ADIYAMAN UNIVERSITY Journal of Educational Sciences

dergipark.org.tr/tr/pub/adyuebd Volume 15 , Issue 1, June 2025, Page 159-184 ISSN: 2149-2727

A Comparative Analysis of Different Machine Learning Models for Classifying Student Achievement*

Tansu Alan[1]

Article Info

Keywords
Education,
Machine learning,
Classification accuracy,
Classification models

Received: 16.09.2024

Accepted: 17.02.2025

Published: 30.06.2025

Research Article

DOI: 10.17984/adyuebd.1551029

Abstract

In the context of teaching and learning, evaluating and classifying student achievement is critical for determining the effectiveness of instructional methods. Categorizing students' academic performance into groups such as "passed," "failed," "successful," and "unsuccessful" provides valuable insights for tracking academic progress and improving instructional strategies. The use of Machine Learning (ML) models in such classifications enables more accurate and objective evaluations, particularly when dealing with large datasets. Therefore, this study aims to examine the accuracy of various ML models in classifying student performance. ML offers enhanced precision and objectivity by analyzing large and complex educational datasets. In this study, the classification accuracies of three machine learning algorithms—Naive Bayes (NB), Support Vector Machines (SVM), and Random Forest (RF)—were evaluated. The research compares the performance metrics of these models in predicting students' academic success and examines the results in detail. As such, the study adopts a descriptive survey design and has an applied nature. A dataset comprising 1,000 samples and variables such as ethnicity, parental education level, and mathematics achievement was used. The analyses were conducted using SPSS and R software. The findings reveal that the Random Forest model achieved the highest classification accuracy. The integration of ML models in education can contribute to improving educational quality by predicting student success, identifying risk of failure, and evaluating the effectiveness of instructional methods and materials.

Introduction

One of the primary goals of education is to enhance students' skills and expand their knowledge (Hart, 2016). Therefore, it is essential to design and implement educational processes tailored to each student's abilities and areas of competence (Corno & Snow, 1986). Given the vital role of education, student achievement becomes a critical component of the learning process and a key indicator of the quality of educational institutions. Academic success can be evaluated not only based on cognitive factors but also through social, economic, individual, and environmental dimensions (Alamri et al., 2020). With the increasing availability of data related to these factors, the role of artificial intelligence (AI) and machine learning (ML) technologies in processing and analyzing large-scale educational data has become increasingly significant (Williamson, 2019; Yang, 2019).

Machine Learning (ML), which lies at the core of Artificial Intelligence (AI) and data science, is one of the fastest-growing fields in computer science and has a wide range of applications (Jordan & Mitchell, 2015; Shalev-Shwartz & Ben-David, 2014). ML refers to a set of algorithms that are capable of detecting complex

---

* A portion of this study was presented by Tansu Alan at the 9th International Congress on Measurement and Evaluation in Education and Psychology, held between October 3–6, 2024, under the title "A Comparative Analysis of Different Machine Learning Models for Classifying Student Achievement"

Corresponding author[1] : Tansu Alan, Adıyaman University, Turkiye, tansualan@adiyaman.edu.tr

Alan, T. (2025). A comparative analysis of different machine learning models for classifying student achievement. *Adiyaman Univesity Journal of Educational Sciences, 15*(1), 159-184.

patterns and making data-driven decisions. These algorithms are designed to learn from data, thereby acquiring the ability to solve problems independently without being explicitly programmed for each task (Gültepe, 2019). ML technology enables the prediction of future outcomes from complex datasets, playing a critical role in situations where the volume and complexity of data exceed human analytical capacities (Türkmenoğlu & Tantuğ, 2014). Regardless of how large the dataset is, ML continues to improve prediction accuracy and enhances the precision of outcomes (IBM, 2024). Moreover, ML can generate new models and forecasts, offering personalized predictions and recommendations based on individual needs and circumstances. Due to these capabilities, ML models are widely used in fields such as medicine, science, and manufacturing to support evidence-based decision-making processes (Kourou et al., 2015; Lin et al., 2011; Wu et al., 2017).

Machine learning (ML) algorithms enhance predictive power by integrating data from various sources and managing large volumes of data efficiently (Özlüer Başer et al., 2021). These algorithms are capable of performing tasks such as clustering, classification, prediction, and estimation (Ülker & Civalek, 2002), and are trained using supervised, unsupervised, or semi-supervised learning models (Brownlee, 2016). In supervised learning, both input and output variables are provided, and the primary goal is to generate predictions for new data based on these variables (Aydın & Özkul, 2015; Ülker & Civalek, 2002). For instance, in a dataset where students are labeled as either successful or unsuccessful, the success level of a new student can be predicted by evaluating the relevant variables. In contrast, unsupervised learning uses only input variables without labeled outputs (Ülker & Civalek, 2002). Semi-supervised learning combines both approaches, utilizing a mix of labeled and unlabeled data (Sarıcaoğlu, 2019). Since the dataset used in this study includes both input (independent) and output (dependent) variables, the analysis was conducted using several supervised machine learning models, and the results were compared accordingly.

Just as machine learning (ML) is utilized across various fields, it can also be effectively applied in education, offering significant advantages in this domain (Aydın, 2007). When implemented in educational contexts, ML enables the adaptation of learning processes to students' individual developmental paces and performance levels, thereby facilitating more personalized learning environments (Kuch et al., 2020). This potential has garnered increasing interest in recent years, as reflected in the growing number of studies in the field of education (Zhou et al., 2018). The diversity of data collected from multiple educational sources further highlights the importance of big data analytics. Analyzing such data allows for applications such as classifying student performance, predicting educational outcomes, identifying necessary interventions during the learning process, recommending appropriate content based on learners' methods and styles, and detecting knowledge gaps based on assessment results (Tosunoğlu et al., 2021).Therefore, ML can support teachers by predicting student performance and enabling process-oriented evaluation (Kucak et al., 2018). Additionally, it offers the potential to eliminate scorer bias and objectively assess students' academic achievements. ML can also aid teachers and education planners in identifying the most effective strategies to improve student success (Alamri et al., 2020). Recent studies indicate that supervised ML models can accurately predict students' grades and hold promise for identifying students at risk of academic failure at an early stage. This makes it possible to conduct student assessments in a more objective and effective manner (Wu et al., 2018).

Numerous studies have investigated the use of machine learning (ML) in education. In a study by Alamri et al. (2020), Random Forest (RF) and Support Vector Machine (SVM) models were employed to predict the Portuguese language and mathematics grades of Portuguese students. The study concluded that both algorithms yielded high accuracy rates. Vandamme et al. (2007) utilized students' demographic, socioeconomic, and academic background data to classify them into different performance groups using an artificial neural network model and examined their academic success. In a study by Şengür and Tekin (2013), artificial neural networks and decision tree models were used to predict students' graduation grades, and the results showed that artificial neural networks outperformed decision trees. Hussain et al. (2018) analyzed data from the Common Entrance Examination (CEE) used for university admissions, comparing various classification and clustering models, and found that artificial neural networks achieved the highest accuracy rate. In a more recent study by Zilyas and Yılmaz (2023), various algorithms such as K-Nearest Neighbors, Random Forest, Linear Regression, Bagged Trees Regression, Gradient Boosting Regressor, and Decision Trees were used to predict achievement in Turkish language courses. The results indicated that the Random

Forest model was the most successful in terms of classification accuracy. Similarly, Quinn and Gray (2020) applied different ML models to student data to predict whether a student would pass a course. These algorithms included Random Forest, Gradient Boosting Regressor, Linear Discriminant Analysis, and K-Nearest Neighbors. Their findings demonstrated that the Random Forest model was more effective than other algorithms in predicting student academic performance.

A review of the relevant literature reveals that various machine learning models have been employed in the field of education, with findings often discussing which models perform better. These studies demonstrate that diverse machine learning techniques can be effectively utilized to classify student academic success. Accurate classification of student performance is a crucial step toward improving educational efficiency and providing individualized educational services. By correctly classifying students, their strengths and weaknesses can be identified, enabling more targeted support and consequently enhancing their academic achievement. Although a wide range of models and algorithms can be employed for this purpose, statistical models such as decision trees, Random Forest (RF), Support Vector Machines (SVM), Naive Bayes, K-Nearest Neighbors (KNN), and logistic regression are among the most commonly used in educational research (Hussain et al., 2018; Rao et al., 2016; Vijayalakshmi & Venkatachalapathy, 2019; Wu et al., 2018). In this study, the advantages of the selected models played a decisive role in their choice. The Random Forest model stands out due to its high accuracy and strong capability to model complex interactions among variables. RF effectively reveals relationships within multidimensional and complex datasets, thereby enhancing the accuracy of predictions. Additionally, its robustness against overfitting offers a significant advantage over other models (Rebai et al., 2020; Cruz-Jesus et al., 2020). Support Vector Machines (SVM), known for their strong generalization ability, perform well in classification and regression tasks. They maintain similar levels of success even on previously unseen data, demonstrating high performance on training datasets (Hastie et al., 2017). The Naive Bayes method attracts attention due to its mathematically interpretable structure and practical utility. Its effectiveness on small datasets and applicability to both categorical and continuous data types (Hamalainen & Vinni, 2011) made it a suitable choice for this study. Given these advantages, these models were preferred with the expectation that they would facilitate more accurate classification and evaluation of student academic success.

A review of previous studies reveals that while some research focuses on analyzing variables that have direct or indirect effects on student outcomes, others attempt to predict student performance based on these variables using different machine learning algorithms. Although a student's academic success largely depends on various socio-economic, school-related, and personal factors (Yassein et al., 2017), this study examines the classification performance of Random Forest, Support Vector Machine, and Naive Bayes models using various variables available in the relevant dataset. An open-access dataset was utilized in the study, where students' mathematics achievement averages served as a reference point. Student success was classified into two categories: below average and above average. The aim of the research is to compare and evaluate the performance of frequently used supervised machine learning models in education and other disciplines—namely Naive Bayes (NB), Random Forest (RF), and Support Vector Machine (SVM)—by analyzing the classification results based on machine learning performance metrics obtained from the respective algorithms. Accordingly, the study seeks to answer the research question: How do the performances of different machine learning models compare in classifying student academic success?

Method

Research Model

In the study, the classification accuracies of different machine learning models were compared using performance metrics. This research, which aims to identify the most successful model through the comparison of various models, is a descriptive survey-type applied study. Applied research involves the experimental application of theoretical knowledge that has been produced or is being produced and can provide concrete contributions toward improving existing practices (Karasar, 2012).

Data Collection

In this study, a dataset collected to understand the effects of various economic, individual, and social variables on the mathematics achievement of high school students living in the United States was used. Although the dataset consists of students from the United States, the primary aim of the study is not to focus on mathematics achievement itself but to test the classification accuracy of machine learning models. The various variables in the dataset were utilized to evaluate the applicability of these accuracy models and to compare the performance of the models. For this purpose, a publicly available dataset consisting of 1000 samples (https://www.kaggle.com/datasets/scientist/students-performance-in-exams) was obtained from an online platform. Kaggle, a platform owned by Google, not only hosts machine learning competitions for data scientists of all levels but also provides a comprehensive collection of datasets and a cloud-based data science environment. These features enable users to effectively conduct data analysis and model development processes (Quaranta et al., 2021). While companies and researchers can share their data, data scientists and researchers can experiment with their own approaches and simultaneously compete to achieve the highest competition score (Wang et al., 2021). In other words, data scientists seek real answers to real problems on this platform (Kuroki, 2023). The dataset includes information on 482 male (48.2%) and 518 female (51.8%) students. The following variables were considered for each group: lunch status, exam preparation course, race/ethnicity, parents' education level, gender, and mathematics score. Since the aim of this study is to compare the classification accuracy of machine learning models, this dataset was used. The dependent and independent variables considered in the analysis are shown in Table 1.

Table 1. Dependent and Independent Variables

Variable	Description	Data Type	Variable Type
Gender	(0 = male, 1 = female)	Categorical	Independent
Race/Ethnicity	(0 = Group A, 1 = Group B, 2 = Group C, 3 = Group D, 4 = Group E)	Categorical	Independent
Parents' Education Level	(0 = some college/no degree, 1 = associate degree, 2 = some high school/ no degree, 3 = high school graduate, 4 = bachelor's degree, 5 = Master's degree)	Categorical	Independent
Lunch	(0 = paid, 1 = free)	Categorical	Independent
Test Preparation Course	(0 = not completed, 1 = completed)	Categorical	Independent
Mathematic Score	(0 = below average, 1 = above average)	Categorical	Dependent

Data Analysis

It is necessary to detect and handle erroneous, missing, and lost data in the dataset to improve the accuracy and integrity of the data. To correctly interpret the research results, missing values and outliers in the data were first examined. Frequencies (percentages) were calculated to summarize the descriptive characteristics. Pearson's chi-square test was used to examine the relationships between the independent variables and mathematics scores. Effect size was also evaluated to determine whether the differences between groups were statistically significant (Sullivan & Feinn, 2012). Cohen's d statistic was used to assess effect size. According to Cohen (1988), an effect size below 0.2 is considered weak, 0.5 moderate, and above 0.8 strong. Statistical analyses were performed using SPSS software. The classification metric results of the machine learning models were analyzed using the R program. Five-fold cross-validation was performed using the entire dataset. In other words, the dataset was divided into five parts for the evaluation and training of the classification models, and the accuracy rates of the models were calculated. The supervised machine learning models used in this study are described in detail below.

