



# Predicting the Academic Performance of Industrial Engineering Students Based on Socioeconomic Background and Past Achievements: A Two-step Blending-based Ensemble Approach

Alvin Muhammad 'Ainul Yaqin<sup>1</sup>, Priskila Destriani Banjarnahor<sup>1</sup>, Vridayani Anggi Leksono<sup>2\*</sup> Bayu Nur Abdallah<sup>3</sup> Mifthahul Janna Rosyid<sup>1</sup>

<sup>1</sup> Systems Modeling and Optimization Research Group, Department of Industrial Engineering, Kalimantan Institute of Technology, Balikpapan, Indonesia

<sup>2</sup> Department of Industrial Engineering, Kalimantan Institute of Technology, Balikpapan, Indonesia

<sup>3</sup> Department of Business Digital, Kalimantan Institute of Technology, Balikpapan, Indonesia  
\*anggi.leksono@lecturer.itk.ac.id

**Abstract.** The academic success of students is of utmost importance for higher education institutions. Therefore, accurately predicting students' academic performance is essential, and early interventions are necessary to improve their achievements. This research focuses on predicting the academic performance of new students and proposes a two-step ensemble learning-based approach for this purpose. A two-step ensemble learning-based approach can improve generalization and predictive ability. It provides a general overview of evaluation actions that both the university and the students can take. Additionally, it aims to contribute to the literature on predicting the academic performance of new students. The results of the conducted research demonstrate that the proposed model outperforms the elastic net model with an accuracy of 87.8% and a kappa value of 81.8%. The high result of the proposed model can be categorized as having good performance. The study also analyzes the variables influencing each model used in this case study. Notably, the top ten variables serving as significant features were identified: first semester GPA, number of failed classes, number of absences, achievement in academic competitions, mother's monthly income, high school GPA, quality of family relationship, scholarship status, hometown location, and travel time to the university.

**Keywords:** Academic Performance Prediction, University Students, Industrial Engineering, Machine Learning, Ensemble Learning.

## 1 Introduction

The academic success of students is vital for the success of a university [1]. The quality of graduates from a tertiary institution influences their employability in the job market [2]. Numerous factors impact students' academic performance, necessitating

proper management to enhance their achievements. Predicting academic performance offers numerous benefits, including guidance and support for students, improving graduation rates, evaluating learning methods, and informing academic policies [3].

However, predicting student academic performance is not easy due to the utilization of complex data, making it almost impossible to manually analyze and make decisions [4]. Due to the rapid advancement of technology, a large amount of information with complex data for decision-making can be carried out [5]. Machine learning has proven effective in analyzing complex data to predict student academic performance. It has been established as a valuable tool for predicting academic performance across various education levels [6].

Previous studies have employed machine learning in various educational contexts, such as predicting drop out of a university and early graduation. This previous study predicts student performance with 44 question attributes in the first semester with an ensemble model approach using heterogeneous multi-models [7]. This study has proposed an ensemble model approach using 2-layer stacking predicts student performance in academic competitions for 4 years with 11 assessment attributes [8]. This research proposes a novel two-step ensemble learning-based approach, using multiple algorithms, to predict student performance. The hybrid approach combines classification and ensemble models to enhance accuracy and efficiency in generating output, primarily through majority voting in the final stage for improved generalization and predictive ability. This is because majority voting yields better accuracy results. The prediction process involves two steps, the first step is to determine whether the GPA will improve or decline in the next semester and the second step is to predict the GPA based on the first step's results, and using majority voting to choose the best algorithm for accuracy.

The proposed approach was tested in a case study at an Indonesian state university, focusing on the industrial engineering study program. This research uses a two-step blending-based ensemble approach to predict academic. This research aims to predict academic achievements and provides insights into evaluation actions that universities and students can undertake to monitor and optimize academic progress. This innovative approach harnesses the collective intelligence of diverse predictive models, thereby enhancing the accuracy and reliability of performance predictions. Through this comprehensive overview, valuable insights can be gleaned to optimize educational practices and foster a conducive learning environment.

## **2 Research Methods**

### **2.1 Ridge Regression**

Ridge regression is a highly effective algorithm for handling multicollinear data. It accomplishes this by evenly shrinking the estimates of all variables. In contrast, least absolute shrinkage and selection operator (LASSO) reduces specific variable estimates to zero based on a particular parameter. Consequently, when the number of variables is greater than the number of observations, ridge algorithm outperforms LASSO [9].

