turkey. Email: elfbahadir@gmail.com

Citation: 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*, 943-964.

Graduate education has become increasingly popular across the spectrum of higher level education. Higher education institutions have always been interested in predicting the paths of students. Thus, they are interested in identifying which students will require assistance as they enter the graduate program. Upon graduation, the students in an educational faculty may either continue in postgraduate programs or become a state or private school teacher. In this way, student performance is critical for ensuring academic success. Student learning in school significantly influences one's future career, particularly for students learning to teach elementary school mathematics. In recent years, prospective teachers have preferred entering postgraduate programs because of having shown more effective teacher performances or having chosen an academic career. A high GPA as an undergraduate is one of the conditions required to be able to enter postgraduate programs. This is important because the ability to predict an undergraduate's success of graduating brings with it the ability to predict their chances of success in being admitted to graduate studies. "To better manage and serve the student population, institutions need better assessment, analysis, and prediction tools to analyze and predict student-related issues." (Sayah & Mehda, 2010, p. 6). These prediction tools can be very helpful in managing and assisting students through their graduate education as well as the four year institutions that serve hundreds of students through various graduate programs. It is possible to determine and guide prospective teachers who plan to have a postgraduate education in accordance with successful prediction methods.

Through literature reviews, several modeling methods were found to have been applied in prior educational researches to predict students' retention. The more frequently used ones were logistic regression, structural equation modeling (SEM), decision trees, discriminant analysis, and neural networks.

Neural Network, Logistic Regression Analysis, and Academic Success

Success, in its most general sense, is progress towards a desired goal (Wolman, 1973). Success is an indication of the extent to which an individual benefits from a certain course or academic program in a school environment (Carter & Good, 1973). When expressing success in education, academic achievement refers to the grades one earns in class as given by teachers, test scores, or both (Carter & Good 1973). In terms of the above-mentioned definitions, academic achievement, as expressed in this study, refers to the achievement of teacher candidates in their designated courses throughout their undergraduate study and their success at being admitted to postgraduate study programs as predicted through their achievements in their courses.

The fact that a prediction method can bring with it success in the decision-making process, thus ensuring maximization of benefits, increases interest in the method of prediction. The studies conducted and methods used regarding prediction methods are becoming increasingly diversified along with such increasing interest. ANN and LRA techniques are the most important of these models (Yurtoglu, 2005). ANN and LRA are also the two most common methods used in predicting academic achievement.

In the literature, research in academic-achievement prediction is focused on two groups. The first group is studies that have been conducted regarding the scores students are expected to get from certain tests; students , are categorized by types of intelligence to determine their student profiles. The other group is studies that have been conducted with data mining techniques, which are based on inferring meaningful information from the pile of data at hand. Many statistical methods are used in tandem in data mining, and such methods are compared in terms of their success. ANN is one of the methods frequently used in data mining.

This study is suitable for modeling the questions with ANN and LRA due to the problem of the uncertainty of academic achievement predictions and the achievement criteria that can only be evaluated based on the data from scores at hand and the hierarchical structure of such criteria. The first reason for modeling our research problem using ANN is that it is an alternative to other conventional statistical methods employed in educational sciences and is one of the most effective methods used for prediction purposes. Furthermore, because it has been effective as a model in the literature regarding prediction analysis, it is also quite significant in this study as we are predicting the academic achievement of students.

ANN can offer linear and nonlinear modeling without the need of any preliminary information on input or output variables. Therefore, ANN is more general and flexible as a prediction tool when compared to other methods (Zhang, Patuwo, & Hu, 1998).

The purpose of using LRA is the same purpose as is in other model structuring techniques used in statistics: to establish a biologically acceptable model that can define the relations between dependent and independent variables in order to obtain an ideal consistency by using the minimum number of variables. Studies analyzing students' performances have been conducted using statistical analysis (Bresfelean, Bresfelean, Ghisoiu, & Comes, 2008; Flitman, 1997; Karamouzis & Vrettos, 2009). Artificial Neural Network (ANN) has been used to predict students' success (Siraj & Abdoulha, 2009), while a comparative study between ANN and statistical analysis for predicting students' final GPA has also been conducted (Naik & Ragotiaman, 2004).

Some researchers (Karamouzis & Vrettos, 2009) have attempted to present the development and performance of Artificial Neural Networks (ANN) for predicting community college graduation outcomes, as well as the results of applying sensitivity analysis on the ANN parameters, in order to identify the factors that result in a successful graduation. The need for disability services, the need for support services, and the student's age when they had applied to college were identified as the three factors that had contributed the most to successful and unsuccessful graduation outcomes. Siraj and Abdoulha (2009) considered the discovery of hidden information within university students' enrollment data. For predictive analysis, three techniques were used: neural network, logistic regression, and the decision tree. Their study showed that the neural network they had obtained gave the most accurate results among the three techniques. Flitman (1997) compared the performance of neural networks, logistic regression, and discriminant analysis for analyzing student failures. Neural networks were found to perform better than other methods. Conversely, Walczak and Sincich (1999) compared the results of a logistic regression analysis to that of a neural network model for modeling student enrollment decision making to show the improvements gained by using neural networks. The authors concluded that the level of performance of the neural network was not significantly higher than that of the other models. SubbaNarasimha, Arinze, and Anandarajan (2000) compared a neural network to regression analysis by introducing skewness in the dependent variable. In one of the two applications, they presented a comparative analysis of the predictions of a group of MBA student's performance. Researchers (Naik & Ragotiaman, 2004) developed a model to predict MBA student performance using logistic regression, probability analysis, and neural networks. The result was that the neural network model had performed better than the statistical models. They concluded that bias had been higher in the neural network model, compared to the regression model, because the absolute percentage error was lower in the case of the regression model. It can be observed from the literature that neither neural networks nor statistical techniques have performed consistently well (Paliwal & Kumar, 2009).

