

Predictions of academic achievements of vocational and technical high school students with artificial neural networks in science courses (physics, chemistry and biology) in Turkey and measures to be taken for their failures

Ali Yağcı^{1a}, and Mustafa Çevik²

¹ Karamanoğlu Mehmetbey University, Institute of Science, Karaman, 70100, Turkey

² Karamanoğlu Mehmetbey University, Education Faculty, Department of Primary Education, Karaman, 70100, Turkey

Abstract. The aim of this study is to predict academic achievements of vocational and technical high school (VTHS) students with artificial neural networks (ANN) in physics, chemistry and biology courses in Turkey and reveal measures to be taken for their failures. The study group consisted of 922 students studying in 10th and 11th grade in VTHS. This study was conducted with the survey method and a 34-item demographic questionnaire was developed in order to collect the data. The parameters in the questionnaire were identified as the items that were considered to influence academic achievements of the students. Opinions of 3 field specialist, 1 measurement and evaluation specialist and 2 technical teachers were taken and it was supported by the literature for the content validity and the KR20 reliability coefficient was found as .90 using SPSS 16.0 package program. The items in the questionnaire that are considered to influence academic achievements of the students were approved as independent variables/input and academic achievement mean scores of the students in physics, chemistry and biology courses in the previous year were approved as dependent variables/output. Academic achievements of the student were predicted with ANN in Matlap R2016a program using these parameters and a model was created. A successful academic achievement prediction system with an average sensitivity of %98 was developed over 922 data and the measures to be taken to prevent the failures of the students were determined at the end of the study.

Keywords: Vocational and technical high schools, artificial neural networks, and science courses academic achievement prediction

^a Corresponding author: ayagci89@gmail.com

1 Introduction

Vocational and technical education institutions, which play an important role in the development of the country, require increasing productivity in education to raise the number and quality of human power that the society needs [1]. The question to be raised in order to increase productivity in education is firstly the relevance of the instructional programs with the personality traits of individuals, job specifications and institutional strategies and to what extent these factors are taken into account when preparing the program. Therefore the need analysis/assessment is a prerequisite for an effective instructional program. Thus, one of the most important issues in research related to education is whether the need analysis/assessment was made before the teaching [2]. It is vital for the vocational schools to carry out the need analysis/assessment. This is because vocational students are expected not only to complete their education that supports their cultural and personal development, but also to complete their vocational training at the best level to ensure their future [3]. Vocational high schools are educational institutions aiming at students to gain work experience at an early age and get involved in business life, meeting the need for intermediate staff in the working segment of the society with lessons that mostly require skills and practice. In this respect, vocational high school graduates need to cultivate themselves in every sense. These characteristics of students ought to be considered when preparing especially instructional programs of science courses besides vocational courses [4]. From this point of view, the success of vocational education, an area that can be regarded as the benchmark of the society, is one of the factors that directly affect the socioeconomic structure of the country. In this context, what need to be done for qualified student training in VTHS and what measures should be taken to avoid failures are the main objectives of educational researches. This situation was statistically determined in the university entrance exam conducted by the Measuring, Selection and Placement Center (MSPC) in Turkey. For instance, the achievement average score of MS (mathematics-science) courses of vocational high schools is below the country average score in 2016. In recent years, the Ministry of National Education (MNE), which has undertaken a task to ensure that young people who have been classified according to vocational skills such as wood technologies, metal, electricity, computer and automotive have the opportunity to study in the direction of their interests and abilities, is far from satisfying needs of the students in verbal and numerical lessons [5-7]. Science courses (physics, chemistry and biology) are among the numerical courses and considered to be very important for vocational high school students to be better prepared for the profession, to increase their working life achievements and to grow in the quality that contemporary technologies desire. It is vital to prevent possible failures in these courses. Regression based methods are techniques used generally in the literature to determine variables that are considered to influence achievements of the students and the effect of these variables. Nevertheless, one of the methods used successfully in every field where data mining is carried out, no matter it is science, health sciences or social sciences, is "Artificial Neural Networks [8]. ANN method aims clustering of given examples and classifying them into specific classes as well as deciding the classes that the following samples would join [9]. ANN's modeling of the human brain emerged with the desire to create machines that function like human brain, which also exists in the modeling of digital computers [10]. ANN is computer systems that perform the learning function, which is the most basic feature of the human brain. A model is created by ANN in the direction of inputs and outputs. Inputs are information received from outside or other nerve cells for an artificial neural cell. These inputs consist of the samples for which learning of the nerve cell is desired [11, 12]. Learning in artificial neural networks takes place by changing the weights between the nerves. Accordingly, nerves whose weights on the inter-nervous links are changeable dynamically with a specific