## 2.2 Elastic Net

Elastic net combines the LASSO and ridge regression models into a single framework. It not only selects variables in the data but also exhibits better performance than LASSO. In this way, elastic net leverages the regularization capabilities provided by ridge regression. As the prediction of student performance cannot be immune to multicollinearity issues, the elastic net method was selected [10].

## 2.3 Support Vector Machine

Support vector machine (SVM) is an algorithm that uses in linear and nonlinear scenarios with the aim of achieving high performance in various contexts. This algorithm is often used because of its capability and accuracy in pattern recognition, classification, and regression. SVM has advantages compared to other models, because it is based on a strong theoretical foundation and high reliability. In addition, research in the field of education has not explored much of the potential use of this algorithm. Therefore, in this study, SVM is used as a decision support system for multiple classes, which aims to predict student performance [11].

## 2.4 Neural Network

Neural network is an algorithm with interconnected input, hidden, and output neurons forming a network. The connection weights between neurons are determined to categorize a data set. Neural network is renowned for its adaptability to data and autonomous learning capability. It has been extensively used to address challenging problems, demonstrating higher efficiency and accuracy compared to other classification methods [12].

## 2.5 Gradient Boosting

Gradient boosting (GB) is one of the most popular ensemble learning methods [13]. This algorithm combines multiple weak learners to create a robust ensemble, resulting in a powerful predictive model. GB helps reduce prediction errors and mitigate bias, making it highly efficient and effective. It has gained popularity due to its effectiveness in handling complex data sets, and researchers often utilize this algorithm to achieve success in various data competitions, including those on Kaggle [14].

# 3 Result and Analysis

## 3.1 Data

This research aims to predict academic outcomes, identify key factors influencing academic performance, and provide recommendations based on predictive models to enhance student achievement. The study population consisted of students from the industrial engineering study program at public university in Indonesia, who success-

fully completed their first year of study. Data collection took place from November 6 to December 16, 2021, and involved administering a questionnaire to students belonging to the 2017, 2018, 2019, and 2020 intake year. A total of 173 questionnaires were collected and edited using Microsoft Excel before analysis. Table 1 provides comprehensive details regarding the data set attributes utilized in the analysis.

During the analysis, it was observed that the variables father's and mother's income had missing values, which were imputed using the mean value to maintain data integrity. The data set was analyzed in R to ensure robust statistical processing. The data set underwent two significant stages of testing, namely regression, and classification, to explore its predictive capabilities. To ensure the validity and generalization of the results, the data set was divided into training data (70%) and testing data (30%). The list of variables in the data set is presented in Table 1.