Naive Bayes (NB)

The Naive Bayes model, used in machine learning, is an approach for determining the class to which an instance belongs. This model calculates the class probabilities for each feature based on Bayes' theorem. Using these probabilities, the class with the highest probability is predicted. The term "Naive" is used because the algorithm assumes that each feature is independent of the others. However, despite this assumption, high accuracy rates are often achieved in practice. Bayes' theorem is expressed as follows (Rish, 2001):

$$P(A\backslash B) = \frac{P(B\backslash A)P(A)}{P(B)} \qquad (1)$$

In this equation:

P(A|B): The probability of event A occurring given that event B has occurred.
P(B|A): The probability of event B occurring given that event A has occurred.
P(A): The probability of event A occurring.
P(B): The probability of event B occurring.

Support Vector Machine (SVM)

Support Vector Machine (SVM) is a widely used classification model in machine learning. This model is employed to classify data samples, i.e., to determine which class they belong to. In classification tasks, training and test datasets containing specific data samples are generally used (Duda & Hart, 1973). The main objective of the SVM model, which has recently gained popularity, is to find the hyperplane that most effectively separates the target variable's classes. In other words, it tries to find a hyperplane that maximizes the margin between the classes. The margin refers to the distance between the separating hyperplane and the nearest data points of each class, and its width directly affects classification accuracy. SVM aims to find the hyperplane that maximizes this margin, thereby optimally separating the classes. To achieve this, SVM employs a training process that minimizes a loss function. This process continues iteratively until the optimal hyperplane is found. The loss function measures the classification errors made by the model and updates the hyperplane to minimize these errors and construct an optimal hyperplane (Cristianini & Shawe-Taylor, 2000; Vapnik, 2013). The classification logic of the support vector machine algorithm is illustrated in Figure 1 (Huang et al., 2006).

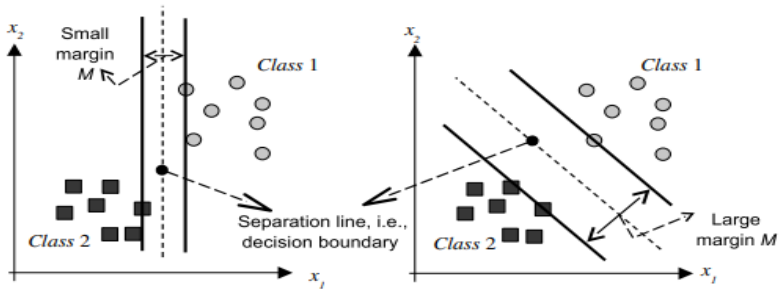


Figure 1. Support Vector Machine Algorithm

Random Forest (RF)

Random Forest (RF) is a machine learning algorithm that provides robust and reliable predictions in classification tasks by using multiple decision trees. The RF algorithm aims to make stronger and more accurate predictions by generating multiple decision trees and combining their results. This ensemble approach enhances model performance and generally yields better results than a single decision tree (Breiman, 2001). Compared to a single decision tree, the RF approach significantly reduces the risk of overfitting and is effective in identifying outliers. Moreover, it is one of the most suitable models for determining the most important variables in a dataset (Iwendi et al., 2020). The classification logic of the Random Forest algorithm is illustrated in Figure 2 (Khan et al., 2021).

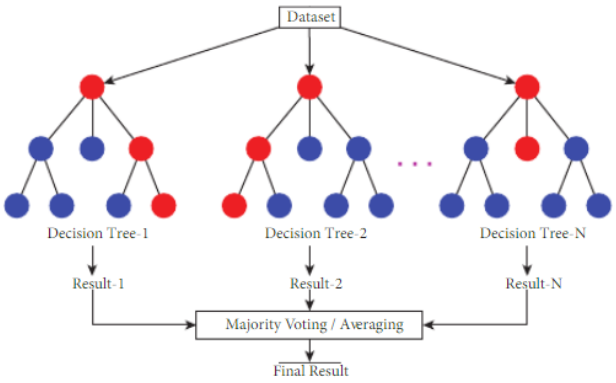


Figure 2. Random Forest Algorithm

Evaluation Criteria for Model Performance

The Confusion Matrix is a tool used to evaluate how well classification models perform, determine model accuracy, and identify errors. In this study, the confusion matrix to be used for measuring the classification performance of the models, along with the evaluation criteria and the formulas related to these criteria, are presented in Table 2 and Table 3.

Table 2. Confusion Matrix

	Prediction	
Reference /Actual	Below Average	Above Average
Below Average	True Negatives (TN)	False Positives (FP)
Above Average	False Negatives (FN)	True Positives (TP)

When examining the matrix in Table 2, the rows represent the reference values (actual outcomes), and the columns represent the predicted values (model's predictions). True negatives (TN) are observations predicted as below average and actually below average. False positives (FP) are observations predicted as above average but actually below average. False negatives (FN) are observations predicted as below average but actually above average. True positives (TP) are observations predicted as above average and actually above average.Table 3.

Table 3. Classification Performance Metrics

Criterion	Formula/Approach
Accuracy	$\frac{TP+TN}{TP+TN+FN+FP}$
Sensitivity	$\frac{TP}{TP+FN}$
Specificity	$\frac{TN}{TN+FP}$
Youden's index	$\frac{TP}{TP+FN}+\frac{TN}{TN+FP}-1$
Positive predictive value (PPV)	$\frac{TP}{TP+FP}$
Negative predictive value (NPV)	$\frac{TN}{TN+FN}$
F1-Score	$\frac{2*PPV*\text{Sensitivity}}{PPV+\text{Sensitivity}}$

Performance Evaluation Criteria

With machine learning algorithms, performance indicators such as accuracy, sensitivity, specificity, Youden's index, F1-score, positive predictive value (PPV), and negative predictive value (NPV) are used to balance correct classification and determine the best approach. The characteristics of these criteria are as follows (Racz et al., 2019):

1. Accuracy: Accuracy, known as the proportion of correctly classified data by the trained machine learning model, can also be defined as the ratio of correct predictions to total predictions across all classes. In other words, it reflects the overall performance of the model. A high accuracy indicates that the model is successful in both positive and negative classifications.

2. Sensitivity: The model's sensitivity indicates how well it identifies true positive cases. On the other hand, an increase in sensitivity may also lead to more false positive errors.

3.Specificity: The specificity level indicates how well the model identifies true negatives. However, an increase in specificity may also lead to a higher chance of false negative errors.

4.Youden's İndex: This index summarizes how well a model performs in terms of both sensitivity and specificity. A value close to 1 indicates that the model performs well on both metrics.

5.Positive predictive value (PPV): This metric indicates the proportion of cases classified as positive by the model that are actually positive. It is an important parameter showing the accuracy of correctly predicting positive data points. Additionally, it can be used to evaluate the performance of classification algorithms. This metric also assesses the model's tendency to make false positive predictions.

6.Negative predictive value (NPV): This metric shows the proportion of cases classified as negative by the model that are actually negative. In other words, it evaluates the model's tendency to make false negative predictions. A high NPV indicates the accuracy of the model's negative predictions. This value is especially important in situations where reducing false negative errors is critical.

7.F1-Score: The metric named after the parameter value $\beta = 1$ is the most commonly used member of the parametric family of F-measures. It is a metric used to evaluate the performance of classification models by balancing sensitivity (recall) and positive predictive value (PPV or precision). The F1-Score harmonizes these two metrics. If the model has both low false positive errors (incorrectly predicting successful students) and low false negative errors (missing unsuccessful students), the F1-Score will be high.

Ethical Disclosure

Ethics Committee Approval: This research was conducted with the permission obtained by the Adıyaman University Scientific Research and Publication Ethics Social and Human Sciences Board's decision dated 20/05/2024 and numbered 62

Findings

According to the examination, it was concluded that there are no outliers or missing data in the dataset. Descriptive statistics for the variables are presented in Table 4, while the performance measurement results of the models used for classifying student success are shown in Table 5.

Table 4. Descriptive Statistics of the Variables

		Group				
		Below Average	Above Average			
Variables	Category	n (%)	n (%)	Effect size (ES)	Effect Magnitude	p
Gender	Female	297 (58.6)	221 (44.8)	0.138	Small Effect	<0.001*
	MAle	210 (41.4)	272 (55.2)			
Race/Ethnicity	Group A	58 (11.4)	31 (6.3)	0.211	Moderate Effect	<0.001*
	Group B	112 (22.1)	78 (15.8)			
	Group C	179 (35.3)	140 (28.4)			
	Group D	116 (22.9)	146 (29.6)			
	Group E	42 (8.3)	98 (19.9)			
Parental Education Level	some college/no degree	107 (21.1)	119 (24.1)	0.115	Small Effect	0.021*
	associate degree	109 (21.5)	113 (22.9)			
	some high school/ no degree	98 (19.3)	81 (16.4)			
	high school graduate	117 (23.1)	79 (16.0)			
	Master's degree	24 (4.7)	35 (7.1)			
	Bachelor's degree	52 (10.3)	66 (13.4)			
Lunch	Discounted	262 (51.7)	383 (77.7)	0.271	Moderate Effect	<0.001*
	Free	245 (48.3)	110 (22.3)			
Test Preparation Course	Not completed	360 (71.0)	282 (57.2)	0.114	Small Effect	<0.001*
	Completed	147 (29.0)	211 (42.8)			

**: Pearson Chi-Square*

When the dataset of 1,000 samples is examined in terms of the gender variable, it is observed that the number of female students whose mathematics achievement scores are below average is 297 (58.6%), while the number above average is 221 (44.8%). For male students, the number of those below average is 210 (41.4%) and above average is 272 (55.2%). This distribution shows that there is a small difference between the number of students below the average (507; 50.7%) and those above the average (493; 49.3%). Since both classes are represented in nearly equal proportions, the classification models minimize the negative effects that can arise from class imbalance. Class imbalance typically becomes a significant issue when one class is significantly larger or smaller than the other, which can adversely affect the training process of the model by causing it to favor the majority class and struggle to accurately identify the minority class (Buda et al., 2018). A statistically significant difference was found between gender and mathematics scores ($p < 0.001$), and this difference had a small effect size (ES = 0.138). Similarly, it was concluded that there were statistically significant differences between mathematics achievement scores and the variables of race/ethnicity, parental education level, lunch status, and test preparation course attendance ($p < 0.001$). When effect sizes were analyzed, race/ethnicity and lunch variables had effect sizes of 0.211 and 0.271 respectively, indicating a

moderate level of effect. The effect sizes for parental education level and test preparation course were found to be 0.115 and 0.114, respectively, suggesting a small level of effect for these variables.

Table 5. Performance Measurement Results of the Models Used for Classifying Student Achievement

Model	NB (95% C.I.)	SVM (95% C.I.)	RF (95% C.I.)
Accuracy	0.65 (0.62-0.68)	0.66 (0.63-0.69)	0.68 (0.65-0.71)
Sensitivity	0.72 (0.67-0.76)	0.53 (0.48-0.57)	0.64 (0.60-0.68)
Specificity	0.59 (0.54-0.63)	0.78 (0.74-0.82)	0.72 (0.68-0.76)
Youden's index	0.30 (0.22-0.39)	0.30 (0.22-0.39)	0.36 (0.27-0.44)
PPV	0.62 (0.58-0.66)	0.69 (0.64-0.74)	0.68 (0.63-0.72)
NPV	0.69 (0.64-0.73)	0.64 (0.60-0.68)	0.68 (0.64-0.72)
F1-Score	0.66 (0.63-0.69)	0.60 (0.57-0.63)	0.66 (0.63-0.69)

NB: Naive Bayes; SVM: Support Vector Machine; RF: Random Forest; PPV: Positive Predictive Value; NPV: Negative Predictive Value; C.I.: Confidence Interval

Table 5 compares the classification outputs of the NB, SVM, and RF models. To evaluate model performance, the metrics used were accuracy, sensitivity, specificity, Youden's index, PPV, NPV, and the F1-score. Based on classification accuracy, the RF model demonstrated superior performance in classifying student achievement, achieving an accuracy of 0.68 (95% C.I.: 0.65–0.71). In contrast, the lowest classification accuracy was obtained with the NB model at 0.65 (95% C.I.: 0.62–0.68).

