Research Article.
Journal of Intelligent Systems: Theory and Applications 5(1) 2022: 9-15
DOI: 10.38016/jista.922663

Nos resultados, o modelo de aprendizado de máquina foi superior à regressão beta na previsão da taxa de sucesso dos alunos. Como proposta, podemos utilizar dos métodos ensemble na regressão beta, no mesmo conjunto de dados, para ver se superamos a capacidade preditiva.

Desafio: Obter a base de dados. [Dionisio Neto]

Estimation of High School Entrance Examination Success Rates Using Machine Learning and Beta Regression Models

Tuba Koç[1], Pelin Akın[2*]

[1,2] Çankırı Karatekin University, Faculty of Science, Department of Statistic, Çankırı/Turkey

tubakoc@karatekin.edu.tr , pelinakin@karatekin.edu.tr

Abstract

Education is the foundation of economic, social, and cultural development for every individual and society as a whole. Students are accepted to secondary education institutions with the high school entrance examination made by the Ministry of National Education in Turkey. In this study, the success rates of the students who took the high school entrance examination in Turkey's 81 provinces in 2019 were handled with the machine learning regression and beta regression model. The present paper aimed to model, predict, and explain students' success rates using variables such as divorce rate, gross domestic product, illiteracy, and higher education populations. Support vector regression, random forest, decision tree, and beta regression model were applied to estimate success rates. Two models with the highest R^2 value were found to be beta regression and random forest models. When the prediction errors of beta regression and random forest model were examined, it seemed to be that the random forest model is relatively superior to the beta regression model in predicting the success rates. While the beta regression model was the best predictor of the success rates of Çanakkale province, the random forest model predicted the success rates of Ankara well. Also, it was seen that the variables found to be significant in the beta regression model for success rates were also crucial in the random forest model. It is recommended to use both the beta and random forest models to estimate the students' success rates.

Keywords: Success rate of exam, Beta regression, Random forest, Classification and regression tree, Support vector regression

Makine Öğrenimi ve Beta Regresyon Modelleri Kullanılarak Lise Giriş Sınavı Başarı Oranlarının Tahmini

Öz

Eğitim, her birey ve bir bütün olarak toplum için ekonomik, sosyal ve kültürel gelişimin temelidir. Ülkemizde orta öğretim kurumlarına Milli Eğitim Bakanlığı tarafından yapılan lise giriş sınavı ile öğrenci kabul edilmektedir. Bu çalışmada, 2019 yılında Türkiye'nin 81 ilinde lise giriş sınavına giren öğrencilerin başarı oranları makine öğrenimi regresyon ve beta regresyon modeli ile ele alınmıştır. Bu makale, boşanma oranı, gayri safi yurtiçi hasıla, okuma yazma bilmeyenlerin sayısı ve yüksek öğretim nüfusu sayısı gibi değişkenleri kullanarak öğrencilerin başarı oranlarını modellemeyi, tahmin etmeyi ve açıklamayı amaçlamaktadır. Başarı oranlarını tahmin etmek için destek vektörü regresyon, rastgele orman, karar ağacı ve beta regresyon modeli uygulanmıştır. En yüksek R^2 değerine sahip iki modelin beta regresyon ve rastgele orman modelleri olduğu bulunmuştur. Beta regresyon ve rastgele orman modelinin tahmin hataları incelendiğinde, başarı oranlarını tahmin etmede rastgele orman modelinin beta regresyon modeline göre nispeten üstün olduğu görülmektedir. Beta regresyon modeli Çanakkale ilinin başarı oranlarının en iyi yordayıcısı iken, rastgele orman modeli Ankara'nın başarı oranlarını iyi tahmin etmiştir. Ayrıca beta regresyon modelinde başarı oranları için anlamlı bulunan değişkenlerin rastgele orman modelinde de önemli olduğu görülmüştür. Öğrencilerin başarı oranlarını tahmin etmek için hem beta hem de rastgele orman modellerinin kullanılması önerilir.