Purpose and Significance of the Study

Given that studies conducted using ANN in educational field have focused on classification of success rather than its prediction, this study intends to introduce a new perspective to predict students' success by using ANN. Considering that the scope of our problem is to predict academic achievement, our objective is to use ANN as an alternative to conventional methods in the educational field and to make an effective prediction of the achievement of students for their postgraduate study. We intend to make this prediction through LRA by using the same variables, comparing the success rates of both methods, and finding out the extent to which the prediction performance of ANN, which offers successful predictions in different fields in the world, can give successful prediction results in the field of education.

The prediction model built using the ANN technique and the model established using the LRA method were compared in terms of their prediction success; the comparison involved analyzing the changes in the performance of the ANN method depending on learning parameters such as size of the education and test data sets, structure of the network used, method of learning and the learning coefficient, momentum, and number of repetitions used for education. The purpose of the study is to use ANN, which has also been employed as an effective prediction method in different sectors, as an alternative to conventional methods in the educational field and to make an effective prediction of the educational success of students for their postgraduate study. It is also intended to make these predictions through LRA by using the same variables, then compare the success rates of both methods and find out the prediction performance of ANN, which has offered successful predictions in different fields in the world.

The significance of this study can be summarized as a comparison of the performances of ANN and LRA methods as prediction models by defining whether the models built by using the ANN method could be an alternative to the LRA method that has been long used in the field of education. In this way, it can contribute to the studies conducted in areas that use these techniques for predicting teacher candidates' postgraduate achievement in the educational field. It can also provide information that may be useful for educational faculty administrators, instructors, and students.

Research Questions of the Study

In this study, predictions about prospective teachers' graduate education success were analyzed. Logistic regression analysis, which is one of the most widely used statistical methods for examining the relations between variables, and the artificial neural network model were used together as predictive models. The success of these models was then compared.

There are three important requirements for admission to graduate education in Turkey. These are one's GPA, foreign language proficiency, and the Academic Personnel and Graduate Education Entrance Exam (ALES) grade. ALES is similar to the Graduate Management Admissions Test (GMAT). Undergraduate success rate is important to students. Students who want to enter postgraduate education must pay attention to their success during the first year.

The importance of this issue for prospective teachers is obvious: school drop outs are more likely to earn less than those who graduate and those who have started postgraduate education. This study wants to apply and compare the back-propagation neural network (BPNN), which is a common class of ANNs, and LRA for accurate predictions and classification of success for the learning effects of prospective teachers during graduate education. This prediction is important for students, teachers, and student career consultants. They appreciate these predictions because they can see their deficiencies. Moreover, the student-learning effect should be watched continuously for improvement. This study aims to determine the prediction success of LRA and BPNN, using General Mathematics, Pure Mathematics, Analysis I, Analysis II, Geometry, Linear Algebra-I, Analysis3, Special Teaching Methods 2, Elementary Number Theory, Algebra, and Problem Solving as variables. These variables can classify and predict students' performance in terms of success and entering postgraduate education.

In this context, answers will be sought to the following questions:

To what extent is ANN successful at predicting teacher candidates' academic achievement and admission to postgraduate programs?

To what extent is the LRA successful at predicting teacher candidates' academic achievements and admission to postgraduate programs?

Which of these two methods yields more effective results?

Neural Networks

The most important reason why we have used Artificial Neural Networks (ANN) in our research for modeling is that it is an alternative to other traditional statistical methods that have been used in educational sciences. Another reason is that it is one of the most effective methods that have been used to predict since the late 1980s. Besides, in this study, which makes predictions for academic achievements, the fact that it has been an effective model in the literature that analyzes predictions is of great importance.

ANNs are computer systems developed for the purpose of automatically realizing certain abilities, such as deriving, producing, and discovering new information by way of learning. This is one of the capabilities of the human brain, which it does without help. Artificial neural networks look into the happenings of events. They generalize related events through these happenings, collect information, and decide upon new happenings that are encountered by using the information that has been learned. ANNs are mathematical systems that consist of numerous process components (neurons) that are interconnected in a weighted manner. Actually, a process component is an equation frequently referred to as a transfer function. Such process components receive signals from other neurons and produce a numerical result by combining and converting these signals. In general, process components roughly correspond to actual neurons and connect each other within a network; such a structure forms neural networks. In most ANNs, neurons that have similar characteristics are structured in layers, and are operated synchronously in terms of transfer functions. Almost all networks have neurons that receive data and neurons that produce outputs. Mathematical functions, which are the key component of ANNs, are by the architecture of the network. Behaviors of the ANNs, in other words how they associate the input data with the output data, are affected firstly by the transfer functions of neurons, how they are interconnected, and the weight of such interconnections.

A neural network is a well-developed modeling technology, and during the past decades it has been widely used in technical applications that involve predictions and classifications. The neural network model is especially attractive for modeling complex systems because of its favorable properties: its abilities to approximate universal functions, accommodate multiple non-linear variables with unknown interactions, and generalize well (Coit, Jackson, & Smith, 1998). More modeling details on applying neural networks to predict student retention in engineering can be found in Imbrie, Lin, and Malyscheff's (2008) study.

Many prior studies involving graduate student performance have used LRA and BPNN. Schwan (1988) found graduates' GPA (GGPA) to be significantly correlated to their GMAT score, undergraduate GPA, and junior/senior year GPA among Murray State University MBA students. Wongkhamdi and Seresangtakul (2010) compared discriminant analyses studies and ANN studies for their ability to predict student graduation outcomes. The average correct classification rate for ANN was higher than for classical discriminant analysis. Gayle and Jones (1973) and Baird (1975) found a significant positive relationship between Graduate Records Examination (GRE) scores and GGPA for graduate students. Paolillo (1982) employed step-wise regression in his study and found that the applicant's junior and senior undergraduate GPA was the first variable entered into their equation. A neural network study by Lee (2010) predicted learning effects in design students with an average accuracy of 93.54%. Deckro and Woundenberg (1977) studied nine variables as possible predictors of academic success among Kent State MBA students. Naik and Ragotiaman (2009) found that the neural network model performs as well as statistical models, and it is a useful tool for predicting MBA student performance. In the study by Ibrahim and Rusli (2007), the demographic profile and cumulative GPA (CGPA) of students in their first semester of undergraduate studies were used as the predictor variable for students' academic performance in their undergraduate degree program. Studies by Jun (2005) and Herrera (2006) have provided a comprehensive overview of theoretical models that describe student continuation and dropout rates in both distance education institutions and institutions attended in person. Levin and Wyckoff (1991), House (1993), Schaeffers, Epperson, and Nauta (1997), Beserfield-Sacre et al. (1997), Zhang and RiCharde (1998), French, Immekus, and Oakes (2005) have all used logistic regression models to study student persistence in colleges.