When examining sensitivity results, which indicate how well the model identifies true positive cases (i.e., successful students), the highest sensitivity was found in the NB model at 0.72 (95% C.I.: 0.67–0.76). The lowest sensitivity was observed in the SVM model at 0.53 (95% C.I.: 0.48–0.57). The RF model ranked second with a sensitivity of 0.64 (95% C.I.: 0.60–0.68).

The specificity values, which indicate how well the models correctly identify negative cases (unsuccessful students), were examined. The highest specificity was obtained from the SVM model at 0.78 (95% CI: 0.74–0.82), while the lowest was from the NB model at 0.59 (95% CI: 0.54–0.63). The RF model ranked second with a specificity of 0.72 (95% CI: 0.68–0.76).

The Youden index, which indicates how well the models perform in terms of both sensitivity and specificity, was examined. The highest value was found in the RF model at 0.36 (95% CI: 0.27–0.44). For the SVM and NB models, the Youden index was calculated as 0.30 (95% CI: 0.22–0.39).

When examining the Positive Predictive Value (PPV), which shows how many of the cases classified as positive are actually positive, the highest PPV was obtained with the SVM model at 0.69 (95% CI: 0.64–0.74), while the lowest was found with the NB model at 0.62 (95% CI: 0.58–0.66). The RF model had a PPV of 0.68 (95% CI: 0.63–0.72).

When examining the Negative Predictive Value (NPV), which indicates how many of the cases classified as negative are truly negative, the highest NPV was found to be 0.69 (0.64–0.73) for the SVM model, while the lowest was 0.64 (0.60–0.68) for the NB model. The RF model had an NPV of 0.68 (0.64–0.72).

The F1 score, which balances sensitivity and positive predictive value (PPV) used to evaluate the performance of classification models, was found to be equal for the NB and RF models at 0.66 (0.63–0.69), while it was 0.60 (0.57–0.63) for the SVM model.

Conclusion and Discussion

Predicting student performance in advance is an important issue in the field of education. Analyzing students' academic achievement and risk of failure can contribute to improving the overall performance of educational institutions. Through such analyses, schools can provide students with the support they need, achieve better outcomes, and increase success rates. In this study, student achievement was categorized and analyzed as "above average" and "below average." The classification accuracy of different machine learning models was examined to ensure accurate predictions. Naive Bayes (NB), Support Vector Machine (SVM), and

Random Forest (RF) machine learning models were used to evaluate the classification performance of students' mathematics achievement. The models were assessed using performance evaluation metrics such as accuracy, F1 score, sensitivity (recall), specificity, Youden's index, negative predictive value (NPV), and positive predictive value (PPV).

When the machine learning (ML) models were tested, it was found that the Random Forest (RF) model had the highest accuracy, Youden's index, and F1 score. The RF model achieved a correct classification rate of 68% for students' mathematics performance, with a 95% confidence interval of 0.65–0.71. The F1 score of the RF model was 0.66 (0.63–0.69), sensitivity 0.64 (0.60–0.68), specificity 0.72 (0.68–0.76), Youden's index 0.36 (0.27–0.44), positive predictive value (PPV) 0.68 (0.63–0.72), and negative predictive value (NPV) 0.68 (0.64–0.72). In previous studies, the correct classification rate of the RF model has been reported to range from 45% to 99% (Gunawan et al., 2021; Alamri et al., 2020; Nurhachita & Negara, 2021; Rodrigues et al., 2020). Especially when both accuracy and overall classification performance are considered, the RF model may be preferred, and the model's overall performance can be improved by including more predictive variables. The findings also indicate that ethnicity and the lunch variable, which reflects socioeconomic status, are important predictors of students' academic achievement. Compared to other models, the RF model has been recommended in some studies due to its ability to produce higher accuracy rates in classification and prediction (Jayaprakash et al., 2020; Rao et al., 2016). Tree-based predictive models, particularly the RF model, offer a strong alternative to traditional techniques in predicting academic success in education. These models can capture non-linear interactions among variables, thereby increasing the predictive accuracy of the model (Geurts et al., 2009; Golino et al., 2014).

When similar studies in the field of education are examined, some research (Gunawan et al., 2021; Alamri et al., 2020; Yağcı, 2022) has shown that the Random Forest (RF) model demonstrates higher classification accuracy, thereby supporting the findings of this study. On the other hand, other studies have reported that the Naive Bayes (NB) and Support Vector Machine (SVM) models (Nurhachita & Negara, 2021; Rodrigues et al., 2020; Nurpiana & Wijayanto, 2022) provide superior classification performance. The accuracy of predictive models can vary depending on the variables used in the prediction process. In studies using the SVM model to predict student performance, important predictive variables have included psychosocial factors such as student interest, study habits, and family support (Sembiring et al., 2011; Shahiri et al., 2015). The higher classification accuracy of the NB model in some similar studies may be due to differences in sample size compared to this study. When the sample size is significantly increased, the NB model can demonstrate higher correct classification performance (Koyuncu, 2018).

The results of this study indicate that the Support Vector Machine (SVM) model achieved the highest specificity in identifying underperforming students. This finding is consistent with the study by Gray et al. (2014), supporting the view in the literature that the SVM model is an effective method for accurately identifying students at risk. Therefore, the use of the SVM model appears to be a more appropriate choice in processes aimed at identifying unsuccessful students.

Although the Naïve Bayes (NB) model demonstrates good performance in capturing successful students due to its high sensitivity, it shows lower performance in other metrics such as overall accuracy, specificity, and PPV. This model may be considered particularly when identifying successful students is of primary importance; however, it lags behind other models in terms of overall performance.

In this study, the results obtained from three different models used in classification processes were compared. A review of the literature reveals that there are many alternative models available for classifying student achievement based on various variables. In other studies, researchers may evaluate the performance of different models such as KNN, decision trees, and logistic regression under similar or varying conditions.

Socio-demographic variables, school-related variables, living conditions, and other behavioral factors enhance the predictive power of models (Cortez & Silva, 2008). To more accurately evaluate the effectiveness of machine learning models, future studies can include a greater number of variables. This would allow machine learning models to analyze student achievement in a more comprehensive and precise manner, providing educational administrators with more effective data-driven strategies to support students. In this

way, a more detailed assessment of the multidimensional factors affecting student achievement would enable a stronger evaluation of the benefits that machine learning models can offer in the field of education.

Machine learning models can guide educational administrators not only by providing classification results but also by revealing which variables reduce the risk of failure or enhance success. The use of data based models to predict students' academic performance in advance increases the practical value of machine learning findings (Rajendran et al., 2022) and can offer a more comprehensive roadmap for efforts to improve educational quality. When a student's performance is classified as low by the model, it should also be able to identify the variables influencing this outcome. By analyzing the factors affecting student performance, the model can generate personalized recommendations. For example, if variables such as socioeconomic status, motivation, and stress levels are found to significantly impact performance classification, financial aid programs and free supplementary lessons could be recommended for students with low socioeconomic status. For students with low motivation, motivational seminars and goal-setting workshops could be proposed. Additionally, for students experiencing high stress levels, offering psychosocial support services and guidance on stress management and relaxation techniques could positively affect their learning processes. These kinds of interventions aim to help students improve their academic performance. Thus, the outcomes of machine learning can form the foundation for educational interventions and contribute to making student support systems more effective.

This study may serve as a valuable resource for education, the field, and researchers, as it compares various machine learning models that can help identify low-performing students, address failures at an early stage, and take appropriate measures to improve student performance.

A notable limitation of this study is that the sample consists solely of participants from a single region. Therefore, the generalizability of the findings to other regions may be limited. The adaptation and implementation of the proposed model in other institutions might yield different results due to observed heterogeneity among students from various academic disciplines and schools with diverse educational strategies.

ADIYAMAN UNIVERSITY Journal of Educational Sciences

dergipark.org.tr/tr/pub/adyuebd Cilt 15 , Sayı 1, Haziran 2025, Sayfa 159-184 ISSN: 2149-2727

Öğrenci Başarısının Sınıflandırılmasında Farklı Makine Öğrenmesi Modellerinin Karşılaştırılması*

Tansu Alan [1]

Makale Bilgisi

Anahtar Kelime
Eğitim,
Makine öğrenmesi,
Sınıflandırma doğruluğu,
Sınıflandırma modelleri

Yükleme: 16.09.2024

Kabul: 17.02.2025

Yayın: 30.06.2025

Araştırma Makalesi

DOI: 10.17984/adyuebd.1551029

Özet

Eğitim-öğretim sürecinde öğrenci başarısının değerlendirilmesi ve sınıflandırılması, öğretim yöntemlerinin etkinliğini belirlemek açısından kritiktir. Öğrencilerin başarı durumlarını "geçti", "kaldı", "başarılı" ve "başarısız" gibi kategorilere ayırmak hem öğrencilerin akademik gelişimini izlemek hem de öğretim stratejilerini iyileştirmek için önemli bilgiler sağlar. Makine Öğrenmesi (MÖ) modellerinin bu sınıflandırmalarda kullanılması, büyük veri setlerinde daha doğru ve objektif değerlendirmeler yapılmasını sağlar. Bu sebeple bu araştırma, MÖ modellerinin öğrenci başarısını sınıflandırmadaki doğruluğunu incelemeyi amaçlamaktadır. MÖ, eğitimde büyük ve karmaşık veri setlerini analiz ederek daha doğru ve objektif değerlendirmeler sağlar. Bu araştırmada, Naive Bayes (NB), Destek Vektör Makineleri (DVM) ve Rastgele Orman (RO) makine öğrenmesi modellerinin sınıflandırma doğrulukları incelenmiştir. Bu araştırma, öğrencilerin başarı durumunu farklı makine öğrenmesi algoritmalarıyla tahmin ederek elde edilen performans ölçütlerini karşılaştırmakta ve model sonuçlarını incelemektedir. Bu yönüyle araştırma, betimsel tarama türünde ve uygulamalı bir çalışmadır. Etnik köken, ebeveyn eğitim düzeyi ve matematik başarısı gibi değişkenlerden oluşan 1000 örneklem içeren bir veri seti kullanılmıştır. Analizler SPSS ve R programı ile gerçekleştirilmiştir. Araştırma bulguları, en yüksek sınıflandırma doğruluğuna sahip modelin RO modeli olduğunu göstermiştir. MÖ modellerinin eğitimde kullanılması öğrenci başarısını, başarısızlık risklerini, öğretim yöntem ve materyallerinin etkinliğini tahmin ederek eğitim kalitesinin iyileştirilmesine katkı sağlayabilir.

Giriş

Eğitimde temel hedeflerden biri öğrencilerin becerilerini geliştirmek ve bilgisini artırmaktır (Hart, 2016). Bundan dolayı, her öğrencinin yetenek ve beceri alanına yönelik eğitim öğretim süreci yürütmek önem taşır (Corno & Snow, 1986). Eğitimin önemi büyük olduğu için öğrencinin başarısı da öğrenme sürecinin kritik bir parçasıdır ve yüksek kaliteli eğitim kurumlarının ölçütlerinden biridir. Başarı, sadece bilişsel faktörlere dayanmaksızın sosyal, ekonomik, bireysel ve çevresel faktörlere dayalı olarak da ölçülebilmektedir (Alamri vd., 2020). Bu faktörlere dayalı toplanan büyük verinin işlenmesinde ise yapay zekâ ve makine öğrenmesi teknolojilerinin katkıları önemli hale gelmiştir (Williamson, 2019; Yang, 2019).

Yapay Zekâ (YZ) ve veri biliminin temelinde yer alan Makine Öğrenmesi (MÖ), bilgisayar biliminin en hızlı gelişen alanlarından biridir ve geniş kapsamlı uygulamalara sahiptir (Jordan & Mitchell, 2015; Shalev-Shwartz & Ben-David, 2014). MÖ, karmaşık örüntüleri algılayabilen ve veriye dayalı kararlar verebilen bilgisayar algoritmalarının genel adıdır. Bu algoritmalara gönderilen verilere öğrenme yeteneği kazandırılıp, kendi kendilerine problem çözme yeteneğine sahip olması sağlanır (Gültepe, 2019). MÖ teknolojisi, karmaşık verilerden gelecekteki sonuçları tahmin etmeyi mümkün kılar. İnsanların karmaşık ve büyük veri kümelerini

*Bu çalışmanın bir kısmı yazar tarafından 3-6 Ekim 2024 tarihleri arasında düzenlenen 9. Uluslararası Eğitim ve Psikolojide Ölçme ve Değerlendirme Kongresi'nde "Öğrenci Başarısını Sınıflandırmak İçin Farklı Makine Öğrenmesi Modellerinin Karşılaştırmalı Analizi" başlığı altında sunulmuştur.