method can be trained. Networks that can learn may recognize new forms or decide the class the given entry belongs to. Although there is a limited number of applications in the field of education regarding ANN today, this method is successfully used in areas such as manufacturing, control, transportation and aviation, medicine, engineering, meteorology, biomedical and pharmaceutical industry, finance, stock market and credit card applications. It is seen that ANN and regression analyzes are used together in the education field especially in the academic achievement estimations. One of the earliest studies in the field of education in which ANN was used in predicting the academic achievements of the students was the work of Gorr, Nagin and Szczypula [13] that compares multi-linear regression and stochastic linear regression analyses with artificial neural network analysis in predicting students' weighted grade averages. Subbanarasimha, Arinzeb & Anandarajanb [14] used two different sets of data to compare artificial neural networks and regression techniques in predicting the academic performance of the students. As a result of the study, it is revealed that the accuracy of the analysis using artificial neural networks is higher. Several researches compared artificial neural networks, decision trees and linear regression methods in predicting students' achievements. As a result of the research, ANN indicated better results than the others. Oladokun, Adebajo and Charles-Owaba [15] aimed to determine the variables that influenced students' achievements and to test the analysis of artificial neural networks in predicting students' performances. At the end of the study, it was concluded that ANN is highly predictive and effective in determining variables that affect the students' achievements. Çırak and Çokluk [16] found in their study in which ANN and logistic regression methods were used in classifying students' achievements in the higher education that the total correct classification rate is 66.10% with the logistic regression analysis and 70.10% with ANN analysis for the proposed model. As seen in the literature, there is a rapid increase in the studies in which ANN is used, especially in the academic achievement predictions in recent years.

1.1 The rationale and aim of the study

When the literature is viewed, it is seen that many methods and techniques were used in predicting academic achievements of students and teacher candidates who are studying at different educational institutions. However there seems to be no study indicating prospective academic achievement prediction of science course of students in VTHS using the ANN analysis, which is the most reliable prediction mechanism of information age and the program with the closest functional feature to the human brain. Furthermore, there seems to be no study determining the factors affecting the academic achievements of these students and necessary measures to prevent their failures. In this context, the research questions of this research are:

1. To what extent does the model emerged with the ANN analysis in the direction of parameters that are considered to affect the academic achievements of VTHS students in physics courses classify students with their achievements?
2. To what extent does the model emerged with the ANN analysis in the direction of parameters that are considered to affect the academic achievements of VTHS students in chemistry courses classify students with their achievements?
3. To what extent does the model emerged with the ANN analysis in the direction of parameters that are considered to affect the academic achievements of VTHS students in biology courses classify students with their achievements?
4. What are the most important factors affecting the academic achievements of VTHS students in science (physics, chemistry and biology) courses and the measures to be taken in order to prevent their failures?

2 Methodology

2.1 The design and study group of the study

The study was structured in the survey model based on the quantitative research model. Survey models are research approaches that aim to describe the past or present instance as it exists. The event, the individual or object that is the subject of the investigation is tried to be defined within their own circumstances [17]. The study group of the research consisted of 922 students studying in 10th and 11th grades at VTHS in Turkey. Face-to-face meetings were held with the students.