Anahtar kelimeler: Sınav başarı oranı, Beta regresyon, Rastgele orman, Sınıflandırma ve regresyon ağacı, Destek vektör regresyonu.

* Corresponding author.
E-mail: pelinakin@karatekin.edu.tr

Recieved : 20 April 2021
Revision : 30 June 2021
Accepted : 22 October 2021

1. Introduction

Education is one of the influential factors that provide social, cultural, and economic development. The economic development of a country and the progress of society take place with qualified workforce. Realizing this is one of the basic functions of education. Increasing the level of education is achieved by making that country more knowledgeable, skilled, and equipped. The education systems of the countries bring some political implications and obligations. Developments in science and technology cause changes in the needs of individuals. It is possible to train a qualified workforce that can adapt to the speed of developing technology as behaviour by innovating countries' education systems. Through innovations in the educational systems, countries can raise qualified human power that can adapt their behaviour to technological developments (MEB, 2018). For this purpose, education programs are prepared to raise qualified individuals who can solve problems, think critically, entrepreneurs, and contribute to society (MEB, 2018). For that purpose, curriculums are developed to raise individuals who can solve problems, think critically, and actively contribute to society. Factors affecting success in education have always been among the issues that are emphasized. At the stage of achieving success, the factors affecting this should be at a level that will create success at the maximum level to be good.

Measurement and evaluation are crucial in education systems as in other systems. Different countries use different variables in student selection for transition to high schools. It is seen that these variables are school graduation exams, central selection exams, school-based selection exams, school grades, and teachers' opinions (Gür et al., 2013). In Turkey, central selection exams are held in student selection for transition to high schools.

In recent years, using machine learning algorithms in the field of education is a general approach. Abbasoğlu (2020) analysed the effects of middle school students' demographic characteristics and socioeconomic status on their year-end general achievement averages using data mining methods. Gök (2017) estimated the end-of-term achievement averages of secondary school students using logistic regression and multi-class machine learning models. According to the results, both logistic regression and classification methods successfully estimate the average success score. Uskov et al. (2019) examined the machine learning predictions of student academic performance in STEM (Science, Technology, Engineering, and Mathematics) education. Abidi et al. (2019) investigated models for predicting confused students who try to do homework using ITS (Intelligent tutoring systems). In their studies, they used naïve Bayes (NB), generalized linear model (GLM), logistic regression (LR), deep learning (DL), decision tree (DT), random forest (RF), and gradient boosted trees (XGBoost) machine learning models. As a result, they showed that the RF, GLM, XGBoost and DL models achieved a high accuracy in predicting students' confusion in the algebra mastery skills in ITS. Using machine learning algorithms, Al Mayahi and Al-Bahri (2020) predicted whether university students would pass a particular course based on previous academic achievement data. In the study, the accuracy rate of the model was found to be 87%. Sethi et al. (2019) used three different machine learning algorithms in the subject/stream selection of middle school students. Rebai et al. (2020) investigated the success of secondary school education in Tunisia with a two-level algorithm. The study used decision trees and random forest algorithms to provide input for the data envelopment analysis (DEA) method. Rajak et al. (2020) applied different classification machine learning algorithms to data sets with characteristics such as family education, father's job, school attendance and calculated the model's performance. Yousafzai et al. (2020) used machine learning and data mining methods to predict students' performance at the secondary education level.

This study aims to show the use of machine learning in the education area. Machine learning algorithms and beta regression models were applied and compared to calculate the success rates of students in the high school entrance examination. The article is divided as follows: In section 2, the beta regression and machine learning algorithms are defined. Section 3 explains the application of the beta regression model and machine learning algorithms with success rate data. Finally, a brief discussion is given in Section 4.

2. Material and Methods

In this section, beta regression and machine learning methods used to estimate the success rates of high school entrance examinations are presented

2.1. Beta Regression Model

The beta regression model is widely used to model variables in the range (0, 1). This model is very flexible and can be used for random variables such as ratio and percentage (Ferrari and Cribari-Neto, 2004). It is commonly used in fields such as education, finance, and social sciences. (Cepeda-Cuervo, 2015).