Sorumlu yazar[1] : Tansu Alan, Adiyaman Üniversitesi, Türkiye, tansualan@adiyaman.edu.tr

Alan, T. (2025). A comparative analysis of different machine learning models for classifying student achievement. *Adiyaman Univesity Journal of Educational Sciences, 15*(1), 159-184

hızlı bir şekilde analiz etme yeteneklerini aşan veri miktarlarıyla, makine öğrenmesi önemli bir rol oynar (Türkmenoğlu & Tantuğ, 2014). Veri miktarı ne kadar büyürse büyüsün, makine öğrenimi daha iyi tahminler yapmayı ve sonuçları daha kesinleştirmeyi sağlar (IBM, 2024). Ayrıca, makine öğrenmesi yeni modeller ve tahminler sunabilir, bireylerin ihtiyaçlarını ve durumlarını özelleştirmek için tahminlerde ve önerilerde bulunabilir. Bu özellikleri sayesinde, makine öğrenmesi modelleri tıp, bilim, imalat gibi alanlarda kanıta dayalı karar verme süreçlerini kolaylaştırmak için yaygın olarak kullanılmaktadır (Kourou vd., 2015; Lin vd., 2011; Wu vd., 2017).

MÖ algoritmaları, çeşitli kaynaklardan gelen verileri birleştirerek büyük miktarda veriyi yönetme yeteneği sayesinde tahmin gücünü artırır (Özlüer Başer vd., 2021). Bu algoritmalar, kümeleme, sınıflandırma, tahmin ve kestirim gibi çeşitli niteliklere sahiptir (Ülker & Civalek, 2002) ve denetimli, denetimsiz ve yarı denetimli modellerle makinelere öğretilirler (Brownlee, 2016). Denetimli öğrenmede, girdi ve çıktı değişkenleri birlikte verilir ve amacı yeni verinin sınıflandırılmasına yönelik tahmin sunmaktır (Aydın ve Özkul, 2015; Ülker ve Civalek, 2002). Örneğin öğrenci başarı düzeyinin, başarılı ve başarısız olarak sınıflandırıldığı bir veri setine yeni bir öğrencinin verisi girildiğinde, bu öğrencinin değişkenleri dikkate alınarak başarı durumu (başarılı ya da başarısız) tahmin edilebilir. Denetimsiz öğrenmede ise sadece girdi değeri verilirken çıktı değeri bilinmez (Ülker ve Civalek, 2002). Yarı denetimli öğrenme ise denetimli ve denetimsiz modellerin bir karışımıdır, bazı veriler etiketlenirken bazıları etiketlenmez (Sarıcaoğlu, 2019). Çalışmada kullanılan veri setinde hem girdi (bağımsız) hem de çıktı (bağımlı) değişken olduğundan bazı denetimli makine öğrenme modelleri kullanılarak sonuçlar karşılaştırılmıştır.

MÖ, birçok alanda kullanıldığı gibi eğitim alanında da kullanılabilir ve bu alanda önemli faydalar sağlayabilir (Aydın, 2007). Eğitimde makine öğrenmesinden yararlanıldığında, öğrenme süreçleri öğrencilerin bireysel gelişim hızlarına ve performanslarına göre esnek bir şekilde şekillendirilebilir, böylece daha kişiselleştirilmiş bir öğrenme ortamı oluşturulabilir (Kuch vd., 2020). Bu durum, son dönemde eğitim alanında artan bir ilgi ve hızla gelişen çalışmalarla kendini göstermektedir (Zhou vd., 2018). Eğitim alanında farklı kaynaklardan toplanan verilerin çeşitliliği, büyük veri analizinin önemini artırır. Bu büyük verinin analizi, öğrenci başarısını sınıflama, eğitim çıktılarını tahmin etme, eğitim sürecinde alınması gereken önlemleri önceden belirleme, bireylerin öğrenme yöntem ve stillerini belirleyerek uygun içerikleri önerme ve ölçme değerlendirme sonuçlarına göre eksik kaldığı konuları belirleme gibi birçok alanda kullanılabilir (Tosunoğlu vd., 2021). Bu nedenle makine öğrenmesi, eğitim alanında öğretmenlere destek sağlama, öğrenci performansını tahmin etme ve süreç odaklı değerlendirme gibi çeşitli amaçlarla kullanılabilmektedir (Kucak vd., 2018). Aynı zamanda, puanlayıcı yanlılıklarını ortadan kaldırarak öğrencilerin notlarını daha nesnel biçimde değerlendirme olanağı sunar. Ayrıca öğretmenlere ve eğitim planlayıcılarına, öğrenci akademik başarısını artırmaya yönelik en etkili stratejileri belirleme konusunda da yardımcı olabilir (Alamri vd., 2020). Son dönem araştırmaları, denetimli makine öğrenmesi modellerinin öğrencilerin ders notlarını başarıyla tahmin edebildiğini ve bu teknolojinin, ders başarısızlığı riski taşıyan öğrencileri erken dönemde tanımlama potansiyeline sahip olduğunu göstermektedir. Bu sayede öğrenci değerlendirmeleri daha objektif ve etkili bir şekilde gerçekleştirilebilmektedir (Wu vd., 2018).

Makine öğrenmesinin eğitimde kullanılmasıyla ilgili birçok çalışma yapılmıştır. Alamri vd., (2020) yaptığı çalışmada Portekizli öğrencilerin Portekizce ve matematik notlarını tahmin etmek için RO ve DVM modellerini kullanmıştır. Her iki algoritmanın da yüksek doğruluk oranı verdiğini sonucuna ulaşılmıştır. Vandamme vd. (2007) öğrencilerin demografik, sosyo-ekonomik ve akademik geçmişlerine ilişkin verileri kullanarak, yapay sinir ağları modeli ile öğrencileri farklı başarı gruplarına sınıflandırmış ve başarılarını incelemişlerdir. Şengür ve Tekin (2013) öğrencilerin mezuniyet notlarını tahmin etmek için yapay sinir ağları ve karar ağacı modellerini kullanarak yaptığı çalışmada, yapay sinir ağlarının karar ağaçlarından daha başarılı olduğu bulunmuştur. Hussain vd. (2018) üniversiteye giriş için kullanılan Ortak Giriş Sınavı (CEE) verilerini kullanarak, farklı sınıflandırma ve kümeleme modellerini karşılaştırmışlar ve yapay sinir ağlarının en yüksek doğruluk oranına sahip olduğunu tespit etmişlerdir. Zilyas ve Yılmaz (2023) Türkçe dersi başarısını tahmin etmek için yaptığı çalışmada, K-En Yakın Komşu, Rastgele Orman, Lineer Regresyon, bir araya getirilmiş karar ağaçları regresyonu (Bagged Trees Regression), Gradyen Arttırıcı Regresyon (Gradient Boosting Regressor) ve Karar Ağaçları gibi algoritmalar kullanılmıştır. Sonuç olarak, doğru sınıflandırma açısından rastgele orman modelinin en başarılı olduğu tespit edilmiştir. Quinn ve Gray (2020) tarafından yapılan çalışmada, bir öğrencinin dersi

geçip geçmeyeceğini tahmin etmek için öğrencilerin verileri üzerinde farklı makine öğrenmesi modelleri uygulamıştır. Bu algoritmalar arasında Rastgele Orman, Gradyen Arttırıcı Regresyon, Doğrusal Diskriminant Analizi ve K-En Yakın Komşu bulunmaktadır. Sonuçlar, Rastgele Orman modelinin öğrencilerin ders başarısını tahmin etmek için diğer algoritmalardan daha etkili olduğunu göstermektedir.

İlgili literatür incelendiğinde, eğitim alanında çeşitli makine öğrenmesi modellerinin kullanıldığı ve hangi modelin daha başarılı olduğuna dair bulguların yorumlandığı görülmektedir. Bu çalışmalar, eğitim alanında öğrenci başarısını sınıflandırmak için çeşitli makine öğrenmesi modellerinin kullanılabileceğini göstermektedir. Öğrenci akademik başarısının doğru bir şekilde sınıflandırılması, eğitim verimliliği artırmanın ve bireysel eğitim hizmetleri sağlamanın bir yoludur. Doğru sınıflandırma sayesinde öğrencinin güçlü ve zayıf yönleri belirlenerek, onlara daha etkili destek sağlanması ve başarılarının artırılması mümkün olur. Bu amaçlar doğrultusunda çok çeşitli model veya algoritmalardan yararlanılmakla birlikte karar ağaçları, Rastgele Orman, Destek Vektör Makineleri, Naive Bayes, K-En Yakın Komşu ve lojistik regresyon gibi modeller sıklıkla kullanılan istatistiksel modellerdir (Hussain vd., 2018; Rao vd., 2016; Vijayalakshmi & Venkatachalapathy, 2019; Wu vd., 2018). Bu çalışma kapsamında kullanılan modellerin avantajları, model seçiminde belirleyici bir rol oynamıştır. Rastgele Orman (RO) modeli, yüksek doğruluk oranı ve değişkenler arasındaki karmaşık etkileşimleri modelleme konusundaki güçlü yeteneği sayesinde öne çıkmaktadır. RO, çok boyutlu ve karmaşık veri setlerinde değişkenler arasındaki ilişkileri etkili bir şekilde ortaya koyarak modellerin doğruluğunu artırmaktadır; Ayrıca, aşırı uyum (overfitting) ile başa çıkma yeteneği, bu modeli diğer modellere kıyasla önemli bir avantaj haline getirmektedir (Rebai vd., 2020; Cruz-Jesus vd., 2020). Destek Vektör Makineleri (DVM), güçlü genelleme yetenekleri sayesinde sınıflandırma ve regresyon problemlerinde önemli bir rol oynamakta; eğitim verisi üzerinde yüksek performans sergilerken, daha önce karşılaşmadığı verilerde de benzer başarı düzeyini koruyabilmektedir (Hastie vd., 2017). Naive Bayes yöntemi ise matematiksel olarak anlaşılır yapısı ve kullanışlılığı ile dikkat çekmektedir. Küçük veri setleriyle etkili bir şekilde çalışabilmesi ve hem kategorik hem de sürekli veriler üzerinde uygulanabilir olması (Hamalainen & Vinni, 2011), bu modelin çalışmada kullanılmasını sağlamıştır. Sunmuş oldukları bu avantajlar sayesinde, öğrenci başarısının daha doğru bir şekilde sınıflandırılmasını ve değerlendirilmesini mümkün kılacakları düşüncesiyle bu modeller tercih edilmiştir.

Yapılan çalışmalar incelendiğinde, bazı çalışmalar öğrencilerin sonuçları üzerinde doğrudan veya dolaylı etkilere sahip olan değişkenlerin analizine odaklanmakta iken, bazıları farklı makine öğrenmesi algoritmalarını kullanarak bu değişkenlere dayalı olarak öğrencilerin sonuçlarını tahmin etmeye çalışmaktadır. Bir öğrencinin akademik başarısı, birçok önemli sosyo-ekonomik, okulla ilgili ve kişisel değişkenlere büyük ölçüde bağlı olmakla (Yassein vd., 2017) birlikte bu araştırmada ilgili veri setindeki çeşitli değişkenler yardımıyla Rastgele Orman, Destek Vektör Makinesı ve Naive Bayes modellerinin sınıflandırma performansları incelenmiştir. Çalışmada açık erişimde yer alan bir veri seti kullanılmıştır. Veri setinde bulunan öğrencilere ait matematik başarı ortalaması referans alınarak, öğrenci başarısı ortalama altı ve ortalama üstü olmak üzere iki kategori olarak sınıflandırılmıştır. Araştırmanın amacı, farklı disiplinlerde ve eğitim alanında sıklıkla kullanılan denetimli makine öğrenmesi modellerinden olan Naive Bayes (NB), Rastgele Orman (RO) ve Destek Vektör Makinesi (DVM) modellerinin öğrencinin başarı durumunu ilgili algoritmalardan elde edilen makine öğrenmesi performans ölçütleri ile karşılaştırarak model sonuçlarını incelemektir. Bu amaçla öğrenci başarısının sınıflandırılmasında farklı makine öğrenmesi modellerinin performansları nasıldır? Araştırma problemine cevap aranmıştır.