Overview of Back-Propagation Neural Network (BPNN)

"The back-propagation neural network is a multi-layer feedforward fully-connected network. This neural network is the most representative model of ANN due to its documented ability to model any function (Funahash, 1989; Hornik, Stinchcombe, & White 1989). The BPNN is composed of three or more layers, including an input layer, one or more hidden layers, and an output layer. Each layer has a number of nodes, called processing units or neurons. One of the most important characteristics of the BPNN is its ability to learn by training samples. Proper training enables the network to memorize the knowledge involved in problem solving in a specific domain. Back-propagation learning uses a gradient-descent algorithm (Rumelhart, Hinton, & Williams, 1986), plus hidden layer and nonlinear transfer function to minimize error function. The training data set is initially collected to develop a BPNN model. Through a supervised learning rule the data set consists of an input and an actual output (target)." (Lee, 2010, p. 256)

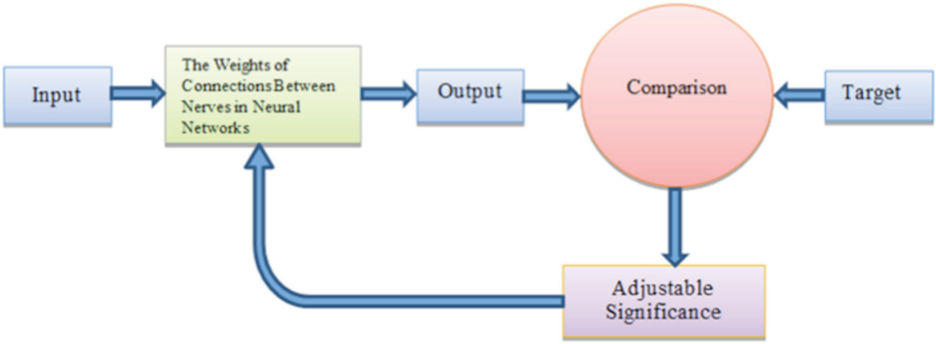

Figure 1. The working principle and training operation of back-propagation neural networks.

The working principle and training operation of a BPNN model (a type of artificial neural network), a back-propagation algorithm (BPA), and a multi-layered network can be seen in Figure 1.

The trained network receives information from outside through entry nerves and gives the produced outcome through output nerves. Although the training of multi-layered network takes a long time, obtaining results from the trained network with new inputs is very quick. The entries in the training set ensure that neurons at the input layer of the network produce outputs. This output constitutes the inputs of the next layer's neurons. Therefore, it provides that neurons at the input level produce the output of the networks. The output produced by the network is compared with the real data from the training set, and the success of the model is displayed by calculating the difference between them.

"The gradient-descent learning algorithm enables a network to improve the performance through self-learning. Two computational phases exist, namely the forwards and backwards phases. In the first phase, the BPNN receives the input data and directly passes it to the hidden layer. Each node of the hidden layer then calculates an activation value by summing the weighted inputs and then transforming them into an activity level using a nonlinear transfer function. One of the most common types of transfer functions is the sigmoid function which is continuous, nonlinear, differentiable everywhere, and monotonically non-decreasing. Each node of the output layer is used to calculate an activation value by summing the weighted inputs attributed to the hidden layer. A transfer function is then used to calculate the network output (i.e. predictive value). In the next phase, the actual network output is compared with the target value. If a difference (i.e. an error term) appears, the gradient-descent algorithm is applied to adjust the connected weights. Meanwhile, if no difference appears, then no learning is processed. This training process is also called supervised training since the target output for each input is known. The training process of BPNN generally involves five steps:

(1) Select representative training samples and turn them into the input layer as the input value.

(2) Calculate the predictive value of the network.

(3) Compare the target value with the predictive value to obtain the error value.

(4) Readjust the weights in each layer of the network according to the error value.

(5) Repeat the above procedure until the error value of each training sample is minimized, meaning that the training is finished." (Lee, 2010, p. 256)

Logistic Regression Analysis (LRA)

Logistic regression analysis is a common method that has been increasingly used particularly in the social sciences. In most socioeconomic researches that have been conducted to reveal causality relations, some of the variables analyzed have consisted of two-level data such as successful-unsuccessful, yes-no, and satisfied-dissatisfied. According to Agresti (1990), in case the dependent variable consists of two-level or multiple-level categorical data, logistic regression analysis plays an important role in analyzing the causality relationship between the dependent variable and independent variable(s).

In logistic regression analysis, which has the objectives of categorization and of investigating the relationships between dependent and independent variables, the dependent variables constitute the categorical data and take discrete values. As for the independent variables, all or some of them need to be continuous or categorical variables (Isigicok, 2003). Normal distribution assumption and continuity assumption are not prerequisites. Risk factors are defined as probabilities by means of obtaining the effects of explanatory variables on the dependent variable as the probability (Hosmer & Lemeshow, 2000; Ozdamar, 2002). Logistic regression analysis, which has been of used recently, is one ofthree methods used in designating observations to groups, the others being clustering analysis and discriminant analysis.