The beta density is given by

$$f(y; p, q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} y^{p-1}(1-y)^{q-1}, \quad 0 < y < 1 \quad (1)$$

Where Γ (.) is the gamma function p, q>0.

Ferrari and Cribari-Neto (2004) proposed a different parameterization with $\mu = \frac{p}{p+q}$ and $\phi = p + q$:

$$f(y; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\varphi)\Gamma\big((1-\mu)\varphi\big)} y^{\mu\phi-1}(1-y)^{(1-\mu)\varphi-1}$$

$$,0<y<1 \quad , 0 < \mu < 1 \text{ and } \phi > 0 \quad (2)$$

$$y \sim B(\mu, \phi) \quad \text{and} \quad E(y) = \mu, \mathrm{Var}(y) = \frac{\mu(1-\mu)}{1+\phi}$$

μ is the mean of the response variable, and ϕ can be interpreted as a dispersion parameter for fixed μ.

Let $y_1, y_2, \dots, y_n$ be independent random variables each $y_t \sim B(\mu_t, \phi_t),\ t = 1, \dots, n$

$$g(\mu_t) = \sum_{i=1}^{k} x_{ti}\beta_i \tag{3}$$

Where $\beta = (\beta_1, \beta_2, \dots\dots\dots, \beta_k)^T$ represent unknown regression parameters. x= (x_{t1},x_{t2},…,x_{tk}) denotes fixed covariates g(.) shows the link function, which strictly monotonic and twice differentiable. The Beta regression model can use different link functions such as log, logit, etc. (Dünder and Cengiz, 2020).

2.2. Machine Learning Regression Models

This study used support vector regression, decision tree, and random forest regression from machine learning regression models.

a) Support Vector Regression

Support vector machines, which have been suggested to resolve classification and regression problems, are supervised learning techniques based on statistical learning theory and the principle of structural risk minimization (Vapnik, 1992). Consider the problem of the set of training data

D= $\{(x^1, y^1), \dots, (x^l, y^l)\}$ with a linear function,

$$y = \langle w, x \rangle + b \tag{4}$$

The optimum regression function is provided from the minimum of the function,

Minimize

$$\frac{1}{2}\|w\|^2 + c\sum(\xi_i + \xi_i^*) \tag{5}$$

Constraints

$$\begin{aligned} &y_i - (wx_i) - b \le \varepsilon + \xi_i \\ &wx_i + b - y_i \le \varepsilon + \xi_i^* \\ &\xi_i, \xi_i^* \ge 0 \end{aligned} \tag{6}$$

where (*) symbolizes both the vector with and without asterisks. ξ_i, ξ_i^* slack variable and c > 0 is a penalty parameter (Gunn, 1998). The constrained optimization problem is then reworded as a dual problem using Lagrange multipliers a_i, a_i^* for each constraint. Lagrange multipliers are determined by solving the issue with quadratic programming. After a_i, a_i^* are determined, the optimal weights w and the base b can be calculated, and the final predictor is given in the equation (Shokry et al., 2015).

$$y = \sum_i^m (a_i^* - a_i)(x_i - x) + b \tag{7}$$

b) Decision Tree

Decision tree algorithms are among the most preferred machine learning techniques because they are easy to interpret, detect errors, and apply easily (Kotsiantis, 2011). Breiman et al. (1998) proposed classification and regression tree (CART). The R function recursive partitioning (RPART) is an application of the CART. The RPART programs construct classification or regression models of a very general structure using a two-step process; the resulting models may represent binary trees (Therneau and Atkinson, 1997). The mean squared error is used for the split data in the RPART algorithm. MSE for a specific node is defined as;

$$\mathrm{MSE}_{node} = \frac{1}{m_{node}}\sum(y_i - \bar{y}_{node})^2 \tag{8}$$

If it is assumed that a binary split on each node on the tree will be divided into left and right. For each division,

$$\mathrm{MSE}_{left} = \frac{1}{m_{left}}\sum(y_i - \bar{y}_{left})^2 \tag{9}$$

$$\mathrm{MSE}_{right} = \frac{1}{m_{right}}\sum(y_i - \bar{y}_{right})^2 \tag{10}$$