Yöntem

Araştırmanın Modeli

Araştırmada farklı makine öğrenmesi modellerinin performans ölçütleri yardımıyla sınıflandırma doğrulukları karşılaştırılmıştır. Farklı modellerinin karşılaştırılması sonucunda en başarılı modelin belirlenmesinin amaçlandığı bu araştırma betimsel tarama türünde uygulamalı bir araştırma niteliğindedir. Uygulamalı araştırmalar üretilmiş veya üretilmekte olan kuramsal bilginin denemeli uygulamasıdır ve var olan uygulamaları iyileştirme yönünde somut katkılar sunabilmektedir (Karasar, 2012).

Verilerin Toplanması

Bu çalışmada, Amerika Birleşik Devletleri'nde yaşayan lise öğrencilerinin matematik başarısı üzerindeki ekonomik, bireysel ve sosyal gibi çeşitli değişkenlerin etkisini anlamak için toplanan bir veri seti kullanılmıştır. Her ne kadar veri seti Amerika'daki öğrencilerden oluşsa da çalışmanın temel amacı matematik başarısına odaklanmak değil, makine öğrenmesi modellerinin sınıflandırma doğruluğunu test etmektir. Veri setindeki çeşitli değişkenler, bu doğruluk modellerinin uygulanabilirliğini değerlendirmek ve modellerin performansını karşılaştırmak için kullanılmıştır. Çalışma kapsamında, kamuya açık (açık erişim) 1000 örnekten oluşan bir veri seti (https://www.kaggle.com/datasets/scientist/students-performance-in-exams) internet sitesinden elde edilmiştir. Google'a ait bir platform olan Kaggle, her seviyeden veri bilimcisine yönelik makine öğrenmesi yarışmalarına ev sahipliği yapmanın yanı sıra, kapsamlı bir veri seti koleksiyonu ve bulut tabanlı bir veri bilimi ortamı sağlamaktadır. Bu özellikler, kullanıcıların veri analizi ve model geliştirme süreçlerini etkin bir şekilde yürütmelerine olanak tanımaktadır (Quaranta vd., 2021). Şirketler ve araştırmacılar verilerini paylaşabilirken, veri bilimcileri ve araştırmacılar kendi yaklaşımlarını deneyebilir ve aynı zamanda birbirleriyle rekabet ederek en yüksek yarışma puanını elde etmeye çalışabilirler (Wang vd., 2021). Bir başka ifadeyle veri bilimciler bu platformda gerçek sorunlara gerçek cevaplar aramaktadır (Kuroki, 2023). Veri seti, 482 erkek (%48.2) ve 518 kadın (%51.8) öğrencinin bilgilerini içermektedir. Her grup için aşağıdaki değişkenler dikkate alınmıştır. Bu değişkenler; öğle yemeği durumu, sınav hazırlık kursu, ırk/etnik köken, ebeveynlerin eğitim düzeyi, cinsiyet ve matematik puanı şeklindedir. Araştırmada amaç makine öğrenmesi modellerinin sınıflandırma doğruluğunu karşılaştırmak olduğundan dolayı bu veri seti kullanılmıştır. Analizde dikkate alınan bağımlı ve bağımsız değişkenler Tablo 1'de gösterilmektedir.

Tablo 1. Bağımlı ve Bağımsız Değişkenler

Değişken	Açıklama	Veri Türü	Değişken türü
Cinsiyet	(0=erkek,1=kadın)	Kategorik	Bağımsız
Irk/etnik köken	(0=Grup A,1=Grup B,2=Grup C,3=Grup D,4=Grup E)	Kategorik	Bağımsız
Ebeveyn eğitim düzeyi	(0=üniversite terk/devam,1=ön lisans,2= lise terk/devam,3=lise,4=üniversite,5=yüksek lisans)	Kategorik	Bağımsız
Öğle yemeği	(0=ücretli,1=ücretsiz)	Kategorik	Bağımsız
Test hazırlık kursu	(0=tamamlamamış,1=tamamlamış)	Kategorik	Bağımsız
Matematik Puanı	(0=ortalama altı,1=ortalama üstü)	Kategorik	Bağımlı

Verilerin Analizi

Veri setinde hatalı, eksik ve kayıp verileri tespit edip işlemek; verilerin doğruluğunu ve bütünlüğünü artırmak açısından gereklidir. Araştırma sonuçlarının doğru bir şekilde yorumlanabilmesi için öncelikle verilerdeki kayıp ve uç değerler incelenmiştir. Betimsel özellikleri özetlemek için frekans (yüzde) hesaplanmıştır. Bağımsız değişkenler ile matematik puanları arasındaki ilişkileri incelemek için Pearson ki-kare testi kullanılmıştır. Çalışmadaki gruplara ait sonuçlar arası farkın önemli olup olmadığını incelemek için etki büyüklüğüne de bakılmıştır (Sullivan ve Feinn, 2012). Etki büyüklüğü için Cohen d istatistiği incelenmiştir. Cohen (1988), bu değerin 0,2'den küçük olması durumunda etki büyüklüğünün zayıf, 0,5 olması durumunda orta düzeyde ve 0,8'den büyük olması durumunda ise güçlü olarak değerlendirilebileceğini belirtmektedir. İstatistiksel analizleri gerçekleştirmek için SPSS programı kullanılmıştır. Makine öğrenmesi modellerinin sınıflandırma metrik kriterlerinin sonuçları, R programı kullanılarak analiz edilmiştir. İlgili veri setinin tamamı kullanılarak 5 katlı çapraz doğrulama yapılmıştır. Bir başka deyişle sınıflandırma modellerinin değerlendirilmesi ve modelin eğitilmesi için veri seti 5 parçaya bölünerek modelin doğru sınıflandırma oranları hesaplanmıştır. Çalışma kapsamında kullanılan denetimli MÖ modelleri aşağıda detaylıca açıklanmıştır.

Naive Bayes (NB)

Makine öğrenmesinde kullanılan Naive Bayes modeli, bir öğenin hangi sınıfa ait olduğunu belirlemek için kullanılan bir yaklaşımdır. Bu model, Bayes teoremi temelinde her değişken için sınıf olasılıklarını hesaplar. Bu

olasılıkları kullanarak en yüksek olasılığa sahip sınıf tahmin edilir. "Naive" kelimesi, algoritmadaki her değişkenin diğer tüm değişkenlerden bağımsız olduğunu varsaydığı için kullanılır. Ancak gerçekte, bu varsayımın aksine yüksek başarı oranları elde edilmektedir. Bayes Teoremi şu şekilde ifade edilir (Rish, 2001):

$$P(A\backslash B) = \frac{P(B\backslash A)P(A)}{P(B)} \quad (1)$$

Bu denklemde:

P(A|B): B olayı gerçekleştiğinde, A olayının gerçekleşme olasılığıdır.

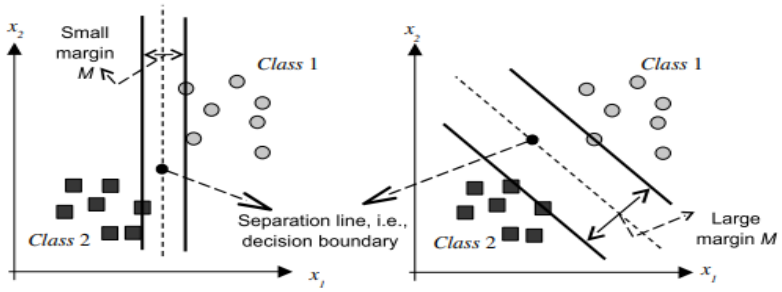
P(B|A): A olayı gerçekleştiğinde, B olayının gerçekleşme olasılığıdır.

P(A): A olayının gerçekleşme olasılığıdır.

P(B): B olayının gerçekleşme olasılığıdır.

Destek Vektör Makinesi (DVM)

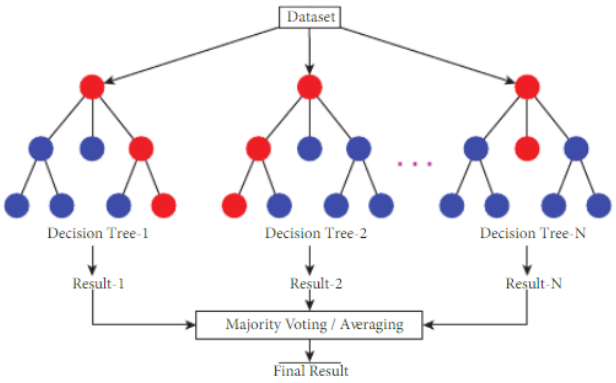
Destek Vektör Makinesi (DVM), makine öğrenmesinde yaygın olarak kullanılan bir sınıflandırma modelidir. Bu model, veri örneklerini sınıflandırmak, yani hangi sınıfa ait olduklarını belirlemek amacıyla kullanılır. Sınıflandırma görevlerinde genellikle belirli veri örneklerini içeren eğitim ve test verileri kullanılır (Duda & Hart, 1973). Son zamanlarda popülerlik kazanan DVM modelinin ana hedefi, hedef değişkenin sınıflarını birbirinden en etkili şekilde ayıracak olan hiper düzlemi bulmaktır. Yani sınıflar arasındaki farkı en büyük yapacak bir hiper düzlem bulmaya çalışır. Bu fark, marj olarak adlandırılır ve marjın genişliği, sınıflandırmanın doğruluğunu doğrudan etkiler. DVM'in amacı, bu marjı maksimize eden, yani sınıfları en iyi şekilde ayıran hiper düzlemi bulmaktır. Bunu yaparken DVM, hata fonksiyonunu minimize eden bir eğitim süreci uygular. Bu süreç, iteratif bir şekilde en iyi hiper düzlemi bulana kadar devam eder. Hata fonksiyonu, modelin yaptığı hataları ölçer ve bu hataları en aza indirmek için hiper düzlem sürekli olarak güncellenir ve optimum bir hiper düzlem oluşturur (Cristianini & Shawe-Taylor, 2000 & Vapnik, 2013). Destek vektör makinası algoritmasının sınıflandırma mantığı Şekil 1'de gösterilmiştir (Huang vd., 2006).



Şekil 1. Destek vektör makinası algoritması

Rastgele Orman (RO)

Rastgele Orman (RO) birçok karar ağacı kullanarak sınıflandırmada güçlü ve güvenilir tahminler yapan bir makine öğrenmesi algoritmasıdır. RO algoritması, birden fazla karar ağacı oluşturarak ve bunların sonuçlarını birleştirerek daha güçlü ve doğru tahminler yapmayı hedefler. Bu yaklaşımla, modelin performansı artırılır ve genellikle tek bir karar ağacından daha iyi sonuçlar elde edilir (Breiman, 2001). RO yaklaşımı, tek bir karar ağacına kıyasla aşırı öğrenme (overfitting) riskini önemli ölçüde azaltır, aykırı değerleri tanımada etkili olmaktadır. Ayrıca, veri setindeki en önemli değişkeni belirlemek için en uygun modellerden biridir (Iwendi vd., 2020). Rastgele orman algoritmasının sınıflandırma mantığı Şekil 2'de gösterilmiştir (Khan vd., 2021).



Şekil 2. Rastgele orman algoritması

Model Performanslarının Değerlendirilme Kriterleri

Sınıflandırma Matrisi, sınıflandırma modellerinin ne kadar iyi çalıştığını değerlendirmek, model doğruluğunu belirlemek ve hataları tespit etmek için kullanılan bir araçtır. Bu çalışmada modellerin sınıflandırma performanslarını ölçmek için kullanılacak sınıflandırma matrisi, kriterler ve bu kriterlerle ilgili formüller Tablo 2 ve Tablo 3'te verilmiştir.

Tablo 2. Sınıflandırma Matrisi

	Tahmin	
Referans	Ortalama altı	Ortalama Üstü
Ortalama altı	Doğru Negatifler (TN)	Yanlış Pozitifler (FP)
Ortalama Üstü	Yanlış Negatifler (FN)	Doğru Pozitifler (TP)

Tablo 2'deki matris incelendiğinde, matrisin satırları referans değerlerini (gerçek sonuçlar), sütunları ise tahmin edilen değerleri (modelin tahminleri) temsil eder. Doğru negatifler (TN) ortalama altı tahmin edilmiş ve gerçekten ortalama altı olan gözlemler, yanlış pozitifler (FP) ortalama üstü tahmin edilmiş ancak gerçekte ortalama altı olan gözlemler, yanlış negatifler (FN) ortalama altı tahmin edilmiş ancak gerçekte ortalama üstü olan gözlemler, doğru pozitifler (TP) ortalama üstü tahmin edilmiş ve gerçekten ortalama üstü olan gözlemler olarak açıklanabilir.