2.2 The data collection instrument

A pool of questions was formed in this study conducted in the survey model for the collection of data. A questionnaire consisting of 40 items selected by the researchers from this pool was developed. While the questionnaire used in the study was being developed, 6 items removed as they were not appropriate with language and measurement criteria according to opinions of 3 science field specialists, 1 measurement and assessment specialist and 2 technical education experts. Eventually, a 34-item questionnaire was created. The content validity of the questionnaire was provided in this way. To perform item factor analysis of the data obtained with the pilot study (250 participants) in the construct validity phase of the questionnaire, the KMO value was calculated as .86, the barlett test value was calculated as .00 and chi-square value was calculated as 2.183, indicating item analysis can be carried out. Item 9, 29, 33 and 34, whose item loads are below .30, were removed from the questionnaire for the item factor analysis. The KR20 reliability coefficient was calculated as .90 in order to determine the reliability. The final questionnaire consisted of 30 items and 3 sub-dimensions. The parameters that are considered to affect academic achievement were grouped under the following headings: 1. Student Features, 2. Teacher-School Features, 3. Parent

2.3 Data analysis

ANN analysis was used in the statistical analysis process. Grade point averages of the students at the end of the academic year were determined as output variables for ANN in the research. A discrete variable with two categories that are unsuccessful (0) for students with a grade average (0-2) and successful (1) for students with a grade average (3-5) was formed. The items in the questionnaire that are considered to affect academic achievement were approved as inputs. The prediction of the academic achievements of the students was realized with ANN in the Matlab R2016a program using the inputs and outputs and a model was created. Bayesian Regularization (trainbr) algorithm was used in training the feed-forward back-propagation network. A three-layer feed-forward network that is input, hidden, and output layer was used in the study. There are 70 neurons in the input layer, 30 neurons in the hidden layer, and 1 neuron in the output layer in the academic achievement prediction model of each course.

3 Result

In this section, the findings and interpretations obtained by analyzing the data for the purpose of the data are discussed.

3.1. Findings on the prediction of students' achievements in science course with the ANN analysis

80% (738 items) of the data set was used in training, 5% (46 items) of the data set was used in validation, and 15% (138 items) of the data set was used in testing. The ANN used

in the study has 70 inputs and 1 output with 30 layers. Findings obtained by analysis of artificial neural network to classify students according to their academic achievements are presented in Table 1.

Table 1. Classification Table obtained with ANN Model for Physics Course

ANN Steps	Actual Situation	Estimated		Accurate Estimation Percentage (%)
		Successful	Unsuccessful	
Training	Successful	363	1	99.7
	Unsuccessful	367	2	99.5
	Total	%49	%51	99.6
Validation	Successful	16	1	94.1
	Unsuccessful	29	0	100
	Total	%37	%63	97.8
Test	Successful	47	16	74.6
	Unsuccessful	64	10	86.5
	Total	%43	%47	81

As seen in Table 1, 363 successful students were correctly classified, 1 was misclassified into 364 from 738 individuals participating in total for physics course and the accurate classification rate is 99.7%. 367 students were correctly classified, 2 were misclassified into 369 unsuccessful students and the accurate classification rate is 99.5%. The total accurate classification rate is 99.6%. 16 of the 17 successful students of 46 individuals who participated in the validation analysis were correctly classified, 1 was misclassified and the accurate classification rate is 94.1 %. 29 of the 29 unsuccessful students were correctly classified, 0 was misclassified and the accurate classification rate is 100%. The total accurate classification rate is %97.8. 47 of the 63 successful students of 138 individuals participating in the test were correctly classified, 16 were misclassified and the accurate classification rate is %74.6. 64 of the 74 unsuccessful students were correctly classified, 10 were misclassified and the accurate classification rate is 86.5%. The total accurate classification rate is 81%.

3.2. Findings on the prediction of students' achievements in chemistry course with the ANN analysis

Findings obtained by analysis of artificial neural network to classify students according to their academic achievements in the chemistry course are presented in Table 2.