For each attribute j, the following formula is calculated,

$$\min(\mathrm{MSE}_{left} + \mathrm{MSE}_{right}) \tag{11}$$

The smallest of the values is chosen. The splits the dataset recursively, which means that the subsets that meet a partition are partitioned until they reach a predetermined expiration criterion (Therneau and Atkinson, 1997).

c) Random Forest Regression

The Random Forests algorithm is one of the ensemble learning algorithms. The ensemble learning algorithms produce a prediction model by combining the strong points of a simpler group of fundamental models (Friedman and Sandow, 2011). The most widely used ensemble learning algorithms are bagging and random forest algorithms. Breiman's random forest classification is an improved version of the bagging technique by adding the randomness feature. The following steps are taken for the random forest algorithm:

i) Draw n bootstrap samples from the original dataset.

ii) For each of the bootstrap samples, grow an unpruned classification or regression tree (CART) is created.

iii) In random classification, two parameters are used, namely the number of variables used in each node (m) and the number of trees to be developed (N) to determine how best to split. A new estimate is made by combining the estimates made by the N number of trees separately. While the class with the majority votes in classification trees is chosen as the final estimate, for regression trees, estimation is made by taking the average of the average votes (Liaw and Wiener, 2002).

2.3. Evaluation Metrics for Regression models

Commonly used metrics to evaluate forecast accuracy are the mean absolute error (MAE), the mean squared error (MSE), the root mean squared error (RMSE), and the coefficient of determination (R^2) (Uğuz, 2019). R^2 is used to measure the wellness of the fit by the trained models. A high R^2 value indicates that the prediction relationship is good. MAE, MSE, RMSE are the average error measure, so low values indicate good performance. The error measures are defined as follows

$$\text{MAE} = \frac{1}{N}(\sum|y_i - \hat{y}|)$$
$$\text{MSE}=\frac{1}{N}\sum(y_i - \hat{y}))$$
$$\text{RMSE}=\sqrt{\frac{1}{N}\sum(y_i - \hat{y})}$$
$$R^2 = 1 - \frac{\sum(y_i-\hat{y})^2}{\sum(y_i-\bar{y})^2} \quad (12)$$

3.Results

In this study, the high school entrance examination data's success rate was used for Turkey's 81 provinces for (in) 2019. The data obtained are available in URL1-2-3. The features of the variables are given in Table 1.

Table 1. Description of the variables

Variable	Description
y (response variable)	Success rate
x_1	Divorce rate
x_2	Gross domestic product (GDP-per city)
x_3	The number of illiteracy
x_4	Number of higher education population
x_5	Households Internet Access rate
x_6	Number of a theatre child audience
x_7	Book reading rate

Table 1 shows the description of the response variable and explanatory variables. The success rate was obtained as the number of questions answered correctly to the total number of questions by students who took the high school entrance examination in 81 provinces. The explanatory variables were chosen among several indicators that may have a potential influence on the success rate.

Table 2. Coefficients for the Beta regression model with Cauchit link

Coefficient	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.0364	0.0884	0.4120	0.6803
x1	0.2084	0.0593	3.5140	0.0004
x2	0.6004	0.2981	2.0140	0.0440
x3	-0.0769	0.3412	-0.2250	0.8217
x4	0.5315	0.3999	1.3290	0.1839
x5	-0.5964	0.1118	-5.3370	9.44e-08
x6	0.7310	0.2877	-2.5410	0.0110
x7	0.2666	0.1394	1.9120	0.0500

Beta regression model was applied to success rates. Choosing the appropriate link function in the beta regression model can significantly improve the model (Koç, 2019). In the beta regression model, the information criteria of different link functions were examined, and the most suitable link function was the Cauchit link. When the beta regression model is applied to the data by selecting the Cauchit link function, the parameter estimates are given in Table 2. According to the results in Table 2, it is seen that the variables x1, x2, x6, and x7 affect the success rate positively and the variable x5 negatively. The results are consistent with the literature (Oral and McGivney, 2014; Yavuz, 2020). Beta regression estimated model is given by

$$\hat{y} = 0.03641 + 0.20837.x1 + 0.60036.x2 - 0.59641.x5 + 0.73103.x6 + 0.26662.x7$$