Tablo 3. Sınıflandırma Performans Ölçütleri

Ölçüt	Formül/Yaklaşım
Doğruluk (Accuracy)	$\frac{TP+TN}{TP+TN+FN+FP}$
Duyarlılık (Sensitivity)	$\frac{TP}{TP+FN}$
Seçicilik (Specificity)	$\frac{TN}{TN+FP}$
Youden İndeksi (Youden's index)	$\frac{TP}{TP+FN}+\frac{TN}{TN+FP}-1$
Pozitif Tahmin Değeri (Positive predictive value- PPV)	$\frac{TP}{TP+FP}$
Negatif Tahmin Değeri (Negative predictive value- NPV)	$\frac{TN}{TN+FN}$
F1-Score	$\frac{2*PPV*\text{Sensitivity}}{PPV+\text{Sensitivity}}$

Performans Değerlendirme Ölçütleri

Makine öğrenmesi algoritmaları ile, doğru sınıflandırmayı dengelemek ve doğru yaklaşımı belirlemek için doğruluk, duyarlılık, seçicilik, Youden indeksi, F1-Skoru, PPV ve NPV gibi performans göstergeleri kullanılır. Ölçütlerin özellikleri aşağıdaki gibidir (Racz vd., 2019):

1.Doğruluk (Accuracy): Eğitilmiş makine öğrenmesi modelinin doğru sınıflandırdığı verilerin oranı olarak bilinen doğruluk, aynı zamanda doğru tahminlerin tüm sınıflar boyunca toplam tahminlere oranı olarak da tanımlanabilir. Yani, modelin genel performansını gösterir. Yüksek bir doğruluk, modelin hem pozitif hem de negatif sınıflandırmalarında başarılı olduğunu gösterir.

2.Duyarlılık (Sensitivity): Modelin duyarlılığı, doğru pozitif durumlarını ne kadar iyi yakaladığını gösterir. Öte yandan, artan duyarlılıkla birlikte yanlış pozitif hatalar da ortaya çıkabilir.

3.Seçicilik (Specificity): Seçicilik derecesi, modelin doğru negatifleri ne kadar iyi tespit ettiğini gösterir. Ancak, seçicilikle birlikte yanlış negatif hataların artma olasılığı da vardır.

4.Youden's İndeksi: Bu indeks, bir modelin hem duyarlılık hem de seçicilik açısından ne kadar iyi performans gösterdiğini özetler. 1'e yakın bir değer, modelin her iki metriği de iyi başardığını gösterir.

5.Pozitif Tahmin Değer (PPV): Modelin pozitif olarak sınıflandırdığı durumların ne kadarının gerçekten pozitif olduğunu gösterir. Pozitif veri noktalarının doğru bir şekilde tahmin edilme oranını gösteren önemli bir parametredir. Ayrıca, sınıflandırma algoritmalarının performansını değerlendirmek için kullanılabilir. Ayrıca bu metrik, modelin yanlış pozitif tahminler yapma eğilimini değerlendirir.

6.Negatif Tahmin Değer (NPV): Modelin negatif olarak sınıflandırdığı durumların ne kadarının gerçekten negatif olduğunu gösterir. Yani, modelin yanlış negatif tahminler yapma eğilimini değerlendirir. Yüksek NPV, modelin negatif tahminlerinin doğruluğunu gösterir. Bu değer, yanlış negatif hataları azaltmak istediğiniz durumlarda çok önemlidir.

7.F1-Skoru (F1-Score): $\beta = 1$ Parametre değerinden adını alan bu ölçüt, parametrik F-ölçüleri ailesinin en yaygın kullanılan üyesidir. Sınıflandırma modellerinin performansını değerlendirmek için kullanılan, duyarlılık ve pozitif tahmin (PPV) değeri arasında denge sağlayan bir metriktir. F1-Skoru, bu iki metriği dengeler. Eğer modelin hem yanlış pozitif tahminleri (başarılı öğrencileri yanlış tahmin etme) hem de yanlış negatif tahminleri (başarısız öğrencileri kaçırma) düşükse, F1-Skoru yüksek olur.

Etik Bildirim

Etik Kurul İzin Bilgisi: Bu araştırma, Adıyaman Üniversitesi Bilimsel Araştırma ve Yayın Etiği Sosyal ve Beşeri Bilimler kurulunun 29/12/2025 tarihli 62 sayılı kararı ile alınan izinle yürütülmüştür.

Bulgular

Yapılan incelemeye göre veri setinde uç değer ve kayıp veri olmadığı sonucuna ulaşılmıştır. Değişkenlere ilişkin betimsel istatistikler Tablo 4'te, öğrenci başarısını sınıflandırmada kullanılan modellerin performans ölçüm sonuçları ise Tablo 5'te sunulmuştur.

Tablo 4. Değişkenlere İlişkin Betimsel İstatistikler

		Grup				
		Ortalama Altı	Ortalama Üstü			
Değişkenler	Kategori	n(%)	n(%)	Etki Büyüklüğü (EB)	Etki Derecesi	p
Cinsiyet	Kız	297 (58.6)	221 (44.8)	0.138	Küçük Etki	<0.001*
	Erkek	210 (41.4)	272 (55.2)			
Irk/Etnik Köken	Grup A	58 (11.4)	31 (6.3)	0.211	Orta Etki	<0.001*
	Grup B	112 (22.1)	78 (15.8)			
	Grup C	179 (35.3)	140 (28.4)			
	Grup D	116 (22.9)	146 (29.6)			
	Grup E	42 (8.3)	98 (19.9)			
Ebeveyn Eğitim Düzeyi	Üniversite Terk/Devam	107 (21.1)	119 (24.1)	0.115	Küçük Etki	0.021*
	Ön Lisans	109 (21.5)	113 (22.9)			
	Lise Terk/Devam	98 (19.3)	81 (16.4)			
	Lise	117 (23.1)	79 (16.0)			
	Yüksek Lisans	24 (4.7)	35 (7.1)			
	Üniversite	52 (10.3)	66 (13.4)			
Öğle Yemeği	İndirimli	262 (51.7)	383 (77.7)	0.271	Orta Etki	<0.001*
	Ücretsiz	245 (48.3)	110 (22.3)			
Test Hazırlık Kursu	Tamamlamamış	360 (71.0)	282 (57.2)	0.114	Küçük Etki	<0.001*
	Tamamlamış	147 (29.0)	211 (42.8)			

**Pearson Ki kare (Pearson Chi-Square)*

Veri setinde yer alan 1000 örneklem cinsiyet değişkeni açısından incelendiğinde, kız öğrencilerin matematik başarı puanlarının ortalamanın altında olduğu kişi sayısı 297 (%58,6), ortalamanın üzerinde ise 221 (%44,8) olarak bulunmuştur. Erkek öğrencilerde ise matematik başarı puanları ortalamanın altında olanlar 210 (%41,4), ortalamanın üzerinde olanlar ise 272 (%55,2) olarak hesaplanmıştır. Bu dağılıma bakıldığında, ortalamanın altında kalan öğrenci sayısı 507 (%50,7) ile ortalamanın üstünde olan öğrenci sayısı 493 (%49,3) arasında küçük bir fark olduğu görülmektedir. Her iki sınıf da birbirine yakın büyüklüklerde temsil edildiği için, sınıflandırma modelleri bu veri seti üzerinde dengesiz sınıfların yol açabileceği olumsuz etkileri minimize eder. Sınıf dengesizliği genellikle, bir sınıfın diğerine göre çok fazla veya çok az olduğu durumlarda büyük bir sorun teşkil eder. Bu dengesizlik, modelin eğitim sürecini olumsuz etkileyerek modelin daha fazla örnek olan sınıfa ağırlık vermesine ve az sayıda örneği olan sınıfı doğru bir şekilde tanımakta zorlanmasına neden olur (Buda vd., 2018). Cinsiyet ve matematik puanları arasında anlamlı bir fark bulunmuştur ($p<0.001$), bu fark küçük bir etkiye sahiptir (EB=0.138). Benzer şekilde, ırk/etnik kökeni, ebeveyn eğitim seviyesi, öğle yemeği ve test hazırlık kursu değişkenleri ile matematik başarı puanı arasında istatistiksel olarak anlamlı bir fark olduğu sonucuna ulaşılmıştır ($p<0.001$). Etki büyüklükleri analiz edildiğinde, ırk/etnik köken ve öğle yemeği değişkenlerinin etki büyüklükleri sırasıyla 0.211 ve 0.271 olarak bulunmuş, bu değişkenlerin orta düzeyde bir etkisi olduğunu göstermiştir. Ebeveynlerin eğitim düzeyi ve test hazırlık kursu değişkenlerinin etki büyüklükleri sırasıyla 0.115 ve 0.114 olarak bulunmuş, bu değişkenlerin küçük düzeyde bir etkisi olduğunu göstermiştir.

Tablo 5. Öğrenci Başarısını Sınıflandırmada Kullanılan Modellerin Performans Ölçüm Sonuçları

Model	NB (95% Güv.A.)	DVM (95% Güv.A.)	RO (95% Güv.A.)
Doğruluk	0.65 (0.62-0.68)	0.66 (0.63-0.69)	0.68 (0.65-0.71)
Duyarlılık	0.72 (0.67-0.76)	0.53 (0.48-0.57)	0.64 (0.60-0.68)
Seçicilik	0.59 (0.54-0.63)	0.78 (0.74-0.82)	0.72 (0.68-0.76)
Youden indeksi	0.30 (0.22-0.39)	0.30 (0.22-0.39)	0.36 (0.27-0.44)
PPV	0.62 (0.58-0.66)	0.69 (0.64-0.74)	0.68 (0.63-0.72)
NPV	0.69 (0.64-0.73)	0.64 (0.60-0.68)	0.68 (0.64-0.72)
F1-Skor	0.66 (0.63-0.69)	0.60 (0.57-0.63)	0.66 (0.63-0.69)

NB: Naive Bayes; DVM: Destek Vektör Makinesı; RO: Rastgele Orman; PPV: Pozitif tahmin değeri; NPV: Negatif tahmin değeri; Güv. A.:Güven Aralığı

Tablo 5'te, NB, SVM ve RO modellerinin sınıflandırma çıktılarını karşılaştırmaktadır. Model performansını karşılaştırmak için doğruluk, duyarlılık, seçicilik, Youden indeksi, PPV, NPV ve F1 skoru ölçütleri kullanılmıştır. Sınıflandırma doğruluğu dikkate alındığında RO'nın öğrenci başarısını sınıflandırmada diğer ML modellerinden daha iyi performans gösterdiği sonucuna ulaşılmıştır. 0.68 (0.65-0.71). En düşük sınıflandırma doğruluğu ise 0.65 (0.62-0.68) ile NB modelinden elde edilmiştir.

Doğru pozitif durumlarını (başarılı öğrencileri) ne kadar iyi yakaladığını gösteren duyarlılık sonuçları incelendiğinde, en yüksek duyarlılığın 0.72 (0.67-0.76) ile NB, en düşük duyarlılığın ise 0.53 (0.48-0.57) ile DVM ile bulunduğu sonucuna ulaşılmıştır. RO ise 0.64 (0.60-0.68) ile ikinci sırada bulunmaktadır.

Modellerin doğru negatifleri (başarısız öğrencileri) ne kadar iyi tespit ettiğini gösteren seçicilik değerleri incelendiğinde, en yüksek seçiciliğin 0.78 (0.74-0.82) ile DVM, en düşük 0.59 (0.54-0.63) ile NB modelinden elde edilmiştir. RO ise 0.72 (0.68-0.76) ile ikinci sırada bulunmaktadır.

Modellerin hem duyarlılık hem de seçicilik açısından ne kadar iyi performans gösterdiğini veren Youden indeksi incelendiğinde en yüksek değerin 0.36 (0.27-0.44) ile RO olduğu sonucuna ulaşılmıştır. Bu indeks DVM ve NB için ise 0.30 (0.22-0.39) olarak bulunmuştur.

Modellerin pozitif olarak sınıflandırdığı durumların ne kadarının gerçekten pozitif olduğunu gösteren PPV değeri incelendiğinde 0.69 (0.64-0.74) ile en yüksek DVM, en düşük ise 0.62 (0.58-0.66) ile NB modelinden elde edilmiştir. RO modelinde ise 0.68 (0.63-0.72) olarak bulunmuştur.

Modellerin negatif olarak sınıflandırdığı durumların ne kadarının gerçekten negatif olduğunu gösteren NPV değeri incelendiğinde 0.69 (0.64-0.73) ile en yüksek DVM, en düşük ise 0.64 (0.60-0.68) ile NB modelinden elde edilmiştir. RO modelinde ise 0.68 (0.64-0.72) olarak bulunmuştur.

