International Journal of Modern Education IJMOE
EISSN: 2637-0905

Volume 7 Issue 24 (March 2025) PP. 887-902
DOI: 10.35631/IJMOE.724064

INTERNATIONAL JOURNAL OF MODERN EDUCATION (IJMOE)
www.ijmoe.com

GLOBAL ACADEMIC EXCELLENCE

ENSEMBLE LEARNING IN EDUCATIONAL DATA ANALYSIS FOR IMPROVED PREDICTION OF STUDENT PERFORMANCE: A LITERATURE REVIEW

Noor Fadzilah Ab Rahman[1], Shir Li Wang[2*], Nurkaliza Khalid[3]

[1] Department of Computing, Universiti Islam Selangor, Malaysia
Email: noorfadzilah@uis.edu.my
[2] Department of Software Engineering and Smart Technology, Universiti Pendidikan Sultan Idris, Malaysia
Email: shirli_wang@meta.upsi.edu.my
[3] Department of Computing, Universiti Islam Selangor, Malaysia
Email: nurkaliza@uis.edu.my
[*] Corresponding Author

Article Info:

Article history:
Received date: 26.01.2025
Revised date: 12.02.2025
Accepted date: 10.03.2025
Published date: 19.03.2025

To cite this document:

Ab Rahman, N. F., Shir, L. W., & Khalid, N. (2025). Ensemble Learning In Educational Data Analysis For Improved Prediction Of Student Performance: A Literature Review. *International Journal of Modern Education*, *7* (24), 887-902.

DOI: 10.35631/IJMOE.724064

Abstract:

The integration of advanced technology and digital platforms in modern education is essential for enhancing educational outcomes. Ensemble learning has emerged as a prominent approach in educational data analysis, demonstrating its effectiveness in improving student performance predictions. The study reviews the application of ensemble learning methods in educational data analysis to improve student performance prediction. The primary objective of this review is to highlight the effectiveness of ensemble approaches in achieving superior prediction accuracy compared to individual classifiers. Additionally, the review examines into the influence of feature selection techniques on optimizing ensemble models by identifying crucial attributes and mitigating the complexity of educational data. The findings show that ensemble learning offers a robust framework for tackling addressing challenges in educational data mining, such as managing high-dimensional datasets and imbalanced classes. By incorporating feature selection methods, ensemble models become more efficient and scalable for various educational datasets. This review concludes that ensemble learning with integrated feature selection is a transformative tool for enhancing student performance prediction. Ensemble learning presents promising opportunities for driving innovation in educational data analysis and addressing the evolving challenges in education. The study offers valuable insights and benefits as a resource for educators, researchers and practitioners, shedding light on the transformative potential of ensemble learning in educational decision-making technology.

International Journal of Modern Education
IJMOE
EISSN: 2637-0905
Volume 7 Issue 24 (March 2025) PP. 887-902
DOI: 10.35631/IJMOE.724064

Keywords:

Machine Learning, Ensemble Learning, Feature Selection, Student Performance, Educational Data Analysis

Introduction

The integration of advanced technology and digital platforms into teaching and learning has become increasingly crucial in modern education to enhance educational outcomes. The rapid growth of educational data sources, encompassing Learning Management Systems (LMS), student information systems, and digital learning tools, has created new opportunities for data-driven decision-making (Kaspi & Venkatraman, 2023; Schildkamp, 2019). However, the expansion volume and complexity of educational data pose challenges in extracting valuable insights (Munshi & Alhindi, 2021). In response, advanced analytical techniques, particularly Artificial Intelligence (AI) and Machine Learning (ML), have emerged as essential tools for processing, analyzing, and interpreting educational data. By leveraging these techniques, institutions can make well-informed, data-driven decisions that significantly enhance the quality of education.