One of the methods used to prepare data for analysis is normalization. The purpose of normalization is to change the values of numeric columns in the data set to use a standard scale without breaking the differences in value ranges or losing information (Han et al., 2005). After normalization, we separated data to 90% of the data going to training and the remaining 10% to test. Shortly, randomly selected data for 73 cities in the data set are train, and eight cities are test data.

Machine learning algorithms and beta regression model was applied to train data to calculate success rates of test data. The analyses were performed using R software's 3.5.2 version.

Firstly, the support vector was applied. When applying support vector machine algorithm, it is essential to determine cost and error parameters. For this reason, the model with the best performance was selected after trying 1100 models in the range of error (0,1) and cost (1,100). The best model was obtained at error 0.2 and cost 2. Then CART, random forest, and beta regression model were performed to train data.

We compared the prediction capabilities of the machine learning algorithm and beta regression model on the test data.

Model validation for the machine learning was performed on the test data and the cities randomly

selected for the test data are extracted out too for the beta model. Performance measurements are given in Table 3.

Table 3. Performance measurement for models

Performance measurements	Beta Regression	Support Regression	Random Forest Regression	CART
MAE	0.0053	0.0603	0.0481	0.0435
MSE	0.0040	0.0055	0.0038	0.0043
RMSE	0.0636	0.0741	0.0617	0.0657
R^2	0.5909	0.5400	0.6638	0.5613

Table 3 shows that the random forest regression algorithm is the best model in machine learning algorithms. Table 4 shows a comparison between forecast success rates and actual success rates.

Table 4. Success rate prediction with models

Cities	Success rate	Beta_predicted	Beta_error	RF_predicted	RF_error
Afyonkarahisar	0.6244	0.6776	0.0532	0.6898	0.0654
Adıyaman	0.6211	0.5143	0.1068	0.5307	0.0904
Ankara	0.7289	0.7799	0.0510	0.7238	0.0051
Bayburt	0.6633	0.6121	0.0512	0.5620	0.1013
Çanakkale	0.7533	0.7462	0.0071	0.7341	0.0192
Kahramanmaraş	0.6078	0.6244	0.0166	0.6192	0.0114
Van	0.4744	0.5088	0.0344	0.4804	0.0060
Hakkari	0.3800	0.5012	0.1212	0.4405	0.0605

When the estimation errors of beta regression and random forest models are examined in Table 4, the random forest model gives less error for five cities. The RF model seems to be relatively superior to the beta regression model in predicting success rates.

In the beta regression model, x1, x2, x5, x6, and x7 variables were found to have a significant contribution to the model. When a random forest algorithm is applied to data, the order of importance of variables is given in Figure 1.

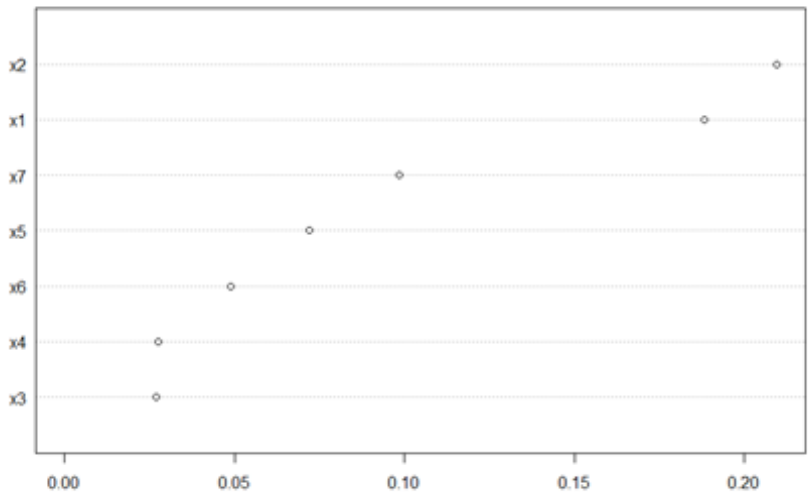

Figure 1. Random forest features importance

When we look at Figure 1, it is seen that the coefficients that are significant for the beta regression model are also important in the RF model. According to the RF model, the three most important variables were x2, x1, and x7.