**Table 1.** List of variables in the data set

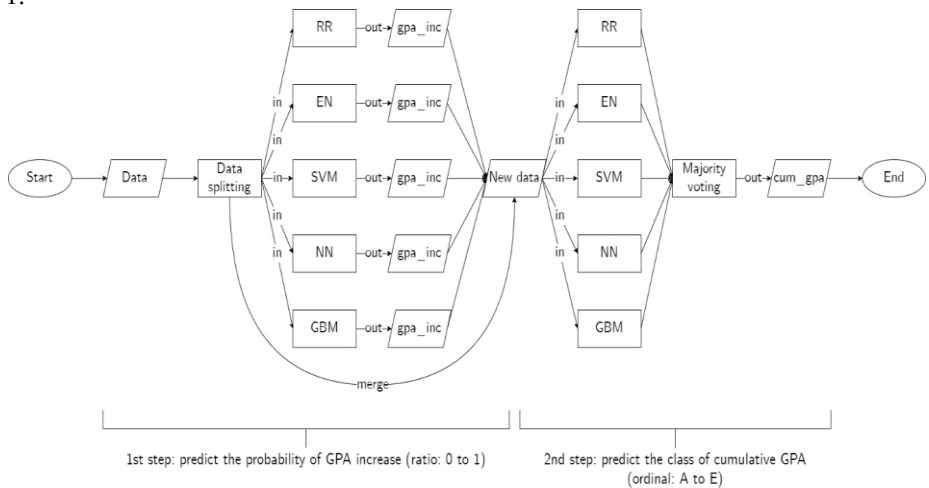
No.	Variable name	Variable definition	description	Variable type and possible values
1	sex	Sex		Nominal: male or female
2	ht_loc	Hometown location		Ordinal: 1 – native, 2 – from outside the city but on the same province, 3 – from outside the province but on the same island, or 4 – from outside the island
3	ht_size	Hometown size		Ordinal: 1 – small city, 2 – medium city, 3 – large city, 4 – metropolitan area, or 5 – megalopolis
4	f_edu	Father's education		Ordinal: 0 – none, 1 – elementary school, 2 – middle school, 3 – high school, 4 – diploma education, 5 – bachelor education, 6 – master's education, or 7 – doctoral education
5	m_edu	Mother's education		Ordinal: 0 – none, 1 – elementary school, 2 – middle school, 3 – high school, 4 – diploma education, 5 – bachelor education, 6 – master's education, or 7 – doctoral education
6	f_job	Father's job		Nominal: none, formal worker, or informal worker
7	m_job	Mother's job		Nominal: none, formal worker, or informal worker
8	f_income	Father's income	monthly	Ordinal: 0 – none, 1 – less than IDR 3 million, 2 – between IDR 3 to 6 million, 3 – between IDR 6 to 9 million, 4 – between IDR 9 to 12 million, 5 – between IDR 12 to 15 million, or 6 – more than IDR 15 million
9	m_income	Mother's income	monthly	Ordinal: 0 – none, 1 – less than IDR 3 million, 2 – between IDR 3 to 6 million, 3 – between IDR 6 to 9 million, 4 – between IDR 9 to 12 million, 5 – between IDR 12 to 15 million, or 6 – more than IDR 15 million
10	p_status	Parental status		Nominal: together or not together (for various reasons)
11	siblings	Number of siblings		Ordinal: 0 – none, 1 – one, 2 – two, 3 – three, or 4 – more than three
12	fam_rel	Perceived quality of family relationships		Interval: 1 – very bad to 5 – very good
13	hs_type	High school type		Nominal: public or private
14	tutoring	Attend tutoring during high school		Nominal: yes or no
15	aca_ach	Level of achievement in academic competitions during high school		Ordinal: 0 – none, 1 – lower than city level, 2 – city level, 3 – provincial level, 4 – national level, or 5 – international level
16	non_aca_ach	Level of achievement in non-academic competi-		Ordinal: 0 – none, 1 – lower than city level, 2 – city level, 3 – provincial level, 4 – national level, or 5 – international level

No.	Variable name	Variable description	Variable type and possible values
		tions during high school	
17	hs_gpa	High school GPA	Interval: 1 to 4
18	admission	University admission scheme	Nominal: SNMPTN (via invitation), SBMPTN (via national test), or SUMMIT (via local test)
19	travel_time	Travel time from residence to university	Ordinal: 1 – less than 10 minutes, 2 – between 10 to 20 minutes, 3 – between 20 to 30 minutes, 4 – between 30 to 40 minutes, 5 – between 40 to 50 minutes, 6 – between 50 to 60 minutes, or 7 – more than 60 minutes
20	allowance	Monthly allowance	Ordinal: 0 – none, 1 – less than IDR 1 million, 2 – between IDR 1 to 2 million, 3 – between IDR 2 to 3 million, 4 – between IDR 3 to 4 million, 5 – between IDR 4 to 5 million, 6 – between IDR 5 to 6 million, or 7 – more than IDR 6 million
21	scholarship	Currently on a scholarship	Nominal: yes or no
22	study_time	Average time spent studying per day	Ordinal: 0 – none, 1 – less than 1 hour, 2 – between 1 to 2 hours, 3 – between 2 to 3 hours, 4 – between 3 to 4 hours, 5 – between 4 to 5 hours, 6 – between 5 to 6 hours, or 7 – more than 6 hours
23	study_method	Preferred study method	Nominal: alone or in groups
24	class_diff	Perception of learning difficulties in the classroom	Interval: 1 – very easy to 5 – very hard
25	class_ben	Perception of learning benefits in the classroom	Interval: 1 – very beneficial to 5 – very useless
26	absences	The average number of absences per class	Ordinal: 0 – none, 1 – one time, 2 – one to two times, 3 – two to three times, 4 – three to four times, or 5 – more than four times
27	failed	Number of failed classes	Ordinal: 0 – none, 1 – one class, 2 – two classes, or 3 – more than three classes
28	out_time	Average time spent hanging out per day	Ordinal: 0 – none, 1 – less than 1 hour, 2 – between 1 to 2 hours, 3 – between 2 to 3 hours, 4 – between 3 to 4 hours, 5 – between 4 to 5 hours, 6 – between 5 to 6 hours, or 7 – more than 6 hours
29	sleep_time	Average time spent sleeping per day	Ordinal: 0 – none, 1 – less than 1 hour, 2 – between 1 to 2 hours, 3 – between 2 to 3 hours, 4 – between 3 to 4 hours, 5 – between 4 to 5 hours, 6 – between 5 to 6 hours, or 7 – more than 6 hours
30	activities	Actively participate in extracurricular activities	Nominal: yes or no
31	friends	Have close friends	Nominal: yes or no
32	romantic	Currently in a romantic relationship	Nominal: yes or no
33	exercise	Have a habit of exercising	Nominal: yes or no
34	phy_health	Perception of physical health condition	Interval: 1 – very unhealthy to 5 – very healthy
35	men_health	Perception of mental health condition	Interval: 1 – very unhealthy to 5 – very healthy
36	first_gpa	First semester GPA	Interval: 1 to 4