AI and ML have revolutionized education, transforming student learning, instructional strategies, and institutional decision-making. AI-powered applications, such as simulation-based learning, learning analytics, and predictive modeling, offer personalized learning experiences, optimize adaptive learning and student engagement, and enhance institutional efficiency (Papadakis, Kiv, Kravtsov, Osadchyi, Marienko, Pinchuk, Shyshkina, Sokolyuk, Mintii, Vakaliuk, Striuk, et al., 2023). The integration of AI in education has facilitated enhanced digital education and supported open learning models by enabling real-time performance monitoring and personalized feedback.

Numerous studies further emphasize the growing adoption of AI-driven technologies in higher education. For instance, research conducted in Greece explores the factors influencing students' intentions and actual use of AI applications in the humanities and social sciences (Lavidas et al., 2024). The study reveals that key determinants such as performance expectancy, social influence, habit formation, and facilitating conditions significantly impact students' acceptance and utilization of AI applications for academic purposes. These findings suggest a substantial reliance on the adoption of AI applications in education significantly influences students' learning experiences, engagement, and academic performance.

Beyond individual learning applications, advancements in cloud-based smart learning environments and digital leadership frameworks further boost AI's role in education. AI-powered, cloud-based smart learning and augmented reality (AR) technologies are revolutionizing instructional methodologies in modernizing education by providing adaptive and immersive learning experiences that meet to the diverse needs of students (Papadakis, Kiv, Kravtsov, Osadchyi, Marienko, Pinchuk, Shyshkina, Sokolyuk, Mintii, Vakaliuk, Azarova, et al., 2023). Furthermore, the increasing prominence of digital leadership and technology-enhanced education has facilitated the widespread adoption of advanced AI in educational settings. Digital transformation, encompassing virtual leadership and AI-driven decision-making, plays a pivotal role in optimizing digitalization in educational environments (Karakose et al., 2022).

International Journal of Modern Education IJMOE EISSN: 2637-0905
Volume 7 Issue 24 (March 2025) PP. 887-902
DOI: 10.35631/IJMOE.724064
As AI continues to transform the educational landscape, the demand for more accurate and interpretable predictive models is growing to improve the accuracy and reliability of student performance predictions. Predicting student performance is among of the most critical applications of educational data analysis. Data-driven decision-making has gains prominence, leading educational institutions to embrace predictive analytics using ML. This approach enables institutions to identify at-risk students, customize learning experiences, and ultimately enhance academic outcomes (Mangat & Saini, 2020). In advance, ML offers substantial opportunities for large-scale educational data analysis (Alam & Mohanty, 2023). By processing vast amounts of educational data, ML algorithms can identify patterns and insights that are crucial for making informed decisions and policies (Munir et al., 2022). This capability is essential for enhancing student performance, optimizing resource allocation, and achieving improved overall educational outcomes (Wu & Zheng, 2021).

In the current digital learning landscape, educational datasets often encompass a wide range of high-dimensional features, including demographic information, behavioral interactions, assessments, and engagement metrics (Hamad, 2023; Putri et al., 2021). However, not all these features are equally valuable in terms of predictive accuracy. Some may introduce noise or redundancy, which can hinder the effectiveness of predictive models, leading to overfitting, increased computational costs, and reduced model interpretability.

Numerous studies highlight the significance of feature selection methods in overcoming the challenges posed by imbalanced educational datasets and high-dimensional educational datasets. Feature selection is the process of identifying the most relevant features for a predictive model. These methods aim to select crucial features and eliminate irrelevant ones to improve the performance of ML algorithms (Hamad, 2023; Kushagra et al., 2023). Another study by Arif et al., (2021) utilized feature selection methods and ML classifiers to identify the significant feature selectors that enhance student performance prediction accuracy.

While feature selection methods enhance the efficiency of conventional ML models, single-learning models still face challenges in effectively handling feature selection within high-dimensional educational datasets. Single learning classifiers, such as decision trees, logistic regression, and support vector machines, may struggle to identify the most relevant features due to their limited ability to capture complex feature interactions (Hilbert et al., 2021). Furthermore, single-learning model approaches may not fully recognize hidden patterns and dependencies in the educational data, resulting in suboptimal model performance and limited generalizability in real-world educational settings (Pawar et al., 2023). To address these challenges, ensemble learning (EL), a technique that combines multiple ML models, has emerged as a more robust and accurate approach for student performance prediction.