4. Conclusion and Discussion

Education is one of the most significant elements that shape the future of society. Nowadays, the application of machine learning, which has successful applications in many fields, is a very favoured approach in education.

In this study, firstly, the beta regression model was used in 81 provinces of Turkey to determine the factors affecting the students who took the exam in 2019. According to the beta regression model, it is seen that the variables of divorce rate, GDP, household's internet access rate, number of theatre child audience, and book reading rate affect success rates. These results are very consistent with the literature (Özeren et al. 2020; Çömlekcioğulları, 2020). Then the data set was separated into train (73 cities) and test (8 cities). SVR, RF, CART, and beta regression models were applied to train data to calculate the success rates of test data. The two models with the highest R^2 values were R^2=0.5909 for beta regression and R^2= 0.6638 for the RF model. Also, it is seen that the best model is RF with the smallest MSE = 0.0038 among machine learning algorithms. When the prediction errors of beta regression and random forest model are examined, it is seen that the RF model is relatively superior to the beta regression model in predicting the success rates (Kikawa et al., 2020). While the beta regression model predicted the success rates of the best Çanakkale province, the RF model predicted the success rates of Ankara. Besides, variables significant in the beta regression model appear to be important in the RF model. According to the RF model, the three most important variables were found as "GDP," "divorce rate," and "book reading rate," respectively. The limitation of this study is that the data of LGS exam results for 2020 are still not disclosed in Turkey due to Covid-19. These two models are likely to provide a scientific basis for predicting students 'high school entrance examination success rates for all provinces in the following years.

References

Abbasoğlu, B., 2020. Ortaokul Öğrencilerinin Akademik Başarılarının Eğitsel Veri Madenciliği Yöntemleri İle Tahmini. Veri Bilimi, 3(1), 1-10.

Abidi, S. M. R., Hussain, M., Xu, Y. L., & Zhang, W., 2019. Prediction of Confusion Attempting Algebra Homework in an Intelligent Tutoring System through Machine Learning Techniques for Educational Sustainable Development. Sustainability, 11(1), 105. doi:ARTN 105 10.3390/su11010105

Al Mayahi, K. and Al-Bahri, M., 2020. Machine Learning Based Predicting Student Academic Success. Paper presented at the 2020 12th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT).

Breiman, L., Friedman, J., Olshen, R., & Stone, C., 1998. CART. In: Chapman and Hall/CRC.

Cepeda-Cuervo, E., 2015. Beta regression models: Joint mean and variance modeling. Journal of Statistical Theory and Practice, 9(1), 134-145.

Çömlekcioğulları, A. (2020). Öğrenci başarısı ile ailelerin sosyo-ekonomik düzeyleri arasındaki ilişki (Denizli ili örneği).

Dünder, E., & Cengiz, M. A., 2020. Model selection in beta regression analysis using several information criteria and heuristic optimization. Journal of New Theory(33), 76-84.

Ferrari, S. L. P., & Cribari-Neto, F., 2004. Beta regression for modelling rates and proportions. Journal of applied statistics, 31(7), 799-815. doi:10.1080/0266476042000214501

Friedman, C., & Sandow, S., 2011. Utility-based learning from data. Boca Raton: Chapman & Hall/CRC.

Gök, M., 2017. Makine öğrenmesi yöntemleri ile akademik başarının tahmin edilmesi. Gazi Üniversitesi Fen Bilimleri Dergisi Part C: Tasarım ve Teknoloji, 5(3), 139-148.

Gunn, S. R., 1998. Support vector machines for Classification and regression. ISIS technical report, 14(1), 5-16.