Logistic regression analysis is an alternative method to discriminant analysis and cross-validation tables in case of failure to establish certain assumptions of regression analysis, such as having normality and common covariance. While it can also be used in cases where the dependent variable is a discrete variable having two or multiple levels (such as 0 and 1), the mathematical flexibility and easy interpretability of this method have increased the interest in this method (Hosmer & Lemeshow, 2000; Tatlidil, 2002).

> The predictor variables may be either numerical or categorical (dummy variables). This model is used for predicting the probability of the occurrence of an event by fitting data to a logistic curve. With a given numerical cutoff (often 0.5), cases with probabilities above this value are categorized as a 1 (success), whereas cases lower than this value are classified as a 0 (failure). Thus, logistic regression is an appropriate statistical procedure to be used in the original study to predict success as an actuarial major." (Schumacher, Olinsky, Quinn, & Smith, 2010, p. 260)

Method

The research methodology in this study aims to determine the utilization of the BPNN model and logistic regression method as a supportive decision-making tool for predicting learning effects of the students of Elementary School Mathematics teaching and to estimate their chances of entering graduate education. The data were analyzed using the neural solution and MATLAB and SPSS programs. We obtained output from logistic regression analysis (to compare with the traditional SPSS logistic regression) and neural networks, which will both be subsequently. Accordingly, the comparative qualitative research method was used in our research.

Data Collection

This study collected the grades of students who had graduated from the Department of Mathematics Education. The information was comprised not only of students' first-year grades for all courses (which included General Mathematics, Pure Mathematics, Analysis I, Analysis II, Geometry, and Linear Algebra-I), but also their professional core course grades at the upperclassman level, which included Analysis3, Special Teaching Methods 2, Elementary Number Theory, Algebra, Problem Solving, and their success at entering a postgraduate program.

The sample group of the research was composed of students from three different universities who were studying or had studied elementary school mathematics teaching. In this way, the researcher selected the purposeful sampling group.

The study group of the research was determined using the easily accessible sampling method. This sampling method was preferred since the score data in these universities was more easily accessible, and two of the universities designated as sample groups had offered postgraduate mathematics study for a higher number of students for many years. This sampling method provides speed and practicality to the research, as the researcher had selected an easily accessible situation (Yıldırım & Simsek, 2006).

Table 1
The Data Set of the Implementation

Data Set Number	3 Different Universities with an Institute of Educational Sciences, Elementary School Mathematics Teaching Department	Input Years	Quantity of Data
1	Those who were working on or had received a master's degree between 2006-2010	2006	4
		2007	6
		2008	5
		2009	11
		2010	13
		2011	15
		2012	14
		2013	12
2	Those who completed their graduate education program during the 2010-2011 academic year	2008- 2011	140
3	Those who had entered a graduate program in 2010 and continued on to postgraduate studies	2010-2014	152
Total Quantity of Data			372

Data Analysis

Data were collected from 3 different universities. Grade information was recorded for a total of 220 students. Afterwards, this information was employed for the training and testing stages of the BPNN. To assess the BPNN model's ability to predict learning effect in students studying in elementary school mathematics teaching, the 176 data sets (80% of the total grades information) were randomly selected from the 220 data sets of the total grade information used for BPNN model building (i.e., the training samples). The remaining 44 data sets (20% of the total grades information) were then used to test the prediction accuracy of the BPNN model; i.e. the testing samples.

The input layer, which included General Mathematics, Pure Mathematics, Analysis I, Analysis II, Geometry, and Linear Algebra-I, were taken as the input variables (input nodes) for the input layer of the BPNN. Therefore, the input layer contained a total of six nodes.

The output layer, which included Analysis3, Special Teaching Methods 2, Elementary Number Theory, Algebra, Problem Solving, and successful entrance to postgraduate education, were used as the output variables (i.e., output nodes). The output layer thus contained six nodes. In the data set, the data was separated randomly into two sections: training and test sets. Training sets are mainly used for learning what is suitable in terms of the significance of the data. Test sets are used entirely for evaluating the performance of a specific classifier technique. The training set was used to train the network, and the test set was used for assessing the performance of training in the implementation. 80% of the data set comprised the training set and

20% comprised the test set. Additionally, the number of neurons in the hidden layers was determined prior to analysis by looking at the significance of the classifier, which is known as the validity set.

In order to evaluate the number of hidden layers in the problem, the performance of the validity data was analyzed. In order to test the performance of the network structure, the mean squared error (MSE) and mean absolute error (MAE) were employed.

For the training of the network, 10,000 iterations were carried out. As a result of the neural network analysis, separate classification tables were obtained for each set. The accuracy percentages obtained from each set were different. It was necessary to combine the results obtained from the three sets to calculate the percentage of general accuracy.

The percentage of classification accuracy for the training set was found for each event. The mean absolute error mean and mean squared error were found. A training set classification table was created for implementing what would be predicted, and the percentage of classification accuracy was found for the test set. The mean absolute error and mean squared error were found. In order to obtain the classification table in accordance with ANN implementation, the training, validity, and test sets were combined. The assigned values (for inhibition maximum as -1.0 and excitation maximum as +1.0) were added. The obtained data have been presented in Table 2 in the results section.

Neural Networks Analysis

An artificial neural network was employed in order to recognize the links between the prediction model for learning effects of prospective teachers and their chances of entering a postgraduate program. The network, shown in Figure 1, is a feed-forward neural network which consists of three layers. The input layer has a total of 6 nodes. Each node represents a graduate course. The output layer has six nodes which represent graduate courses and successful entrance to a postgraduate program.