No.	Variable name	Variable description	description	Variable type and possible values
37	sec-ond_gpa	Second GPA	semester	Interval: 1 to 4
38	cum_gpa	Cumulative GPA		Ordinal: 1 – F to 7 – A

### 3.2 Proposed Model

In this study, a hybrid model is used based on a multi-output blending-based technique. This approach builds accurate and efficient machine learning for generating output. The combination of classification with ensemble models enhances generalization and prediction capabilities. In the final step, an ensemble learning-based model is used with majority voting as the ultimate decision based on the majority vote, as it provides better accuracy results [15]. The proposed models are presented in the Figure 1.



**Fig. 1.** The framework of the proposed model

The framework of the proposed model is structured as follows: in the first step, we collect information related to industrial engineering students who have passed their first year of study. There are 36 variables that will be tested with five different models to predict the probability of a student's semester grade either increasing or decreasing in the subsequent semester. In the second step, the test results from the first step will be processed again with five different models to predict cumulative GPA. Based on second step result, majority voting is carried out to select the best model in prediction accuracy.

### 3.3 Performance Metrics

#### Accuracy.

The effectiveness of the classification model can be evaluated through accuracy measurement meaning that models with high accuracy are more reliable, while those with poor accuracy have less impact on building trust [16]. Accuracy is calculated as the number of correct predictions divided by the total number of predictions sampled [17]

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive + False\ Negative} \quad (1)$$

In predictive modelling and classification tasks, understanding the terminology associated with different prediction outcomes is crucial. True positive indicates correct predictions where the predicted value corresponds to the actual occurrence of the event. Conversely, true negative represents correct predictions where the event did not happen as predicted. False positive refers to incorrect predictions where the event occurred despite the prediction suggesting otherwise. Finally, false negative occurs when the predicted value is wrong, and the event does not happen as predicted.

#### Cohen's kappa.

Cohen's kappa is another performance measurement tool for classification algorithm. Research about deep learning recommends using Cohen's kappa for dealing with class imbalance problems and multiclass data types in data sets [17]. It is defined as:

$$K = \frac{P_0 - P_e}{1 - P_e} = 1 - \frac{1 - P_0}{1 - P_e} \quad (2)$$

Where  $P_0$  represents the percentage of observed calculation, and  $P_e$  is the percentage of expected calculation. Cohen's kappa indicates the level of agreement:  $< 0$  indicates there is no agreement;  $0 - 0.20$  indicates low agreement,  $0.21 - 0.40$  indicates reasonable agreement,  $0.41 - 0.60$  indicates adequate agreement,  $0.61 - 0.80$  indicates fair agreement, and  $0.81 - 1$  indicates very strong agreement [17]

### 3.4 Discussion.

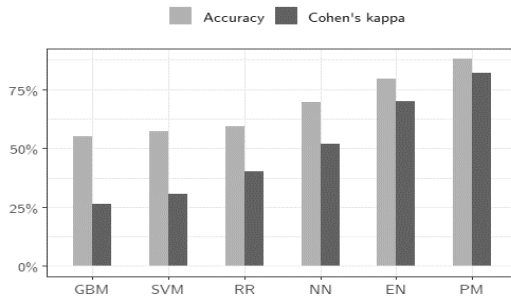
In this study, we processed a data set containing 36 variables. In the first step is to predict the probability of a student's semester grade either increasing or decreasing in the subsequent semester. The second step is to predict cumulative GPA. Based on the result from the second step majority voting is carried out to select the best model in prediction accuracy. Qualitative variables, such as sex, parental status, and admission status, were converted into quantitative data using dummy variables. Research about machine learning highlights the use of dummy variables in predictive models, as they have been proven to increase accuracy [18]. During the data set splitting process, 70% of the data was used for model training, while the remaining 30% was used for testing model performance [19].