Sınıflandırma modellerinin performansını değerlendirmek için kullanılan, duyarlılık ve pozitif tahmin (PPV) değeri arasında denge sağlayan F1 skoru NB ve RO modelinde eşit olup bu değer 0.66 (0.63-0.69) iken DVM'de 0.60 (0.57-0.63) olarak bulunmuştur.

Sonuç ve Tartışma

Öğrenci performansının önceden tahmin edilmesi eğitim alanında önemli bir konudur. Öğrencilerin başarı durumu, başarısızlık riskleri gibi durumunun analiz edilmesi eğitim kurumlarının genel performansını artırmada katkı sağlayabilir. Bu analizler sayesinde okullar, öğrencilere ihtiyaç duydukları desteği sağlayarak daha iyi sonuçlar elde edebilir ve başarı oranlarını artırabilir. Bu çalışmada, öğrenci başarısı "ortalamanın üstünde" ve "ortalamanın altında" olarak kategorize edilip sınıflandırma doğrulukları incelenmiştir. Doğru sınıflandırma için çeşitli makine öğrenmesi modellerinin performans metrikleri incelenmiştir. Çalışmada öğrencilerin matematik başarısının sınıflandırma performansını değerlendirmek için NB, DVM ve RO makine öğrenmesi modelleri kullanılmıştır. Modeller doğruluk, F1 skoru, duyarlılık, seçicilik, Youden indeksi, NPV ve PPV performans değerlendirme ölçütleri açısından incelenmiştir.

MÖ modelleri test edildiğinde, en yüksek doğruluk, Youden indeksi ve F1 skoruna sahip modelin RO modeli olduğu sonucuna ulaşılmıştır. RO modelinin öğrencilerin matematikteki başarılarının doğru sınıflandırma

oranı %68 olup, bu oran 0.65-0.71 güven aralığı içindedir. RO modelinin F1-skoru 0.66 (0.63-0.69), duyarlılığı 0.64 (0.60-0.68), seçiciliği 0.72 (0.68-0.76), Youden indeksi 0.36 (0.27-0.44), pozitif tahmin değeri 0.68 (0.63-0.72) ve negatif tahmin değeri 0.68 (0.64-0.72) olarak bulunmuştur. Yapılan çalışmalarda, RO modelinin doğru sınıflandırma oranı %45 ile %99 arasında değişmektedir (Gunawan vd, 2021; Alamri vd., 2020; Nurhachita ve Negara, 2021; Rodrigues vd., 2020). Özellikle sınıflandırmanın hem doğruluk hem de genel performansını göz önünde bulundurulduğunda RO modeli tercih edilebilir ve daha fazla yordayıcı değişken eklenerek modelin genel performansı arttırılabilir. Elde edilen sonuçlar, etnik köken ve sosyo ekonomik düzeyi ölçen öğle yemeği değişkenlerinin öğrencilerin akademik başarısını tahmin etmek için önemli değişkenler olduğunu da ortaya koymaktadır. Diğer modellerle karşılaştırıldığında, RO modeli sınıflandırma ve tahminde daha yüksek doğruluk oranları ürettiği için bazı çalışmalar tarafından da önerilmektedir (Jayaprakash vd., 2020; Rao vd., 2016). Ağaç tabanlı tahmin modelleri, özellikle RO modeli eğitim alanında akademik başarıyı tahmin etme konusunda geleneksel tekniklere güçlü bir alternatif sunmaktadır. Bu modeller, değişkenler arasındaki doğrusal olmayan etkileşimleri yakalayabilmekte ve bu sayede modelin tahmin doğruluğunu artırmaktadır (Geurts vd., 2009; Golino vd., 2014).

Eğitim alanında yapılan benzer çalışmalar incelendiğinde, bazı araştırmalar (Gunawan vd., 2021; Alamri vd., 2020; Yağcı, 2022) RO modelinin daha yüksek sınıflandırma doğruluğu sergilediğini ve böylece bu çalışmanın bulgularını desteklediğini göstermektedir. Öte yandan, bazı çalışmalarda ise NB ve DVM modellerinin (Nurhachita & Negara, 2021; Rodrigues vd., 2020; Nurpiana & Wijayanto, 2022) daha üstün sınıflandırma performansı sağladığı bildirilmektedir. Modellerin tahmin doğruluğunun sonuçları, tahmin sürecinde kullanılan değişkenlere bağlı olarak değişebilmektedir. DVM modeliyle öğrencilerin performansını tahmin etmede önemli değişkenler öğrenci ilgisi, çalışma davranışları ve aile desteği gibi psikososyal değişkenler olabilmektedir (Sembiring vd., 2011; Shahiri vd., 2015). Benzer çalışmalarda NB modelinin daha yüksek sınıflandırma doğruluğu göstermesi, bu çalışmada kullanılan verinin örneklem büyüklüğündeki farklılıktan kaynaklanabilir. Örneklem önemli düzeyde artırıldığında NB modeli daha yüksek doğru sınıflandırma performansı gösterebilmektedir (Koyuncu, 2018).

Bu çalışmanın sonuçları, başarısız öğrencilerin belirlenmesinde Destek Vektör Makinesi (DVM) modelinin en yüksek seçicilik değerine ulaştığını göstermektedir. Bu bulgu, Gray vd (2014) çalışmasıyla uyumlu olup, literatürde DVM modelinin risk altındaki öğrencileri yüksek doğrulukla tanımlamada etkili bir yöntem olarak öne çıktığını desteklemektedir. Dolayısıyla, başarısız öğrenci belirleme süreçlerinde DVM modelinin kullanımı daha uygun bir tercih olarak öne çıkmaktadır.

NB modeli, yüksek duyarlılık ile başarılı öğrencileri yakalamada iyi performans göstermesine rağmen, genel doğruluk, seçicilik ve PPV gibi diğer ölçütlerde daha düşük performans göstermektedir. Bu model, özellikle başarılı öğrencileri yakalamak önemli olduğunda düşünülebilir, ancak genel performans açısından diğer modellerin gerisinde kalmaktadır.

Bu çalışmada, sınıflandırma işlemlerinde kullanılan üç farklı modelle elde edilen sonuçlar karşılaştırılmıştır. Alan yazın incelendiğinde, öğrencilerin başarılarını çeşitli değişkenler açısından sınıflandırmak için birçok alternatif model bulunduğu görülmektedir. Başka çalışmalarda araştırmacılar, KNN, karar ağaçları ve lojistik regresyon gibi farklı modelleri kullanarak, benzer veya farklı koşullar altında bu modellerin performanslarını değerlendirebilir.

Sosyo-demografik değişkenler, okul değişkenleri, yaşam koşulları ve diğer davranışsal değişkenler modellerin tahmin gücünü artırmaktadır (Cortez & Silva, 2008). Makine öğrenmesi modellerinin etkisini daha doğru değerlendirebilmek için, başka çalışmalarda değişken sayısı artırılabilir. Böylelikle makine öğrenmesi modellerinin öğrenci başarısını daha kapsamlı ve hassas bir biçimde analiz etmesini sağlayacak ve eğitim yöneticilerine, öğrencileri desteklemek için daha etkili veri odaklı stratejiler geliştirme imkânı sunacaktır. Böylece, öğrenci başarılarını etkileyen çok boyutlu faktörlerin daha ayrıntılı değerlendirilmesi, makine öğrenmesi modellerinin eğitim alanında sağladığı faydaların daha güçlü bir şekilde test edilmesini mümkün kılacaktır.

Makine öğrenmesi modelleri, yalnızca sınıflandırma sonuçlarını değil aynı zamanda hangi değişkenlerin başarısızlık riskini azalttığını veya başarıyı artırdığını ortaya koyarak eğitim yöneticilerine öğrenciyi destekleme

konusunda rehberlik edebilir. Öğrencilerin başarı düzeylerini önceden tahmin edebilmek amacıyla veri tabanlı modellerin kullanımı, makine öğrenmesi bulgularının uygulamadaki değerini artırarak (Rajendran vd., 2022) eğitimde kalite iyileştirme çalışmalarına yönelik daha kapsamlı bir yol haritası sunabilir. Öğrencinin başarısı makine öğrenmesi modeli tarafından düşük olarak sınıflandırıldığında, bu model öğrencinin hangi değişkenlerden etkilendiğini belirleyebilmelidir. Model, öğrencinin başarısını etkileyen değişkenleri analiz ederek, öğrenciye özel öneriler geliştirebilir. Örneğin sosyoekonomik düzey, motivasyon ve stres düzeyi gibi değişkenlerin öğrencilerin başarı sınıflandırması üzerinde etkili olduğu belirlenirse, düşük sosyoekonomik düzeye sahip öğrenciler için mali destek programları ve ücretsiz ek dersler önerilebilir. Motivasyon eksikliği yaşayan öğrenciler için motivasyon artırıcı seminerler ve hedef belirleme atölye çalışmaları önerilebilir. Ayrıca, yüksek stres düzeyine sahip öğrenciler için psiko-sosyal destek hizmetleri sunmak, stres yönetimi ve rahatlama teknikleri konusunda rehberlik sağlamak, öğrenme süreçlerini olumlu yönde etkileyebilir. Bu tür uygulamalar, öğrencilerin akademik başarılarını artırmalarına yardımcı olmayı amaçlar. Böylece, makine öğrenmesi sonuçları, eğitim müdahaleleri için bir temel oluşturur ve öğrencinin başarısını artırmaya yönelik destek sistemlerinin daha etkili hale gelmesine katkıda bulunur.

Bu çalışma, başarı düzeyi düşük öğrencileri belirlemeye, başarısızlıkları erken düzeltmeye ve öğrencilerin performansını iyileştirmek için uygun önlemler almaya yardımcı olabilecek çeşitli makine öğrenmesi modellerini karşılaştırdığı için eğitime, alana ve araştırmacılara kaynak olarak uygun olabilir.

Bu çalışmanın dikkate değer bir sınırlaması, örneklemin yalnızca bir bölgede yaşayan katılımcılardan oluşmasıdır. Bu nedenle, elde edilen sonuçların başka bölgelere genellenebilirliği sınırlı olabilir. Önerilen modelin diğer kurumlara uyarlanması ve uygulanması, farklı bilim alanlarındaki ve çeşitli eğitim stratejilerine sahip okullardaki öğrenciler arasındaki gözlemlenen heterojenlik nedeniyle farklı sonuçlara yol açabilir.

References

Alamri, L., S. Almuslim, R., S. Alotibi, M., K. Alkadi, D., Ullah Khan, I., & Aslam, N. (2020). Predicting Student Academic Performance using Support Vector Machine and Random Forest [Paper Presentation]. 3rd International Conference on Education Technology Management. London, United Kingdom.

Aydın, S. (2007). Veri madenciliği ve Anadolu Üniversitesi uzaktan eğitim sisteminde bir uygulama. (Publication No. 220873) [Yayınlanmamış Doktora tezi]. Anadolu Üniversitesi, Eskişehir.

Aydın, S. ve Özkul, AE (2015). Veri madenciliği ve anadolu üniversitesi açık öğretim uygulaması bir uygulama [A data mining application and Anadolu University open education system: A case study]. *Eğitim ve Öğretim Araştırmaları Dergisi, 4* (3), 36-44. http://www.jret.org/FileUpload/ks281142/File/05a.sinan_aydin.pdf

Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5-32. http://dx.doi.org/10.1023/A:1010933404324.

Brownlee, J. (2016). Machine learning mastery with python. *Machine Learning Mastery Pty Ltd*, *527*, 100-120.

Buda, M., Maki, A., & Mazurowski, M. A. (2018). A systematic study of the class imbalance problem in convolutional neural networks. *Neural networks*, *106*, 249-259. https://doi.org/10.1016/j.neunet.2018.07.011

Cristianini, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods*. Cambridge University Press.

Cruz-Jesus, F., Castelli, M., Oliveira, T., Mendes, R., Nunes, C., Sa-Velho, M., & Rosa-Louro, A. (2020). Using artificial intelligence methods to assess academic achievement in public high schools of a European Union country. *Heliyon*, *6*(6). https://doi.org/10.1016/j.heliyon.2020.e04081

Cohen, J. (1988). *Statistical power analysis fort he behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Corno, L., & Snow, R. E. (1986). *Adapting teaching to individual differences among learners. In Handbook of research on teaching* (pp. 605-629). Chicago, IL: Rand McNally.

Cortez, P., & Silva, A. M. G. (2008). Using data mining to predict secondary school student performance. https://hdl.handle.net/1822/8024

Duda, R. O., & Hart, P. E. (1973). *Pattern classification and scene analysis*. New York: A Wiley-Interscience Publication.