The network was trained and tested. The first 176 observations were used as the training set for the training process while the other 70 were used for testing. Therefore, 81% of the samples were used for training and 19% were used for testing; this is considered to be a good proportion for modeling nonlinear functions, according to Granger (1993). A logarithmic sigmoid function was used as a transfer (activation) function to connect the neurons of the input layer with those in the hidden layer. Subsequently, a linear activation function connected the hidden layer nodes with the output layer nodes. The learning algorithm for training follows the back-propagation learning rule (Rumelhart et al., 1985). The performance goal for error was set to 0.01 and the training epochs to 10,000. Finally, a normalization took place before the training, which scaled the variable range of the survey into a scale of 0–1, as required for neural network training. A complex network is more difficult to train and usually takes more time, as more epochs are required to complete the tasks. On the other hand, most of the problem domain involves large amounts and variables. Removing any of the data or variables, even what some consider less relevant, could affect the system knowledge. Even small amounts of information can affect the whole process.

The academic achievement prediction was made by analyzing the 292 students' grades from the field courses they had attended in their first years in the faculty. The network pattern of the BPNN, with its 6 input and 6 output layers as used in this study, can be seen in Figure 2. The information provided in the figure is as follows: the input and output layers of the ANN, the number of examples and transfer function of the BPNN model together with the number of operation elements in the hidden layer, and the learning theorem and threshold level which had been identified for obtaining the value of expected error. The mean squared error (MSE) value continued learning until it reached a threshold level. The minimum function was employed while doing this. By the time the expected error value had been reached, the number of training epochs had stopped before reaching 10,000.

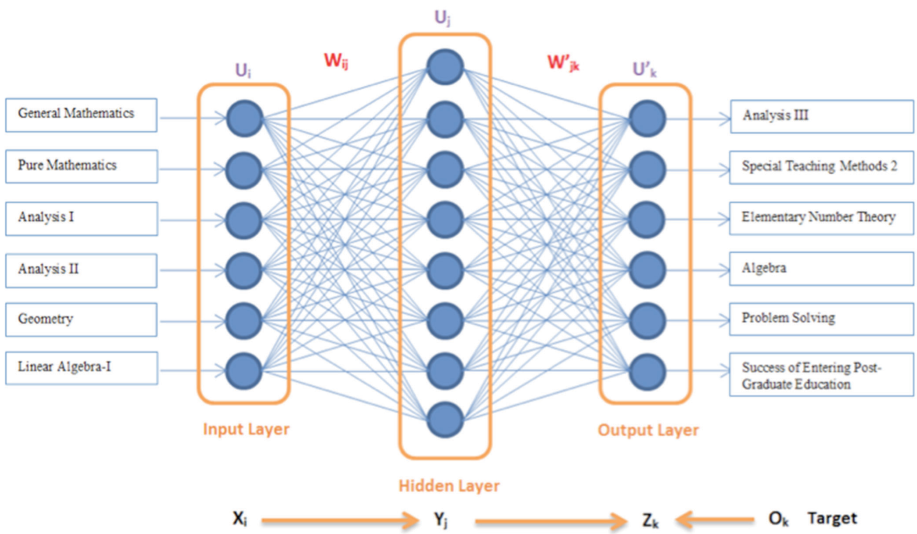

Figure 2. Architecture of the three-layer BPNN used in this study.

For the testing data, 58 of the 292 data were used. The data from the training set and the network were trained and the possibility of generalizing the results produced by the network was tested. For testing the data, 29 data from the data table were selected as Successful (1), and the other 29 data were selected as Unsuccessful (0). The total number of data for the test was 58.The significance analysis was carried out after all the samples in the training set had been shown to the network.

Logistic Regression Analysis

The logistic regression model was generated using SPSS 17.0. This study may be useful in categorizing students in terms of their level of achievements in undergraduate programs by means of using the achievement variables of prospective teachers in courses such as General Mathematics, Pure Mathematics, Analysis I, Analysis II, Geometry, Linear Algebra-I, Analysis3, Special Teaching Methods 2, Elementary Number Theory, Algebra, and Problem Solving.

Considering the prospective teachers' academic achievement from the university where we had access to data from the years 2006-2010 and the passing grades they had received from specific courses during their undergraduate education in the Elementary School Mathematics Teaching department in the Faculty of Education, their success at being accepted in a postgraduate program was identified as the predictable variable. The technique of logistic regression analysis was used because the dependent variable had a categorical structure. The results of the research showed that eleven variables were statistically significant.

Results

The results of the study have been given under three headings: the ANN results, LRA results, and the comparison of both models according to the research questions.

Results of the ANN

In our case, a back-propagation network with three layers seemed to be the most appropriate technique. The input layer had 6 neurons, and the output layer had 6 neurons. The number of nodes in the hidden layer was 8.

Table 2
Performance When Tested with the Tested Data

Performance	Successful	Unsuccessful
MSE	0.25	0.16
NMSE	1.35	1.36
MAE	0.31	0.29
Min Abs Error	0.00	0.00
R	0.21	0.21
Percent Correct	98	88.04

After optimizing the network's structure and training the data within 10,000 epochs, we tested the network's prediction power on the data set. The accurate classification success of the network was found as 93.02%.

Results of the LRA

General Mathematics, Pure Mathematics, Analysis I, Analysis II, Geometry, Linear Algebra-I, Analysis 3, Special Teaching Methods 2, Elementary Number Theory, Algebra, and Problem Solving courses were used as independent variables. These courses were used as variables for all steps of the logistic regression analysis.

Table 3
LRA Classification Matrix

		Predicted		
		Fail	Pass	Percentage of accuracy
Observed	Fail	114	8	94.7
	Pass	10	88	86.8
Percentage of accuracy				90.75

It was observed that the rate of correct classification was 90.75% with the given logistic regression model. The findings of the study reveal that 86,8% of the students who were successful in entering graduate education and 94,7% of students who were not successful were estimated correctly.

Comparison of Two Models: LRA and ANN

To compare the performance of the neural network approach with the LRA approach, the Mean Correct Classification Rate (Mean CCR) of predictive accuracies in the neural network and LRA models are shown in Table 4. Clearly, the neural network method demonstrated a superior ability to predict the student graduation outcomes.