Studies indicate that combining the predictions of multiple ML models, known as ensemble methods, can overcome the limitations of individual classifiers. Ensemble techniques possess the capability to outperform individual models in terms of performance (Vanneschi & Silva, 2023), and reduce overfitting and enhance model stability (Dong et al., 2020). Furthermore, the relevance of EL in data science has experienced substantial growth, leading to significant improvements in predictive capabilities and the reliability of predictive models (Shah et al., 2023).

International Journal of Modern Education IJMOE EISSN: 2637-0905
Volume 7 Issue 24 (March 2025) PP. 887-902
DOI: 10.35631/IJMOE.724064
The integration of various ML algorithms enhances the analysis of educational data, resulting in increased prediction accuracy and reliability, making this ensemble method particularly advantageous in educational contexts (Zeineddine et al., 2021). As educational institutions increasingly rely on AI-driven analytics, the integration of feature selection with EL presents a powerful approach to enhance student success through data-driven decision-making.

Literature Review

Ensemble techniques have demonstrated remarkable effectiveness across various domains, including healthcare and finance, and have recently been increasingly employed in education. These methods are particularly important in educational data analysis due to the ability in addressing challenges such as data sparsity, noise, and class imbalance, ultimately enhancing the reliability of predictions.

Ensemble methods are widely employed in student performance prediction research to improve model accuracy, robustness, and generalizability. These methods combine multiple learning models to improve predictions and mitigate the limitations of single-learning approaches. The ensemble methods can be broadly categorized into three main types: bagging, boosting, and stacking.

Bagging, also known as Bootstrap Aggregating, is a technique used to train multiple instances of a model on different subsets of data. By aggregating the predictions from these instances, it helps reduce the variance in the final predictions. For instance, Manoharan et al., (2023) employed ensemble approaches, such as Random Forest and boosting, that effectively address the issue of sparse datasets by integrating predictions from several models and help in minimizing the impact of limited data availability. Similarly, Sahlaoui et al., (2021) combined ensemble models with the Synthetic Minority Oversampling Technique (SMOTE) to handle class imbalance in student performance prediction, resulting in more reliable assessments. These studies highlight the potential of ensemble techniques in refining predictive accuracy in educational settings.

Boosting techniques play a crucial role in predicting student performance by training models sequentially, allowing each subsequent model to correct the errors of its predecessor. This iterative process reduces bias and enhances predictive accuracy by focusing on residual errors from previous models, ultimately improving overall performance. Gradient boosting algorithms, such as XGBoost, CaBoost, and LightGBM, have demonstrated considerable effectiveness in educational data analysis. For instance, CaBoost has achieved a prediction accuracy of 92.16% in determining final student status (Mashagba et al., 2023). By continuously refining predictions, these algorithms facilitate more precise identification of at-risk students, enabling timely interventions that support academic success. The application of boosting techniques in education is particularly valuable, as it contributes to data-driven decision-making and personalized learning strategies.

Sequential modeling techniques, such as Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU), have been widely utilized to analyze student performance over time. These models are particularly effective in handling sequential data, achieving prediction accuracies exceeding 90% in various educational contexts (Kurniawati & Maulidevi, 2022). However, mitigating biases in student performance predictions remains a critical challenge.

International Journal of Modern Education
IJMOE
EISSN: 2637-0905
Volume 7 Issue 24 (March 2025) PP. 887-902
DOI: 10.35631/IJMOE.724064

To address this issue, attention-based personalized federated learning approaches have been proposed to optimize predictions across diverse demographic groups (Chu et al., 2022). By incorporating self-supervised learning and attention mechanisms, these methods enhance both model accuracy and fairness, ensuring that underrepresented student populations are adequately accounted for in predictive analyses.