Geurts, P., Irrthum, A., & Wehenkel, L. (2009). Supervised Learning with Decision Tree-Based Methods in Computational and Systems Biology. Molecular BioSystems, 5, 1593-1605.

Gunawan, G., Hanes, H., & Catherine, C. (2021). C4. 5, K-nearest neighbor, naïve bayes, and random forest algorithms comparison to predict students' on time graduation. *Indonesian Journal of Artificial Intelligence and Data Mining, 4*(2), 62-71.

Golino, H. F., Gomes, C. M. A., & Andrade, D. (2014). Predicting academic achievement of high-school students using machine learning. *Psychology*, *5*(18), 2046-2057. 10.4236/psych.2014.518207

Gray, G., McGuinness, C., & Owende, P. (2014). An application of classification models to predict learner progression in tertiary education. In *2014 IEEE international advance computing conference (IACC)* (pp. 549-554). IEEE.

Gültepe, Y. (2019). Makine öğrenmesi algoritmaları ile hava kirliliği tahmini üzerine karşılaştırmalı bir değerlendirme [A comparative evaluation on air pollution forecasting with machine learning algorithms]. *Avrupa Bilim ve Teknoloji Dergisi,* (16), 8-15. https://doi.org/10.31590/ejosat.530347.

Hämäläinen, W., & Vinni, M. (2011). Classifiers for educational data mining. *Handbook of Educational Data Mining, Chapman & Hall/CRC Data Mining and Knowledge Discovery Series*, 57-71.

Hart, S. A. (2016). Precision education initiative: Moving toward personalized education. *Mind Brain and Education, 10*(4), 209–211. https://doi.org/10.1111/mbe.12109.

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2, pp. 1-758). New York: springer.

Huang, T. M., Kecman, V., & Kopriva, I. (2006). *Kernel based algorithms for mining huge data sets: Supervised, semi-supervised, and unsupervised learning*. Heidelberg: Springer.

Hussain, S., Dahan, N. A., Ba-Alwib, F. M., & Ribata, N. (2018). Educational data mining and analysis of students' academic performance using WEKA. *Indonesian Journal of Electrical Engineering and Computer Science*, *9*(2), 447-459. http://doi.org/10.11591/ijeecs.v9.i2.pp447-459.

IBM. (2024, June 15). *Makine Öğrenmesi.* https://www.ibm.com/docs/tr/rpa/21.0?topic=classification-text-algorithms.

Iwendi, C., Bashir, A. K., Peshkar, A., Sujatha, R., Chatterjee, J. M., Pasupuleti, S., Mishra, R., Pillai, S., Jo, O. (2020). COVID-19 patient health prediction using boosted random forest algorithm. *Frontiers in public health, 8*, 357. https://doi.org/10.3389/fpubh.2020.00357.

Jayaprakash, S., Krishnan, S., & Jaiganesh, V. (2020). Predicting students academic performance using an improved random forest classifier [Paper Presentation]. International Conference On Emerging Smart Computing and Informatics (ESCI), 238-243, IEEE. Pune, India

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science, 349*(6245), 255–260. https://doi.org/10.1126/science.aaa8415.

Karasar, N. (2012). *Bilimsel araştırma yöntemi [Scientific research method]*. Ankara: Nobel Akademik Yayıncılık.

Khan, M. Y., Qayoom, A., Nizami, M. S., Siddiqui, M. S., Wasi, S., & Raazi, S. M. K. U. R. (2021). Automated prediction of good dictionary examples (GDEX): A comprehensive experiment with distant supervision, machine learning, and word embedding-based deep learning techniques. *Complexity*, *2021*(1), 2553199. https://doi.org/10.1155/2021/2553199.

Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. *Computational and Structural Biotechnology Journal,* 13, 8–17. https://doi.org/10.1016/j.csbj.2014.11.005.

Koyuncu, İ. (2018). Öğrencilerin PISA matematik başarılarının yordanmasında veri madenciliği yöntemlerinin karşılaştırılması. (Yayınlanmamış Doktora Tezi). Hacettepe Üniversitesi, Ankara.

Kucak, D., Juricic, V. & Dambic, G. (2018). Machine learning in education- a survey of current research trends. 29th DAAAM International Symposium, Vienna, Austria.

Kuch, D., Kearnes, M., & Gulson, K. (2020). The Promise of precision: Datafication in medicine, agriculture and education. *Policy Studies, 41*(5), 527-546. https://doi.org/10.1080/01442872.2020.1724384.

Kuroki, M. (2023). Integrating data science into an econometrics course with a Kaggle competition. *The Journal of Economic Education*, *54*(4), 364-378. http://hdl.handle.net/10.1080/00220485.2023.2220695

Lin, W. Y., Hu, Y. H., & Tsai, C. F. (2011). Machine learning in financial crisis prediction: A Survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 42*(4), 421-436. 10.1109/TSMCC.2011.2170420.

Nurhachita, N. ve Negara E. S. (2021). A comparison between deep learning, naïve bayes and random forest for the application of data mining on the admission of new students. *IAES International Journal of Artificial Intelligence (IJ-AI) 2*(10), 324-331. http://doi.org/10.11591/ijai.v10.i2.pp324-331.

Nurpiana, A., & Wijayanto, A. W. (2022). Comparison of models for classification of learning achievement of middle school students in ındonesia in 2019 using the support vector machine algorithm, conditional inference trees, and random forest. *Jurnal Matematika, Statistika dan Komputasi*, *18*(3), 447-455. https://doi.org/10.20956/j.v18i3.19208.

Özlüer Başer, B., Yangın, M. & Sarıdaş, E. S. (2021). Makine öğrenmesi teknikleriyle diyabet hastalığının sınıflandırılması [Classification of diabetes using machine learning techniques]. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 25* (1), 112-120. https://doi.org/10.19113/sdufenbed.842460.

Quaranta, L., Calefato, F., & Lanubile, F. (2021, May). KGTorrent: A dataset of python jupyter notebooks from kaggle. In *2021 IEEE/ACM 18th International Conference on Mining Software Repositories (MSR)* (pp. 550-554). IEEE. https://doi.org/10.1109/MSR52588.2021.00072

Quinn, R.J., & Gray, G. (2020). Prediction of student academic performance using Moodle data from a Further Education setting, *Irish Journal of Technology Enhanced Learning, 5*(1), 17-29. https://doi.org/10.22554/ijtel.v5i1.57.

Rácz, A., Bajusz, D., & Héberger, K. (2019). Multi-level comparison of machine learning classifiers and their performance metrics. *Molecules*, *24*(15), 2811 https://doi.org/10.3390/molecules24152811.

Rajendran, S., Chamundeswari, S., & Sinha, A. A. (2022). Predicting the academic performance of middle-and high-school students using machine learning algorithms. *Social Sciences & Humanities Open*, *6*(1), 100357. https://doi.org/10.1016/j.ssaho.2022.100357

Rao, K. P., Sekhara, M. C., & Ramesh, B. (2016). Predicting learning behavior of students using classification techniques. *International Journal of Computer Applications*, *139*(7), 15-19. https://www.ijcaonline.org/research/volume139/number7/rao-2016-ijca-909188.pdf.

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

Rish, I. (2001). An empirical study of the naive Bayes classifier. In *IJCAI 2001 Workshop On Empirical Methods in Artificial Intelligence*, 3(22), 41-46. https://www.researchgate.net/publication/228845263.

Rodrigues, I., Parayil, A., Shetty, T., & Mirza, I. (2020). Use of linear discriminant analysis (LDA), k nearest neighbours (KNN), decision tree (CART), random forest (RF), gaussian naive bayes (NB), support vector machines (SVM) to predict admission for post-graduation courses [Paper Presentation]. International Conference on Recent Advances in Computational Techniques (IC-RACT).

Sarıcaoğlu, C. (2019). Sözcüksel analiz kullanarak kötü niyetli url'leri derin öğrenme teknikleri ile tespit etme (Publication No.655610) [Yayınlanmamış Yüksek Lisans Tezi]. Gazi Üniversitesi, Ankara.

Sembiring, S., Zarlis, M., Hartama, D., Ramliana, S., & Wani, E. (2011, April). Prediction of student academic performance by an application of data mining techniques International Conference on Management and Artificial Intelligence IPEDR (Vol. 6, pp. 110-114).

Shahiri, A. M., & Husain, W. (2015). A review on predicting student's performance using data mining techniques. *Procedia Computer Science*, *72*, 414-422. https://doi.org/10.1016/j.procs.2015.12.157

Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding machine learning: From theory to algorithms*. Cambridge University Press.

Sullivan, G. M., & Feinn, R. (2012). Using effect size—or why the P value is not enough. *Journal of graduate medical education*, *4*(3), 279-282. https://doi.org/10.4300/JGME-D-12-00156.1.

Şengür, D., & Tekin, A. (2013). Öğrencilerin mezuniyet notlarının veri madenciliği metotları ile tahmini [Prediction of students' graduation grades using data mining methods]. *Bilişim Teknolojileri Dergisi*, *6*(3), 7-16. https://dergipark.org.tr/tr/download/article-file/75330.

Tosunoğlu, E., Yılmaz, R., Özeren, E., Sağlam, Z. (2021). Eğitimde makine öğrenmesi: Araştırmalardaki güncel eğilimler üzerine inceleme [Machine learning in education: A review of current trends in research]. *Ahmet Keleşoğlu Eğitim Fakültesi Dergisi, 3*(2), 178-199. https://dergipark.org.tr/en/download/article-file/1873550

Türkmenoglu, C., & Tantug, A. C. (2014, June). Sentiment analysis in Turkish media [Paper Presentation]. International Conference on Machine Learning (ICML). Beijing, China.

Ülker, M., & Civalek, O. (2002). Yapay sinir ağları ile eksenel yüklü kolonların burkulma analizi [Buckling analysis of axially loaded columns using artificial neural networks]. *Turkish J. Eng. Env. Sci*, 26, 117-125.

Vandamme, J. P., Meskens, N., & Superby, J. F. (2007). Predicting academic performance by data mining methods. *Education Economics, 15*(4), 405. https://doi.org/10.1080/09645290701409939.

Vapnik, V. (2013). *The nature of statistical learning theory*. Springer science & business media.

Vijayalakshmi, V., & Venkatachalapathy, K. (2019). Comparison of predicting student's performance using machine learning algorithms. *International Journal of Intelligent Systems and Applications, 10*(12), 34. https://doi.org/10.5815/ijisa.2019.12.04.

Yağcı, M. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, *9*(1), 11. https://doi.org/10.1186/s40561-022-00192-z

Yang, S. J. H. (2019). Precision education: New challenges for AI in education [Paper Presentation]. 27th International Conference on Computers in Education (ICCE) (pp. XXVII-XXVIII). Kenting, Taiwan.

Yassein, N. A., Helali, R. G. M., & Mohomad, S. B. (2017). Predicting student academic performance in KSA using data mining techniques. *Journal of Information Technology & Software Engineering*, *7*(5), 1-5. DOI:10.4172/2165-7866.1000213

Zhou, Y., Huang, C., Hu, Q., Zhu, J., & Tang, Y. (2018). Personalized learning full-path recommendation model based on LSTM neural networks. *Information Sciences,* 444, 135-152. https://doi.org/10.1016/j.ins.2018.02.053.

Zilyas, D. ve Yılmaz, A. (2023). Makine öğrenmesi yöntemleri ile eğitim başarısının tahmini modeli [A model for predicting academic success using machine learning methods]. *DÜMF MD*, 14(3), 437–447. https://doi.org/10.24012/dumf.1322273.

Wang, A. Y., Wang, D., Drozdal, J., Liu, X., Park, S., Oney, S., & Brooks, C. (2021, May). What makes a well-documented notebook? a case study of data scientists' documentation practices in kaggle. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1-7). https://hdl.handle.net/1721.1/146030

Williamson, B. (2019). Digital policy sociology: Software and science in data-intensive precision education. *Critical Studies in Education, 62*(2), 1-17. https://doi.org/10.1080/17508487.2019.1691030.

Wu, D., Jennings, C., Terpenny, J., Gao, R. X., & Kumara, S. (2017). A Comparative study on machine learning algorithms for smart manufacturing: Tool wear prediction using random forests. *Journal of Manufacturing Science and Engineering, 139*(7), 071018. https://doi.org/10.1115/1.4036350.

Wu, J. Y., Hsiao, Y. C., & Nian, M. W. (2018). Using supervised machine learning on large-scale online forums to classify course-related Facebook messages in predicting learning achievement within the personal learning environment. *Interactive Learning Environments*, *28*(1), 65–80. https://doi.org/10.1080/10494820.2018.1515085.