Table 4
Prediction Results from Two Different Prediction Methods for Two Models

Model	Methods	Overall Accuracy	Sum of Mean Square Error (MSE)	Mean Absolute Error (MAE)
Grades from all courses and success of entering graduate education total data from eleven variables	Neural Networks	93.04	0.20	0.30
	Logistic Regression	90.75	0.21	0.33

Table 4 shows the classification results for the LRA technique and the neural network model as described in this research. The percentage of average correct classifications for the Artificial Neural Networks was higher than LRA. The success of prediction for the BPNN was 93.02%, while the success of prediction for the LRA was 90.75%.

Discussion

Although logistic regression is a popular method for predicting a categorical variable, neural networks as an alternative technique are more effective for prediction. Because logistic regression ignores many cases with missing data in the predicted variables, neural networks can include all data with very promising results.

The aim of this research was to determine the effectiveness of artificial neural networks in forecasting the chances of entering postgraduate education for prospective elementary mathematics teachers. In our approach, we trained two models, a logistic regression analysis model and a three-layer supervised neural network model based on the back-propagation learning algorithm. This study proved that artificial neural networks are able to significantly improve the chances of accurately predicting entrance to a postgraduate program, when compared to logistic regression analysis. The prediction success rate of BPNN was 93.02%, while LRA's was 90.75%. These findings are quite consistent with other published study results, such as Lee (2010), Naik and Ragotiaman (2009), Ibrahim and Rusli (2007), Turhan, Kurt, and Engin (2013), and Hardgrave, Wilson, and Walstrom (1994).

To gain more insight and implications from the research, we identified the factors that had the most significant impact on student's chances of entering graduate education through the values of standardized coefficients in the logistic regression analysis. The results indicated that the most influential factors were the courses taken in their respective undergraduate programs. This factor is in agreement with the results obtained by Hedjazi and Omidi (2008) and Diaz (2003). The findings of the study help understand which students need preliminary assistance from their advisors.

Course contents used as input data in the study qualify as the continuation of the course contents used as output data. Because the course contents (see the Appendix) used in the research were the continuation of each other, it can be seen that achievements in these courses input data taken in the first years of the undergraduate study also determined their achievement in courses taken later. The predictions can be understood from their high achievements. It has been concluded that achievement in the courses that were similarly designated as inputs and outputs affected each other in other studies where the academic achievement was predicted using ANN (Shell, Vrooman, Renner, & Dawsey, 2001; Schumacher et al., 2010). A similar relationship between course contents has also been seen in research conducted in the field of medical education (Monique & Claude, 2000; Vedlinski, 2002).

The back-propagation neural network applied in this study showed high prediction accuracy for predicting the learning effect in students of Primary School Mathematics Teaching. The prediction results of the BPNN model provide insight to the educators of mathematics education, enabling them to tailor their teaching strategies in order to meet their students' individual needs. Additionally, mathematics education educators or student career consultants may also consider other factors related to a student's choice of major. Similarly, the analyses performed using the ANN were discovered to yield more effective results when compared to other prediction methods, and they are powerful guides in career management as seen from studies conducted to predict the academic achievement of engineering students, design students, MBA students, and medical students (Drecko & Woundenberg, 1977; Herrera, 2006; Jun, 2005; Lee 2010; Lykourentzou, Giannoukos, Nikolopoulos, Mpardis, & Loumos, 2009).

Although the difference in error values produced by prediction models as mentioned in studies was not high, the fact that the neural network model has been more successful than all other data sets shows that the artificial neural network technique could be an alternative for classical statistical methods in educational studies that rely on prediction. It is possible to determine academic success and guide prospective teachers who plan to enter a postgraduate program in accordance with the successful forecasts of prediction methods.

To achieve improved results in follow-up studies, a holistic scope that includes student learning effects and their first-year grades should be developed. Other factors might include the financial condition of students' families, educational background of the parents, parents' supportiveness, and parents' social status. All of these factors could be quantified and then input into the BPNN.

Limitations of the Study and Recommendations for Further Studies

The research has limitations in terms of the data collected from the three designated universities. In this context, the study may be enriched with data obtained from other universities that offer postgraduate study. The prediction of academic achievement in the research was limited to the use of the ANN, BPNN and logistic regression analysis methods. Models such as cluster analysis and decision trees can be used in addition to this study's methods to compare prediction success.

It must be noted that when using the ANN technology in predicting success in education, the coefficients related to the established model stay above the weights within the network and the weights cannot be interpreted yet, the network architecture with optimal characteristics can only be found through trial and error as there is no specific method that can be followed when developing ANN models. Furthermore, models obtained through the ANN technology are in a closed box; therefore, certain questions related to the model such as how and why have no answers. Such limitations of the ANN model should also be taken into consideration.

Limitations of the study come from the fact that the research was only carried out on students of mathematics education, and that other groups might show different characteristics. In future studies, the relationships between the number of attributes and the size of the training set, the type of neural network and the number of hidden layers, the number of nodes, and so on can be studied intensively.

Additionally, a more accurate prediction network might be designed by adjusting learning rate and momentum.

Future studies can have their directions set in determining an accurate, reliable prediction network for helping instructors, students, and parents in decision-making. Although limited data sets were used in this study, future studies might continue to collect information on students' grades and follow student job performance to demonstrate that the BPNN model is a useful method for predicting the learning effect of students of elementary mathematics teaching.

References

Agresti, A. (1990). *Categorical data analysis*. New York, NY: J.VV Iley & Sons.

Baird, L. L. (1975). Comparative prediction of first year graduate and professional school grades in six fields. *Educational and Psychological Measurement*, *35*(4), 941–946.

Bresfelean, V. P., Bresfelean, M., Ghisoiu, N., & Comes, C. A. (2008, June). Determining students' academic failure profile founded on data mining methods. In Information Technology Interfaces, *2008 30th International Conference on* (pp. 317–322). IEEE.

Carter, V., & Good, E. (1973). *Dictionary of education* (4th ed.). New York, NY: McGraw Hill Book.