Stacking, also known as stacked generalization, is an effective EL technique that merges multiple base learners to enhance predictive performance through the creation of a meta-model. This approach has demonstrated remarkable potential in predicting student performance by capitalizing on the strengths of various ML algorithms. Common base learner algorithms utilized in this technique include Random Forest, KNN, and XGBoost, which are combined to form a robust ensemble model (Abiodun & Wreford, 2024). Logistic regression is commonly used as the meta-learner, although innovative methods such as computational geometry have been introduced to enhance interpretability and reduce the need for hyperparameter tuning (Tong & Li, 2025).

Stacked models have demonstrated superior predictive accuracy, as evidenced by performance metrics such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R^2 scores, which indicate strong correlations between predicted and actual student performance (Abiodun & Wreford, 2024). Stacking consistently outperforms individual ML models by achieving lower error rates and higher precision in predictions (Muhammad Ricky Perdana Putra & Ema Utami, 2024). However, while stacking has proven highly effective, blending techniques may yield slightly better results in certain applications, such as predicting student dropout rates in massive open online courses (MOOCs). This suggests that the selection of EL method should be carefully considered based on the specific educational context and predictive objective (Muhammad Ricky Perdana Putra & Ema Utami, 2024).

However, despite their strengths, ensemble models alone do not always yield optimal results, particularly when dealing with high-dimensional educational data containing redundant or irrelevant features. Feature selection techniques play a crucial role in further enhancing the efficiency and accuracy of ensemble models by identifying the most relevant predictors of student performance. Feature selection methods are essential for optimizing ML models by reducing dimensionality and improving performance. The integration of feature selection techniques with ensemble ML models has gained increasing attention in educational contexts due to its potential to enhance predictive accuracy while reducing computational complexity (Rahul & Katarya, 2023). Feature selection techniques help eliminate irrelevant or redundant attributes, allowing ensemble models to focus on the most significant predictors of student success.

Several studies have explored the effectiveness of combining feature selection with EL. For example, Evangelista and Sy (2022), applied feature selection with ensemble models to predict student academic performance, demonstrating improved accuracy compared to single models. Other study by Sengupta (2023) used feature selection and EL to identify key factors affecting student success in a specific course, achieving high prediction accuracy. Alija et al., (2023) employed data mining techniques, including feature selection and ensemble models, to not only predict student academic performance but also identify influential features that impact learning outcomes.

International Journal of Modern Education
IJMOE
EISSN: 2637-0905
Volume 7 Issue 24 (March 2025) PP. 887-902
DOI: 10.35631/IJMOE.724064
Babu et al., (2023) demonstrated how combining feature selection with EL improves student success prediction in online courses, highlighting its applicability across varied educational settings. While ensemble methods and feature selection techniques have been widely recognized for their effectiveness in student performance prediction, existing studies lack comprehensive comparative analyses. Many studies focus on a single ensemble approach without evaluating its performance against other ensemble techniques or baseline ML models in educational settings.

This study aims to systematically review the existing literature on EL for student performance prediction, with a particular emphasis on comparing the effectiveness of different ensemble methods, including bagging, boosting, and stacking, in educational data analysis. Additionally, the study examines the impact of integrating feature selection techniques on ensemble models, specifically in terms of improving accuracy, interpretability, and computational efficiency. By synthesizing these findings, the study seeks to provide valuable insights for researchers, educators, and practitioners, highlighting the transformative potential of EL in educational technology. By addressing these gaps, this study seeks to provide a comprehensive understanding of the role of EL in education, offering a resource for future research and practical implementation in academic settings.

Methodology

The primary objective of this study is to investigate the effectiveness of EL methods in enhancing the accuracy of student performance predictions. The study employed a literature review methodology to explore and identify the role and practices of EL techniques in educational data analysis for predicting student performance. The methodology of this study divided into two major stages: data collection and data analysis, which each aimed to ensure the rigor and relevance of the review.

In the initial stage of the study, data collection was undertaken. The study identified and retrieved relevant literature from the Scopus database. This database provide access to top and recent publications related to disciplines such as machine learning and education. The search strategy employed a combination of keywords, including “ensemble learning,” “educational data analysis,” and “student performance prediction,” to the article title, abstract, and keywords within the Scopus database, as presented in Table 1. These search keywords were employed to identify studies that focus on the application of EL techniques in educational contexts. During this stage, the present study successfully accumulated a comprehensive total of 28 relevant research papers.