In this case, many machine learning models are built with certain types of parameters. The SVM model is built by tuning the radial basis function kernel parameters. To measure the performance of each model, we utilized accuracy and kappa metrics. The

results of the model testing process and their respective performance are presented in the Table 2.

**Table 2.** Comparison of model performance

Model	Accuracy (%)		Cohen's kappa (%)	
	Training set	Testing set	Training set	Testing set
Ridge regression	50.80	59.20	28.30	40.10
Elastic net	73.10	79.60	61.30	69.80
SVM	53.70	57.10	25.90	30.70
Neural network	60.10	69.40	38.30	52.00
GB	55.80	55.10	28.20	26.30
Proposed model	<b>87.10</b>	<b>87.80</b>	<b>81.20</b>	<b>81.80</b>



**Fig. 2.** Comparison of model performance in the testing set

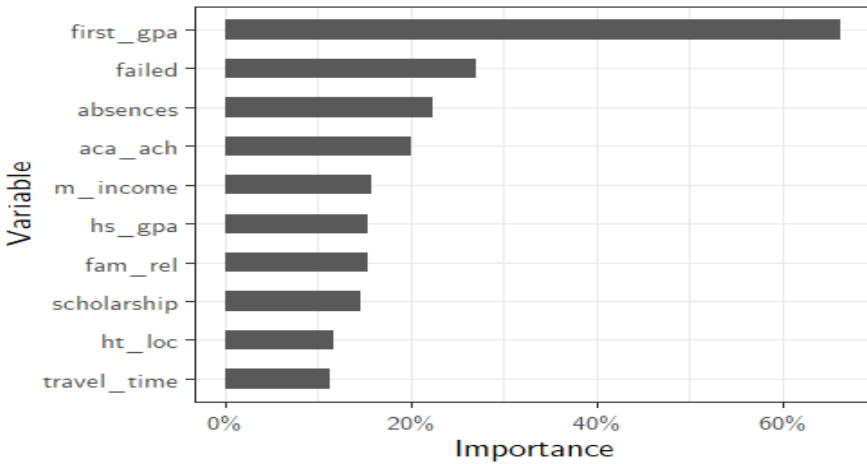
The results in the Figure 2 demonstrate that the elastic net model is better performance compared to other machine learning models with an accuracy of 79.6% and a kappa value of 69.8%. However, it is noteworthy that the proposed model outperformed the elastic net model with an accuracy of 87.8% and a kappa value of 81.8%. The high result of the proposed model can be categorized as having good performance.

The GB algorithm obtained the poorest performance with an accuracy of 55.1% and a kappa value of 26.3%. Therefore, this study recommends the use of the proposed model for predicting student academic performance. The superiority of the proposed model can be attributed to its two-step ensemble approach. In the first step, the proposed model combines the regression results from several algorithms used in this case study, including ridge regression, elastic net, SVM, neural network, and GB. In the second step, the results from the first step are further processed using the blending technique. The previous study recommends using a stacking model that combines several models for enhanced learning process and predictive performance. In their research, this approach achieved over 80% accuracy [20].



### 3.5 Feature Importance

The variables influencing the performance of student's academic achievement at each classification level can be analyzed from each constructed algorithm. In this case study, the variables influencing each algorithm model were analyzed, and the top ten significant features were identified. Figure 3 presents the significant variables that impact each model.