Gür, B., Çelik, Z., & Coşkun, İ., 2013. Türkiye'de ortaöğretimin geleceği: Hiyerarşi mi eşitlik mi. Seta analiz, 69, 1-26.

Han, H., Wang, W.-Y., & Mao, B.-H., 2005. Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning. Paper presented at the International conference on intelligent computing.

Kikawa, C. R., Ngungu, M. N., Ntirampeba, D., & Ssematimba, A. (2020). Support vector regression and beta distribution for modeling incumbent party for presidential elections. Appl. Math, 14(4), 721-727.

Koç, T., 2019. Türkiye'de boşanma oranlarini etkileyen faktörlerin beta regresyon modeli ile belirlenmesi. Avrasya Uluslararası Araştırmalar Dergisi, 7(16), 1111-1117.

Kotsiantis, S. B., 2011. Decision trees: a recent overview. Artificial Intelligence Review, 39(4), 261-283. doi:10.1007/s10462-011-9272-4

Liaw, A., & Wiener, M., 2002. Classification and regression by randomForest. R news, 2(3), 18-22.

MEB. ,2018). 2018 yılı performans programı Ankara Retrieved from sgb.meb.gov.tr.

Oral, I., & McGivney, E. J. ,2014. Türkiye eğitim sisteminde eşitlik ve akademik başarı, araştırma raporu ve analiz. İstanbul: Sabancı Üniversitesi Yayınları.

Özeren, E., Çiloğlu, T., Yilmaz, R., & Özeren, A. (2020). Öğrencilerin akademik kariyer hedefi seçiminde etkili olan faktörlerin veri madenciliği yöntemi ile belirlenmesi: Bartın başarı takip araştırması sonuçları üzerine bir inceleme. Bilgi ve İletişim Teknolojileri Dergisi, 2(2), 182-210.

Rajak, A., Shrivastava, A. K., & Vidushi. ,2020. Applying and comparing machine learning classification algorithms for predicting the results of students. Journal of Discrete Mathematical Sciences & Cryptography, 23(2, 419-427. doi:10.1080/09720529.2020.1728895

Rebai, S., Ben Yahia, F., & Essid, H. ,2020. A graphically based machine learning approach to predict secondary schools performance in Tunisia. Socio-Economic Planning Sciences, 70, 100724. doi:ARTN 100724 10.1016/j.seps.2019.06.009

Sethi, K., Jaiswal, V., & Ansari, M. D. ,2020. Machine learning based support system for students to select stream ,subject). Recent Advances in Computer Science and Communications ,Formerly: Recent Patents on Computer Science), 13(3), 336-344.

Shokry, A., Audino, F., Vicente, P., Escudero, G., Moya, M. P., Graells, M., & Espuña, A. ,2015. Modeling and simulation of complex nonlinear dynamic processes using data-based models: Application to photo-Fenton process. In Computer Aided Chemical Engineering (Vol. 37, pp. 191-196): Elsevier.

Therneau, T. M., & Atkinson, E. J. ,1997. An introduction to recursive partitioning using the RPART routines. Retrieved from

Uğuz, S. ,2019. Makine Öğrenmesi Teorik Yönleri ve Python Uygulamaları ile Bir Yapay Zeka Ekolü. Ankara: Nobel.

Uskov, V. L., Bakken, J. P., Byerly, A., & Shah, A. ,2019. Machine learning-based predictive analytics of student academic performance in STEM education. Paper presented at the 2019 IEEE Global Engineering Education Conference (EDUCON).

Vapnik, V. ,1992. Principles of risk minimization for learning theory. Paper presented at the Advances in Neural Information Processing Systems.

Yavuz, A., 2020. Ortaöğretime geçiş sınavında öğrenci başarısını etkileyen etmenler.

Yousafzai, B. K., Hayat, M., & Afzal, S. , 2020. Application of machine learning and data mining in predicting the performance of intermediate and secondary education level student. Education and Information Technologies, 25(6), 4677-4697. doi:10.1007/s10639-020-10189-1