Table 2. Classification Table obtained with ANN Model for Chemistry Course

ANN Steps	Actual Situation	Estimated		Accurate Estimation Percentage (%)
		Successful	Unsuccessful	
Training	Successful	423	0	100
	Unsuccessful	309	0	100
	Total	%58	%42	100
Validation	Successful	27	0	100
	Unsuccessful	19	0	100
	Total	%59	%41	100
Test	Successful	83	5	94.3
	Unsuccessful	45	5	90
	Total	%60	%32.8	92.8

As seen in Table 2, 423 of the 423 successful students of 738 individuals participating in total in the chemistry course were correctly classified, 0 was misclassified and the accurate classification rate is 100%. 309 of the 309 unsuccessful students were correctly classified, 0 was misclassified and the accurate classification rate is 100%. The total accurate classification rate is 100%. 19 of the 19 successful students of 46 individuals

participating in the validation analysis were correctly classified, 0 was misclassified and the accurate classification rate is 100%. The total accurate classification rate is 100%. 83 of the 88 successful students of 138 individuals participating in the test analysis were correctly classified, 5 were misclassified and the accurate classification rate is 94.3%. 45 of the 50 unsuccessful students were correctly classified and 5 were misclassified and the accurate classification rate 90%. The total accurate classification rate is %92.8

3.3. Findings on the prediction of students achievements in biology course with the ANN analysis

Findings obtained by analysis of artificial neural network to classify students according to their academic achievements in the biology course are presented in Table 3.

Table 3. Classification Table obtained with ANN Model for Biology Course

ANN Steps	Actual Situation	Estimated		Accurate Estimation Percentage (%)
		Successful	Unsuccessful	
Training	Successful	425	3	99.3
	Unsuccessful	305	1	99.7
	Total	%58	%42	99.5
Validation	Successful	30	1	96.8
	Unsuccessful	15	0	100
	Total	%67	%33	97.8
Test	Successful	82	0	100
	Unsuccessful	51	3	94.4
	Total	%62	%38	97.8

As seen in Table 3, 425 of the 428 successful students of 738 individuals participating in total in the biology course were correctly classified, 3 were misclassified and the accurate classification rate is 99.3%. 305 of the 306 unsuccessful students were correctly classified, 1 was misclassified and the accurate classification rate is 99.7%. The total accurate classification rate is 99.5%. 30 of the 31 successful students of 46 individuals participating in the validation analysis were correctly classified, 1 was misclassified and the accurate classification rate is 96.8%. 15 of the 15 unsuccessful students were correctly classified, 0 was misclassified and the accurate classification rate is 97.8%. 82 of the 82 successful students of 138 individuals participating in the test analysis were correctly classified, 0 was misclassified and the accurate classification rate is 100%. 51 of the 54 unsuccessful students were correctly classified, 3 were misclassified and the accurate classification rate is 94.4%. The total accurate classification rate is 97.8%.

Table 4. Classification Table obtained with ANN Model for Study-wide

Courses	Actual Situation	Estimated	Accurate Estimation Percentage (%)
Physics	Successful	426	95.9
	Unsuccessful	460	97.5
	Total		96
Chemistry	Successful	533	99.1
	Unsuccessful	373	98.7
	Total		98.9
Biology	Successful	437	99.3
	Unsuccessful	371	98.9
	Total		99.1

As seen in Table 4, the total accurate estimation percentage is 96% for physics, 98.9% for chemistry and 99.1% for biology in the study-wide.

3.4. The most important factors affecting the academic achievements of VTHS students in science (physics, chemistry and biology) courses and the measure to be taken in order to prevent their failures

The most important factors with significance at $p < .01$ affecting the academic achievements of VTHS students in physics, chemistry and biology courses are given in Table 5.

Table 5. The Most Significant Factors of Academic Achievement with ANN

Courses	Order of Importance	Factors	Significance	Normalized Significance Percentage (%)
Physics	1	Class size	.030	98.2
	2	Use of different method in the course	.029	95.2
	3	Participating in activities	.029	95
	4	Having a study room	.024	79.7
	5	The economic situation of the family	.023	74.2
Chemistry	1	Class size	.032	100
	2	The father being literate	.028	88.9
	3	The father having an occupation	.026	82.3
	4	Regular studying situation	.024	74.5
Biology	1	Enthusiasm for Biology	.032	100
	2	Regular preparation for exams	.029	92.4
	3	Coexistence of parents	.026	83.3
	4	Liking the biology teacher	.026	82.9
	5	Class size	.026	82.1