Table 1: The Search Strategy

Database	Search string keyword
Scopus	TITLE-ABS-KEY ("ensemble learning" OR "educational data analysis" AND "student performance prediction")

The study further examined the most relevant studies based on the inclusion and exclusion criteria defined in Table 2. The first criterion was the timeline, which was limited to the scope of the search procedure only included publications published between 2020 and 2024. Consequently, the authors decided to focus only on journal sources as the primary source of literature (research publications). Other document types were not included, such as

International Journal of Modern Education
EISSN: 2637-0905
IJMOE
Volume 7 Issue 24 (March 2025) PP. 887-902
DOI: 10.35631/IJMOE.724064
conference paper, book chapter, and conference review. Furthermore, this review's analysis was restricted to English-language literature and final publication stage.

In this stage, the collected studies underwent a rigorous screening process. Titles and abstracts were reviewed to identify relevant studies, and the full texts of shortlisted articles were evaluated to confirm their suitability for inclusion. The final compilation of literature included 15 studies that demonstrated the application of EL techniques to enhance predictive accuracy and model performance in educational data analysis.

Table 2: Inclusion and Exclusion Criteria

Criteria	Inclusion	Exclusion
Timeline	Between 2020-2024	< 2020
Source Type	Journal (only research article)	Conference proceeding, Book series, Book
Document Type	Article	Conference paper, Book Chapter, Conference Review
Language	English	Non-English
Publication stage	Final	Article in press

The second stage involved analyzing and synthesizing information from the collected studies. Initially, each study was systematically evaluated and categorized based on key research dimensions, such as EL techniques, feature selection strategies, and model performance evaluation metrics. The effectiveness of ensemble techniques used to analyze educational data was also given special attention. Model performance evaluation metrics, including accuracy, precision, recall, and F1-score, were reviewed to assess the impact of ensemble methods on predictive outcomes.

Subsequently, the study further synthesized the findings to identify patterns and trends, evaluate the effectiveness of various ensemble approaches, assess the impact of feature selection techniques, and highlight their practical implications for educational settings. The present study categorizes the findings into three main themes: ensemble techniques, feature selection methods, and ML algorithms, as well as performance ensemble models.

Results and Discussion

The findings of this review highlight the transformative potential of EL techniques in educational data analysis, particularly in predicting student performance. The results provide valuable insights into the effectiveness and adaptability of ensemble methods and feature selection techniques in the education domain. Consequently, the present study categorized the findings into two major themes: ensemble learning in model performance and the impact of feature selection on model performance.

Ensemble Learning in Model Performance

Ensemble learning has emerged as a robust methodology for analyzing educational data and predicting student performance. Ensemble techniques employed in predictive models consistently demonstrated superior accuracy compared to single-model approaches due to the synergistic combination of multiple ML algorithms. Numerous studies have demonstrated the

International Journal of Modern Education
IJMOE
EISSN: 2637-0905
Volume 7 Issue 24 (March 2025) PP. 887-902
DOI: 10.35631/IJMOE.724064

efficacy of integrating ensemble methods and a diverse range of ML algorithms within educational contexts, as presented in Table 3.