**Fig. 3.** The ten most important variables based on the weighted average value of their standardized coefficient values in the proposed model

**Table 3.** The ten most important variables based on the weighted average value of their standardized coefficient values in the proposed model

No.	Variable name	Variable description	Importance (%)
1	first_gpa	First semester GPA	66.30
2	failed	Number of failed classes	27.00
3	absences	The average number of absences per class	22.37
4	aca_ach	Level of achievement in academic competitions during high school	20.01
5	m_income	Mother's monthly income	15.79
6	hs_gpa	High school GPA	15.45
7	fam_rel	Perceived quality of family relationships	15.41
8	scholarship	Currently on a scholarship	14.63
9	ht_loc	Hometown location	11.65
10	travel_time	Travel time from residence to university	11.38

Table 3 displays the top ten variables influencing student academic achievement. These variables include first semester GPA, number of failed classes, number of absences, achievement in academic competitions, mother's monthly income, high school GPA, quality of family relationship, scholarship status, hometown location, and travel

time to the university. These factors are crucial in predicting the academic performance of first year students in the industrial engineering program. In the previous research, the variables that most influence student academic achievement are grade point average scores, class absences, and participation in academic competitions [8]

The cumulative GPA, number of absences, and achievement in academic competitions significantly influence student academic achievement [21]. The results of this study have practical implications for students, helping them identify weaknesses and improve their learning methods. Additionally, this research can assist lecturers and study programs in evaluating learning methods and academic policies to enhance student outcomes.

## 4 Conclusion

This research aimed to predict the academic performance of new students in the industrial engineering program at a university in Indonesia using a two-step blending. The results of this study demonstrate that the proposed model outperforms the elastic net, achieving an accuracy of 87.8% and a kappa value of 81.8%. The high result of proposed model can be categorized as having good performance. The analysis of variables influencing each algorithm model revealed the top ten significant features. The first semester GPA, number of failed classes, number of absences, achievement in academic competitions, mother's monthly income, high school GPA, quality of family relationship, scholarship status, hometown location, and travel time to the university were identified as the most important factors influencing the prediction model for academic performance. This research can help students in recognize academic weaknesses such as cumulative GPA, number of failed classes, and number of absences so as to improve their study methods. It also assists lecturers and study programs in evaluating study methods and academic policies. For further research, it is recommended to collect data from different study programs to achieve more comprehensive results and gain a better understanding of first-year students' academic performance. Additionally, future studies may consider adding or updating more relevant input and output variables to enhance prediction accuracy.

## 5 Authors' Contributions

**Alvin Muhammad 'Ainul Yaqin:** Conceptualization, Methodology, Validation, Investigation, Resources, Writing – original draft preparation, Writing – review and editing, Supervision, Project administration, Funding acquisition. **Priskila Destriani Banjarnahor:** Methodology, Software, Formal analysis, Investigation, Data curation, Writing – original draft preparation, Visualization. **Vridayani Anggi Leksono:** Validation, Writing – review and editing, Supervision, Project administration, Funding acquisition. **Bayu Nur Abdallah:** Validation, Project administration, Funding acquisition. **Mifthahul Janna Rosyid:** Writing – original draft preparation. **Alvin Muhammad 'Ainul Yaqin** and **Priskila Destriani Banjarnahor** are equal contributors to this work and designated as co-first authors.

## 6 Acknowledgments

This work was supported by the Institute for Research and Community Service of the Kalimantan Institute of Technology under grant number 2674/IT10/PPM.05/2022.