As seen in Table 5, class size (%98.2) is forefront in the affecting factors in academic achievements in the physics course. This is followed by different teaching methods (95.2%), participating in activities (%95%), having a study room (79.7%) and the economic situation of the family (74.2%), respectively. One of the most important factors affecting academic achievements in the chemistry course is class size (100%). In addition, factors such as father's literacy status (88.9%), father's occupation status (82.3%) and regular studying (74.5%) are among the most important factors affecting academic achievements in the chemistry course. The most important factors affecting the academic achievements of the students in the biology course are: enthusiasm for the biology course (100%), regular preparation for exams (92.4%), coexistence of parents (83.3%), liking biology teacher (82.9%), and class size (82.1%). It can be considered that eliminating the negative impacts of these factors can increase academic achievements of the students.

4 Conclusion and discussion

As a result of the study, a model that can predict the academic achievements of the students with ANN was created. When it is desired to predict the future academic achievements of a new student included in the model, with the emerged ANN model in science courses, the probability of being predicted correctly is 96% for physics, 98.9% for chemistry and 99.1% for biology. The multilayered model created by ANN correctly classifies 95.9% of the successful students and 97.5% of the unsuccessful students for the physics course. This suggests that the performance of predicting students who failed in physics is higher than the performance of predicting successful students in this course and therefore the model created by ANN can indicate better results especially in predicting unsuccessful students. The multilayered model created by ANN correctly classifies 99.1% of the successful students

and 98.7% of the unsuccessful students in the chemistry course. In this context, this suggests that the performance of predicting successful students in the chemistry course is higher than the performance of predicting the students who failed in this course and therefore the model created by ANN can indicate better results especially in predicting successful students. For the biology course, the multilayered model created by ANN correctly classifies 99.3% of the successful students and 98.9% of the unsuccessful students. According to this finding, the performance of predicting the students who succeeded in the biology course is higher than the performance of predicting the students who failed in this course and therefore the model created by ANN can indicate better results in predicting successful students. It is possible to pay attention to factors such as setting the class size in the pedagogical sense, delivering the lesson with different teaching methods and techniques, increasing participation in activities, the fact that the student has a study room and the fact that the economic situation of the family is well in order to influence academic achievements of the students in physics positively. When considering the factors that affect the academic achievements of the students in the chemistry course, it is possible to interpret that appropriate class size, the father being literate and having an occupation, and regular studying situation of the student affect the academic achievement positively. Furthermore, it is possible to consider that having enthusiasm for the biology courses, regular preparation for exams, coexistence of parents, liking the biology teacher, and appropriate class size are influential on the positive direction of the academic achievements in the biology course. One of the strengths of the model created by ANN is high accuracy of the prediction of academic achievements of the students. However, this research is a specific implementation since the study was carried out with a certain number of data obtained from VTHS students. For this reason, it is useful to evaluate the results obtained from the study, taking into account the limitations. It is considered that comparing predictions by using logistic regression analysis or other techniques together with the ANN technique carried out in the study can be beneficial to reach more accurate predictions.

This study forms a part of a thesis study titled « Prediction of Academic Achievements of Vocational and Technical High School Students with Artificial Neural Networks in Science Courses (Physics – Chemistry – Biology) and Measures to be Taken for Failure (Turkey- Malaysia Comparison) » conducted by Ali YAĞCI.