Table 3: Summary of Ensemble Learning in Model Performance

Ensemble method	ML algorithm	Result	Implication for Education Institutions	Reference
stacking	Decision tree, k-nearest neighbor, Naïve Bayes, and One vs. Rest support vector machine	Prediction accuracy of 93%.	Improves student success rates through early interventions and personalized learning.	(Saluja et al., 2023)
stacking	Naïve Bayes,Random Forest, Decision tree, Adaboost, SVM, K-NN, Logistic Regression	Prediction accuracy of 98.4%.	Identifies at-risk students proactively to improve overall success.	(Pek et al., 2023)
stacking	RF, LR, SVM, NB, and K-Neighbors Classifier	Improved model accuracy.	Enhances educational evaluation and student monitoring.	(Xuan Lam et al., 2024)
bagging, boosting stacking, voting	Naïve Bayes, J48, MLP, Random Forest, Logistic Regression	Boosting ensemble improved model accuracy.	Enhancing assessments, identifying at-risk students early, and offering targeted support to improve student success.	(Butt et al., 2023)
stacking	SVM, RF, AdaBoost	High precision, recall, and low error rate in prediction.	Enhancing the identification and support of talented students' performance in academic competitions.	(Yan & Liu, 2020)
boosting, stacking	Gradient Boosting, Extreme Gradient Boosting, Light Gradient Boosting Machine	Ensemble learning (boosting & stacking) improves predictive results and efficiency.	Supports proactive educational strategies and optimized resource distribution.	(Keser & Aghalarova, 2022)
bagging, boosting,	DT, RF, GBT, NB, and KNN	Stacking five classifiers achieved the	Improves student monitoring and	(Saleem et al., 2021)

International Journal of Modern Education
IJMOE
EISSN: 2637-0905
Volume 7 Issue 24 (March 2025) PP. 887-902
DOI: 10.35631/IJMOE.724064

stacking, voting		highest F1 score (0.8195).	support in e-learning environments	
bagging, boosting	RF, FDT, BN, SVM, Naïve Bayesian, Linear Regression.	Ensemble algorithm (bagging and boosting) achieved high accuracy (98.25% for binary and 89.47% for multi-class) in classification tasks.	Helps institutions personalize learning experiences and improve success rates.	(Latif et al., 2023)

As presented in Table 3, the study's findings highlight the effectiveness of combining ensemble methods with ML in predicting student performance and its implications for educational institutions.

Stacking emerged as the most frequently used ensemble technique, demonstrating high predictive accuracy across various studies (Pek et al., 2023; Saluja et al., 2023; Xuan Lam et al., 2024; Yan & Liu, 2020). Pek et al. (2023) achieved the highest prediction accuracy of 98.4% using a stacking model with multiple classifiers, showcasing its ability to effectively identify at-risk students and support early interventions. Other studies that used stacking also reported improved accuracy, which contributed to enhanced educational evaluations and personalized learning strategies (Xuan Lam et al., 2024). Furthermore, Yan and Liu (2020) discovered that stacking was beneficial in identifying and supporting talented students, which ultimately enhanced their performance in academic competitions.

Beyond stacking, the combination of bagging, boosting, and voting also significantly improved model performance. Keser and Aghalarova (2022) demonstrated that integrating boosting with stacking enhanced predictive efficiency, enabling educational institutions to adopt more proactive strategies and optimize resource allocation. Similarly, Saleem et al., (2021) reported that employing bagging, boosting, stacking, and voting resulted in the highest F1 score (0.8195), making it particularly valuable for student monitoring and support in e-learning environments. Latif et al., (2023) achieved an impressive accuracy of 98.25% in binary classification using bagging and boosting, emphasizing the potential of these methods to personalize learning experiences and enhance student success rates.

The discussion of these findings emphasizes the growing significance of EL methods in education. Stacking models have demonstrated remarkable effectiveness in identifying at-risk students at an early stage, enabling timely interventions that can significantly enhance academic outcomes. Furthermore, these models play a crucial role in talent identification, allowing institutions to provide targeted support to high-achieving students. The combination of boosting and bagging further enhances predictive accuracy, making data-driven educational strategies more efficient and reliable. Institutions that implement these ensemble techniques

International Journal of Modern Education
IJMOE
EISSN: 2637-0905
Volume 7 Issue 24 (March 2025) PP. 887-902
DOI: 10.35631/IJMOE.724064

can enhance student monitoring, optimize resource allocation, and design more personalized learning environments.

Overall, the findings indicate that EL methods, particularly stacking, bagging, and boosting, are valuable tools for improving student performance prediction. Their successful implementation in educational settings can lead to the development of more effective early intervention strategies, providing support for both struggling and high-achieving students. Ultimately, this can result in enhanced academic success. Future research should explore the integration of these models into real-world educational systems and assess their long-term impact on student learning outcomes.