## References

1. E. Alyahyan and D. Düstegör, “Predicting academic success in higher education: literature review and best practices,” *Int. J. Educ. Technol. High*, vol. 17, no. 1, Dec. 2020, DOI: 10.1186/s41239-020-0177-7.
2. H. Alhakami, T. Alsubait, and A. Aljarallah, “Data mining for student advising,” *Int. J. Adv. Comput. Sci. Appl*, vol. 11, no. 3, pp. 526–532, 2020, DOI: 10.14569/ijacsa.2020.0110367.
3. F. Ofori, E. Maina, and R. Gitonga, “Using machine learning algorithms to predict students’ performance and improve learning outcome: A literature-based review,” *J. Inf. Technol*, vol. 4, no. 1, pp. 33–55, 2020, DOI: <https://stratfordjournals.org/journals/index.php/Journal-of-Information-and-Techn/article/view/480>.
4. E. B. Belachew and F. A. Gobena, “Student performance prediction model using machine learning approach: The case of Wolkite University,” *Int. J. Adv. Res. Comput. Sci. Softw. Eng*, vol. 7, no. 2, pp. 46–50, Feb. 2017, DOI: 10.23956/IJARCSSE/V7I2/01219.
5. S. Umadevi and K. S. J. Marseline, “A survey on data mining classification algorithms,” *Proceedings of IEEE International Conference on Signal Processing and Communication, 2017*, DOI: 10.1109/CSPC.2017.8305851.
6. P. Balaji, S. Alelyani, A. Qahmash, and M. Mohana, “Contributions of machine learning models towards student academic performance prediction: A systematic review,” *Appl. Sci*, vol. 11, no. 21, Nov. 2021, DOI: 10.3390/app112110007.
7. O. W. Adejo and T. Connolly, “Predicting student academic performance using multi-model heterogeneous ensemble approach,” *J. Appl. Res. High. Educ.*, vol. 10, no. 1, pp. 61–75, 2018, doi: 10.1108/JARHE-09-2017-0113.
8. L. Yan and Y. Liu, “An ensemble prediction model for potential student recommendation using machine learning,” *Sym*, vol. 12, no. 5, May 2020, doi: 10.3390/SYM12050728.
9. Chan, J.Y. Le, S.M.H Leow, K.T Bea, W. K Cheng, S. W. Phoong, Z. W. Hong, and Y. L. Chen, “Mitigating the multicollinearity problem and its machine learning approach: A review,” *J. Math*, vol. 10, no. 8, Apr. 2022, DOI: 10.3390/math10081283.
10. J. E. Yoo, “TIMSS 2011 student and teacher predictors for mathematics achievement explored and identified via elastic net,” *Front Psychol*, vol. 9, no. 317, Mar. 2018, DOI: 10.3389/fpsyg.2018.00317.
11. M. G. Asogbon, O. W. Samuel, M. O. Omisore, and B. A. Ojokoh, “A multi-class support vector machine approach for students academic performance prediction,” *J. Inter. Multi. Res. Liter*, vol. 4, DOI: <http://ijmcr.com>.
12. N. M. Suhaimi, S. Abdul-Rahman, S. Mutalib, N. H. A. Hamid, and A. M. A. Malik, “Review on predicting students’ graduation time using machine learning algorithms,” *Int. J. Mod. Educ. Comput. Sci*, vol. 11, no. 7, pp. 1–13, Jul. 2019, DOI: 10.5815/ijmecs.2019.07.01.
13. J. H. Friedman, “A gradient boosting machine,” *Ann. Statist*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001, DOI: 10.1214/aos/1013203451.

14. O. H. T. Lu, A. Y. Q. Huang, J. C. H. Huang, A. J. Q. Lin, H. Ogata, and S. J. H. Yang, "Applying learning analytics for the early prediction of students' academic performance in blended learning," *Educ. Technol. Soc.*, vol. 21, no. 2, pp. 220–232, Apr. 2018, DOI: <https://www.jstor.org/stable/26388400>.
15. A. Siddique, A. Jan, F. Majeed, A. I. Qahmash, N. N. Quadri, and M. O. A. Wahab, "Predicting academic performance using an efficient model based on fusion of classifiers," *Appl. Sci.*, vol. 11, no. 24, Dec. 2021, DOI: 10.3390/app112411845.
16. Y. K. Salal, S. M. Abdullaev, and M. Kumar, "Educational data mining: Student performance prediction in academic," *Int J Eng Adv Technol*, vol. 8, no. 4C, pp. 54–59, 2019.
17. S. Hussain, Z. F. Muhsin, Y. K. Salal, P. Theodorou, F. Kurtoglu, and G. C. Hazarika, "Prediction model on student performance based on internal assessment using deep learning," *Int. J. Emerg. Technol. Learn*, vol. 14, no. 8, pp. 4–22, 2019, DOI: 10.3991/ijet.v14i08.10001.
18. S. Jolly and N. Gupta, "Understanding and implementing machine learning models with dummy variables with low variance," *Adv. Intell. Syst. Comput*, 2021, pp. 477–487. DOI: 10.1007/978-981-15-5113-0\_37.
19. M. Kuhn and K. Johnson, *Applied Predictive Modeling*. Springer, 2013. DOI: 10.1007/978-1-4614-6849-3.
20. F. Saleem, Z. Ullah, B. Fakieh, and F. Kateb, "Intelligent decision support system for predicting student's e-learning performance using ensemble machine learning," *J. Math.*, vol. 9, no. 17, 2021, doi: 10.3390/math9172078.
21. E. Fernandes, M. Holanda, M. Victorino, V. Borges, R. Carvalho, and G. Van Erven, "Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil," *J. Bus. Res.*, vol. 94, pp. 335–343, Jan. 2019, DOI: 10.1016/j.jbusres.2018.02.012.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