Coit, D. W., Jackson, B. T., & Smith, A. E. (1998). Static neural network process models: Considerations and case studies. *International Journal of Production Research*, *36*(11), 2953–2967.

Drecko, R. F., & Wounderberg, H. W. (1977). MBA admission criteria and academic success. *Decision Sciences*, *8*, 765–769.

Flitman, A. M. (1997). Towards analysing student failures: Neural networks compared with regression analysis and multiple discriminant analysis. *Computers & Operations Research*, *24*(4), 341–347.

French, B. F., Immekus, J. C., & Oakes, W. C. (2005). An examination of indicators of engineering students' success and persistence. *Journal of Engineering Education*, *94*(4), 419–425.

Gayle, J. B., & Jones, T. H. (1973). Admission standards for graduate study in management. *Decision Sciences*, *July*, 421–425.

Hardgrave, B. C., Wilson, R. L., & Walstrom, K. A. (1994). Predicting graduate student success: A comparison of neural networks and traditional techniques. *Computers & Operations Research*, *21*(3), 249–263.

Hedjazi, Y., & Omidi, M. (2008). Factors affecting the academic success of agricultural students at University of Tehran, Iran. *Journal of Agricultural Science and Technology, 10*(3), 205–214.

Herrera, O. L. (2006). *Investigation of the role of pre-and post-admission variables in undergraduate institutional persistence, using a Markov student flow model* (Doctoral dissertation). North Carolina State University, USA.

Hosmer, D. W., & Lemeshow, S. (1989). *Applied logistic regression.* New York, NY: John Wiley & Sons.

House, J. (1993). The relationship between academic self-concept and school withdrawal. *Journal of Social Psychology, 133*, 125–127.

Ibrahim, Z., & Rusli, D. (2007, September). *Predicting students' academic performance: comparing artificial neural network, decision tree and linear regression*. Paper presented at the 21st Annual SAS Malaysia Forum, Shangri-La Hotel, Kuala Lumpur.

Imbrie, P. K., Lin, J. J., Reid, K., & Malyscheff, A. (2008). Using hybrid data to model student success in engineering with artificial neural networks. *Proceedings of the Research in Engineering Education Symposium,* 7-10.

Işığıçok, E. (2003, October). *Bebeklerin doğum ağırlıklarını ve boylarını etkileyen faktörlerin lojistik regresyon analizi ile araştırılması* [Analyzed by logistic regression analysis of the factors affecting birth weight and length of the Baby]. Paper presented at VI Ulusal Ekonometri ve İstatistik Sempozyumu, Gazi Üniversitesi İktisadi ve İdari Bilimler Fakültesi Ekonometri Bölümü, Ankara, Turkey.

Jun, J. U. S. U. N. G. (2005). *Understanding dropout of adult learners in e-learning* (Doctoral dissertation). The University of Georgia, Athens, GA.

Karamouzis, S. T., & Vrettos, A. (2009, March). Sensitivity Analysis of Neural Network Parameters for Identifying the Factors for College Student Success. *In Computer Science and Information Engineering, 2009 WRI World Congress on* (Vol. 5, pp. 671–675). IEEE.

Lee, Y. J. (2010). Neural network based approach for predicting learning effect in design students. *International Journal of Organizational Innovation*, *2*(3), 250. Retrieved from http://www.ijoi-online.org

Levin, J., & Wycokoff, J. (1991). Predicting persistence and success in baccalaureate engineering. *Education, 111*(4), 461–468.

Lykourentzou, I., Giannoukos, I., Nikolopoulos, V., Mpardis, G., & Loumos, V. (2009). Dropout prediction in e-learning courses through the combination of machine learning techniques. *Computers & Education*, *53*(3), 950–965.

Monique, F., & Claude, F. (2000). Decision-support and intelligent tutoring systems in medical education. *Clinical & Investigative Medicine, 23*(4), 266–269.

Naik, B., & Ragotiaman, S. (2004). Using neural networks to predict MBA student success. *College Student Journal*, *38*(1), 143–149.

Paliwal, M., & Kumar, U. A. (2009). A study of academic performance of business school graduates using neural network and statistical techniques. *Expert Systems with Applications: An International Journal*, *36*(4), 7865–7872.

Paolillo, J. G. (1982). The predictive validity of selected admissions variables relative to grade point average earned in a master of business administration program. *Educational and Psychological Measurement*, *42*(4), 1163–1167.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). *Parallel distributed processing: Explorations in the microstructures of cognition* (Vol. 1). Cambridge, MA: MIT Press.

Schaeffers, K. G., Epperson, D. L., & Nauta, M. M. (1997). Women's career development: Can theoretically derived variables predict persistence in engineering majors? *Journal of Counseling Psychology, 44*, 173–183.

Schumacher, P., Olinsky, A., Quinn, J., & Smith, R. (2010). A comparison of logistic regression, neural networks, and classification trees predicting success of actuarial students. *Journal of Education for Business*, *85*(5), 258–263. http://dx.doi.org/10.1080/08832320903449477

Schwan, E. S. (1988). MBA admissions criteria: An empirical investigation and validation study. *Journal of Education for Business, 63*(4), 158–162.

Siraj, F., & Abdoulha, M. A. (2009, May). Uncovering hidden information within university's student enrollment data using data mining. In *2009 Third Asia International Conference on Modelling & Simulation* (pp. 413–418). IEEE.

SubbaNarasimha, P. N., Arinze, B., & Anandarajan, M. (2000). The predictive accuracy of artificial neural networks and multiple regression in the case of skewed data: Exploration of some issues. *Expert Systems with Applications*, *19*(2), 117–123.

Turhan, K., Kurt, B., & Engin, Y. Z. (2013). Yapay sınır ağları ile öğrenci başarısı tahmini [Student achievement estimated with artificial neural networks]. *Eğitim ve Bilim*, *38*(170), 112–120.