References

1. Erol, N. Mesleki Ortaöğretimde Yeni Müfredat Doğrultusunda Öğretmenlerin Eğitim İhtiyaçlarının Belirlenmesi. Yayınlanmamış Yüksek Lisans Tezi. Gazi Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara (2010)
2. Aycan, Z. Balcı, H. Hizmet İçi Eğitimin Etkililiğini Yordayan Bireysel ve Kurumsal Faktörler. Türk Psikolojisi Dergisi, **16**, 48, 15 (2001)
3. Eker, G. Endüstri Meslek Lisesi Öğrencilerinin Tükenmişlik Düzeyi. Yüksek Lisans Tezi. İstanbul: Marmara Üniversitesi Fen Bilimleri Enstitüsü (2007)
4. Çevik, M. (2014). Mevcut Biyoloji Öğretim Programının Mesleki ve Teknik Liselerde Görevli Yönetici, Öğretmen ve Öğrenci Görüşlerine Göre Değerlendirilmesi ve Yeni Bir Taslak Program Önerisi (Fotosentez Konusu Örneği). Yayınlanmış Doktora Tezi. Gazi Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
5. Binici, H., Arı, N. Mesleki ve teknik eğitimde arayışlar. Gazi Üniversitesi Eğitim Fakültesi Dergisi, **24**, 3, (2004)
6. Şahin, İ. Fındık, T. Türkiye’de mesleki ve teknik eğitim: Mevcut durum, sorunlar ve çözüm önerileri. *TSA/Yıl:12*, **3**, 65-86 (2008)

7. Yörük, S. Dikici, A., Uysal, A. Bilgi toplumu ve Türkiye’de mesleki eğitim. *Fırat Üniversitesi Sosyal Bilimler Dergisi*, **24**, 3, 229-312 (2002)
8. Han, J., Kamber, M. Data Mining, Southeast Asia Edition: Concepts and Techniques. Morgan Kaufmann, **81** (2006)
9. Öztemel, E. *Yapay sinir ağları*. (1.Basım). İstanbul: Papatya Yayıncılık (2003)
10. Tosun, Ö. Sınıflandırmada yapay sinir ağları ve karar ağaçları karşılaştırması: öğrenci başarıları üzerine bir uygulama, Yüksek Lisans Tezi, İstanbul teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul (2007)
11. Yurtoğlu, H. Yapay sinir ağları metodolojisi ile öngörü modellemesi: bazı makroekonomik değişkenler için türkiye örneği, Doktora Tezi, Ankara (2005)
12. Baş, N. Yapay Sinir ağları Yaklaşımı ve Bir Uygulama. Yayımlanmamış Yüksek Lisans Tezi. Mimar Sinan Güzel Sanatlar Üniversitesi Fen Bilimleri Enstitüsü. İstanbul (2006)
13. Gorr, W. L., Nagin, D., Szczygula, J. Comparative study of Artificial Neural Network and Statistical Models for Predicting Student Grade Point Averages. *International Journal of Forecasting*, **10**, 17-34 (1994).
14. Subbanarasimha, P. N., Arinzeb, B., Anandarajanb, M. 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**, 117-123 (2000)
15. Oladokun, V. O., Adebajo, A. T., Charles-Owaba, O. E. Predicting Students’ Academic Performance using Artificial Neural Network: A Case Study of an Engineering Course. *The Pacific Journal of Science and Technology*, **9**, 1 (2008)
16. Çırak, G., Çokluk Ö. Yükseköğretimde öğrenci başarılarının sınıflandırılmasında yapay sinir ağları ve lojistik regresyon yöntemlerinin kullanılması. *Mediterranean Journal of Humanities*, **3**, 2, 71-79 (2013)
17. Karasar, N. Bilimsel araştırma yöntemi. Ankara: Nobel Yayın Dağıtım (2005)
18. Güneri, N., Apaydın, A. Öğrenci Başarılarının Sınıflandırılmasında Lojistik Regresyon Analizi ve Sinir Ağları Yaklaşımı. *Ticaret ve Turizm Eğitim Fakültesi Dergisi*, **1**, 170-188 (2004)
19. İbrahim, Z., Rusli, D. “Predicting Students’ Academic Performance: Comparing Artificial Neural Network, Decision Tree and Linear Regression. *21st Annual SAS Malaysia Forum*, Shangri-La Hotel, Kuala Lumpur, September, 5 (2007)
20. MEB. Matematik ve meslek matematiği dersi öğretim programı. Milli Eğitim Bakanlığı Çıracılık ve Yaygın Eğitim Genel Müdürlüğü, Ankara (2005)