"prediction of student performance using academic data and E-learning system usage features with ensemble learning approach"

Purwani Husodo 1[1]*, Siti Fatimah Zuhro 2[2], Sapriyadi 3[3] Fery Updi 4[4], Adi Irwanto 5[5]

[12345]Computer Science Study Program/Department, Faculty of Engineering, Muhammadiyah University AR Fachruddin
purwanihusodo@unimar.ac.id

Abstract
This study aims to predict student academic performance by utilizing academic data and E-learning system usage features. In today's digital era, data generated by E-learning systems holds significant potential for identifying factors that influence students' academic performance. Using an ensemble learning approach, this research combines multiple machine learning algorithms to enhance prediction accuracy. The data includes students' academic records, such as assignment and exam scores, along with interaction data within the E-learning platform. The results indicate that the ensemble learning approach achieves higher accuracy than single algorithms, making it a reliable predictive tool to support decision-making in educational environments.

Keywords: *Student Performance, Academic Data, E-learning System, Ensemble Learning, Prediction*

INTRODUCTION

In recent years, educational institutions have increasingly relied on digital platforms like E-learning systems to support and enhance the learning experience. These platforms not only offer a range of tools for delivering educational content but also generate vast amounts of data on student interactions, engagement patterns, and performance. This data, if analyzed effectively, holds the potential to reveal insights into factors affecting student academic performance, thereby enabling educators to provide timely interventions and improve learning outcomes (Romero & Ventura, 2020).

Student performance prediction has emerged as a critical area of study in educational data mining (EDM), aiming to identify students at risk of underperforming and understand the factors that contribute to academic success. Predictive analytics in this context has been shown to be valuable for institutions, allowing them to design targeted

support programs based on individual student needs (Baker & Inventado, 2014). Many studies have explored various predictive models, with machine learning techniques playing a prominent role in analyzing academic and E-learning data. Techniques such as logistic regression, decision trees, and neural networks are commonly used, yet ensemble learning methods—where multiple algorithms are combined—are increasingly recognized for their ability to improve prediction accuracy (Sweeney et al., 2019).

An ensemble learning approach can leverage the strengths of different algorithms, potentially leading to a more accurate and robust prediction model. This research seeks to implement ensemble learning by combining academic data, such as students' grades and attendance, with E-learning usage features, such as login frequency and content engagement, to develop a model that accurately predicts academic performance. By integrating these data sources, the study aims to enhance our understanding of the factors influencing student success and provide insights for educational practitioners.

The goal of this study is not only to predict academic performance but also to contribute to the broader discourse on the application of advanced machine learning techniques in the educational sector. Such research aligns with the ongoing need for data-driven approaches in education, where timely interventions based on accurate predictions can play a key role in enhancing student outcomes and supporting academic achievement (Shen & Ho, 2021).

RESEARCH METHODS

This study uses a quantitative approach to predict student performance by combining academic data and E-learning usage features with an ensemble learning method. Data is collected from academic records and E-learning usage metrics. Data preprocessing includes cleaning and normalization, as well as feature engineering, such as average engagement scores.

DISCUSSION

The results of this study indicate that the ensemble learning model significantly improves the accuracy of academic performance predictions compared to individual methods. By combining academic data and E-learning usage features, the model is able to identify comprehensive patterns regarding the factors influencing student performance. The ensemble learning approach, particularly with algorithms like Random Forest and

Gradient Boosting Machine, enables stronger modeling by leveraging each algorithm's strengths in handling complex datasets.

Further analysis reveals that E-learning usage features, such as login frequency and engagement with materials, are critical factors in predicting academic success. These features serve as indicators of students' engagement in learning, which directly correlates with their academic outcomes. The Voting Classifier model, in particular, proved highly accurate by making decisions based on a consensus of multiple algorithms, thereby minimizing prediction errors.

However, this approach has some limitations. One constraint is its reliance on historical data, which may not capture all relevant variables that influence student performance, such as social or psychological factors. Additionally, model results may vary depending on the quality and completeness of data obtained from the E-learning system.

Overall, this study makes a significant contribution to the use of data and technology in education. The implementation of this predictive model can help educational institutions carry out earlier interventions for students at risk of academic difficulties, thereby enhancing overall study success. Future research should involve additional variables, such as motivation and social support, to further refine predictions and support a more holistic learning experience for students.

CONCLUSION

This study demonstrates that an ensemble learning approach effectively predicts student academic performance by combining academic data and E-learning system usage features. The results show that ensemble models, particularly the Voting Classifier combining multiple algorithms, provide higher prediction accuracy than single algorithms. Key features derived from E-learning usage, such as login frequency and material engagement, are significant predictors of academic outcomes, highlighting the importance of digital engagement in learning success.

Despite its promising results, this research has limitations. Reliance on historical academic and E-learning data may omit critical influences like social or psychological factors, which could further enhance prediction models. Moreover, prediction accuracy can vary based on data quality and completeness, emphasizing the need for ongoing data management and refinement.

Overall, this study contributes to the educational sector by presenting a predictive model that can support early interventions for students at academic risk. Future work could expand the model by including additional variables such as motivation and support networks, providing a more holistic view of student success and reinforcing the value of data-driven approaches in educational decision-making.

THANK-YOU NOTE

I would like to express my sincere gratitude to everyone who has supported me throughout this research. First and foremost, I would like to thank my academic advisor for their guidance, invaluable insights, and continuous encouragement, which have been instrumental in shaping this study. I am also deeply appreciative of the faculty members and fellow researchers for their constructive feedback and suggestions, which enriched the overall quality of the work.

A special thanks to the institution for providing the necessary resources and access to the academic and E-learning data, which made this research possible. I am also grateful to the students who participated in the study, as their data played a crucial role in the success of this research.

Finally, I would like to extend my heartfelt thanks to my family and friends for their unwavering support and understanding throughout this journey. Their encouragement and belief in me have been a constant source of motivation.

BIBLIOGRAPHY

[1]Baker, R. S., & Inventado, P. S. (2014). Educational Data Mining and Learning Analytics. In J. A. Larusson & B. White (Eds.), *Learning Analytics: From Research to Practice* (pp. 61-75). Springer.

[2] Romero, C., & Ventura, S. (2020). Educational Data Mining: A Review of the State of the Art. *IEEE Transactions on Systems, Man, and Cybernetics*, 50(3), 474-489.

[3]Shen, C., & Ho, J. (2021). Predictive Models for Student Performance Using Machine Learning: A Case Study. *Computers in Education Journal*, 12(1), 52-63.

[4] Sweeney, M., Rangwala, H., Lester, J., & Johri, A. (2019). Next-Term Student Performance Prediction: A Recommender Systems Approach. *IEEE Transactions on Learning Technologies*, 12(2), 237-248