Walczak, S., & Sincich, T. (1999). A comparative analysis of regression and neural networks for university admissions. *Information Sciences*, *119*(1-2), 1–20.

Wongkhamdi, T., & Seresangtakul, P. (2010, July). 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,* 29–34.

Yıldırım, A., & Simsek, H. (2006). *Sosyal bilimlerde nitel arastirma yöntemleri* [Qualitative research methods in the social sciences]. Ankara, Turkey: Seçkin Yayıncılık.

Yurtoglu, H. (2005) *Yapay sinir ağları metodolojisi ile* öngörü *modellemesi: Bazı makroekonomik değişkenler için Türkiye* örneği [Artificial neural networks and predictive modeling methodology: some macroeconomic variables for Turkey] (Uzmanlık Tezi, Ekonomik Modeller ve Stratejik Araştırmalar Genel Müdürlüğü).

Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, *14*, 35–62.

Zhang, Z., & RiCharde, R. S. (1998, May). *Prediction and analysis of freshman retention*. Paper presented at the Annual Forum of the Association for Institutional Research, Minneapolis, MN. Retrieved from http://files.eric.ed.gov/fulltext/ED422814.pdf

Appendix

Brief Descriptions of the Courses in the Elementary Mathematics Education Bachelor Program

Semester	Courses	Contents
1st Semester	General Mathematics	The sets and properties of natural numbers, integers, rational numbers, and real numbers. Quadratic equations and inequalities, analytical studies of lines, circles and related applications. Concepts of functions, polynomials, rational functions, trigonometric functions, hyperbolic functions, antilogarithms, logarithmic functions, and elementary functions that occur as inverses of antilogarithm and logarithmic functions. Function graphs. Principle of induction, properties of sum and product symbols, fundamental concepts on sequences and series. Complex numbers and their properties.
2nd Semester	Pure Mathematics	Conceptual explanations of axioms, theorems, sets, and methods of direct and indirect mathematical proof. Axioms and theorems of symbolic logic, applications on symbolic logic. Universal and existence quantitatives, conceptual operations on sets. Cartesian product of sets, graph plotting, relational concept and properties, varieties of relation, ordered and equivalence relations and their properties. Construction of numbers assisted equivalence relation. Function concept, inverse functions, types of functions, composite functions, operations with functions. Power concept in mathematics, finite and infinite sets.
2nd Semester	Geometry	Definition of geometry, its structure and real life applications. Axiom, undefined terms, explanation of theorem. Euclidean and non-Euclidean geometries. Basic axioms of Euclidean geometry. Relations between point, line, and plane in space. Concept of the angle, types of angles, congruent angles, congruence axioms, applications on angles. Definition of polygon. Definition of triangle, types of triangles, basic and aide-components of a triangle. Congruence axioms and theorems about triangles, applications of congruent triangles, similarity theorems about triangles, applications of similarity on triangles. Proof of theorems regarding to trapezoids, parallelograms, equilateral quadrangles, rectangles, and squares. Applications on quadrangles. Concepts of circle and ellipse, theorems and proofs of angle and length for circles and ellipses, applications of angle and length for circles and ellipses. Properties of objects in space, area and volume of solid objects.
3rd Semester	Analysis I	Concept of limit, its applications on single variable functions. Continuity and applications, varieties of discontinuity on single variable functions. Concept of derivatives, derivative methods on single variable functions. Derivative of polynomial, trigonometric, logarithmic, exponential, and hyperbolic functions, their inverses, and derivative of closed functions. Higher order derivatives. Extremum and absolute extrema points of functions, extremum problems and practice on different fields. Rolle and Cauchy Mean Value Theorems. Finite Taylor Theorem. L'Hospital Rule and limit practices by the help of this rule. Differential and linear rise. Concept of integral, indefinite integrals, integral methods, definite integral, calculating area and volume with the help of definite integrals, practices in different areas.
3rd Semester	Linear Algebra I	Vectors in R^2 and R^3, m x n matrices; addition and scalar products in matrices, linear independence in matrices, introduction to concept of vector space. Linear equation systems, Gauss elimination, subspaces. Linear independence and dimension. Linear transformations, relationship between linear transformations and matrices, matrix product, inverse matrix and applications.

4th Semester	Analysis II	Concept of multi-variable functions, definition of function and value sets, function drawings. Limit concept in two value functions and applications, concept of continuity. Partial derivative in two value functions, chain rule, differential increase and linearization, local extreme values, absolute extreme values and applications, Lagrange factors, concepts of two-multiple integrals, calculating volumes with two-multiple integrals.
5th Semester	Analysis III	Concept of sequence and applications. Concepts of series, series with positive term, divergent sequence, convergent sequence, alternating series, criteria convergence, power series. Function series, point and regular convergence on function series, convergence tests, Taylor series, applications in real life. Fourier series.
5th Semester	Introduction to Algebra	Binary operations, definition of group, subgroups, permutation groups, homomorphism, cyclic groups, residue classes, normal subgroups, quotient groups, definition of ring, sub-rings, ideals.
5th Semester	Special Teaching Methods II	What is a problem? Solving problems. Importance of problem solving, classification of problems, purpose of teaching problem solving and process of problem solving; teaching problem solving which needs four-operations, strategies of extraordinary problem solving. Natural numbers and operations in natural numbers, fractions and teaching fractions, measurements and teaching, data analysis, teaching geometry. Learning based project. Prepare a subject plan, presentation and evaluation.
7th Semester	Elementary Number Theory	Divisibility on integers, prime numbers, important functions on number theories, congruent, linear congruent, separating prime products on integers, primitive roots and indexes, quadric residues (second degree), cryptography subjects and usage areas in real life, continuous fractions.
7th Semester	Problem Solving	The student will be able to make a presentation according to the mathematical problem-solving process, evaluate the problem-solving process in the mathematics curriculum, have positive attitudes and beliefs towards problem solving, will be able to use problem-solving strategies, pose and model mathematics problems and comprehend problems and problem solving processes.