Impact of Feature Selection on Ensemble Model Performance

Feature selection is an essential step in developing predictive models. This technique identifies the most relevant attributes that contribute to prediction accuracy and efficiency. When combined with ensemble models, feature selection enhances prediction robustness and effectiveness by reducing data complexity and concentrating on key variables. Ensemble models, which combine the predictions of multiple ML classifiers, greatly benefit from well-selected features, resulting in enhanced performance and reliability.

Table 4 presents the impact of various feature selection techniques on ensemble model performance in predicting student academic outcomes. Each technique demonstrates significantly improves prediction accuracy and has practical implications for educational institutions.

Table 4: Impact of Feature Selection on Ensemble Model Performance

Feature selection Technique	Result	Implication for Education Institutions	Reference
Random forest	Achieved prediction accuracy with 82.4%	Enables accurate academic performance prediction, helping educators identify struggling students early and tailor data-driven interventions effectively.	(Xiaoming et al., 2022)
Filter-based and wrapper-based feature selection	Achieved higher predictive accuracy	Enables educational institutions make informed decisions to enhance student performance and engagement.	(Evangelista, E. 2021)
Genetic Algorithm (GA)	Achieved higher classification accuracy	Insights from predictions assist institutions refine teaching strategies and improve student outcomes.	(Farissi et al., 2020)
Principal Component Analysis (PCA)	Achieved high accuracy	Enable develop effective academic warning systems, driving better decision-making in education management.	(Li & Li, 2024)

International Journal of Modern Education
IJMOE
EISSN: 2637-0905
Volume 7 Issue 24 (March 2025) PP. 887-902
DOI: 10.35631/IJMOE.724064

Binary Particle Swarm Optimization (BPSO)	Achieved accuracy of 96.6%	Provide educators actionable insights and enhancing educational management by accurately identifying factors influencing student performance.	(Begum & Padmannavar, 2022)
SelectKBest	Achieved accuracy of 83.16%	Enables educational institutions enhanced decision-making to support at-risk students early, improving their outcomes.	(Yacoub et al., 2022)
Backward selection with Pearson correlation	Achieved high accuracy of 93%	Enables early identification of weak students, facilitating timely interventions and providing actionable insights for students, parents, and institutions to improve academic outcomes.	(Fida et al., 2022)

As shown in Table 4, various feature selection techniques have been integrated into ensemble methods to enhance model performance. Effective feature selection plays a crucial role in improving the accuracy and efficiency of ensemble models. The number of studies that incorporated feature selection demonstrated superior model prediction accuracy, which has practical implications for educational institutions.

Xiaoming et al., (2022) reported the effectiveness of feature selection techniques, particularly Random Forest, in enhancing the accuracy of predicting academic performance to an impressive 82.4%. The proposed feature selection method in the ensemble model facilitates early identification of students who are struggling and provides data-driven interventions. Similar to a study conducted by Fida et al., (2022), which employed Backward Selection along with Pearson Correlation, achieving an impressive 93% accuracy rate, the study successfully identified weak students and provided timely interventions to enhance their academic performance.

Other feature selection techniques are demonstrated to enhance predictive accuracy across various scenarios. For instance, the binary particle swarm optimization (BPSO) method, which achieved an impressive accuracy of 96.6% (Begum & Padmannavar, 2022) and SelectKBest, which yielded an accuracy of 83.16% (Yacoub et al., 2022). These techniques significantly contribute to educational management. By providing valuable insights into performance-influencing factors, enabling early identification of at-risk students and facilitate targeted interventions that can improve their outcomes.

In essence, the reviewed studies highlight the significance of feature selection techniques in enhancing the predictive accuracy of ensemble models for predicting student performance. These methods offer several beneficial implications for education, including enhanced prediction capabilities that enable educational institutions to make precise predictions,

Volume 7 Issue 24 (March 2025) PP. 887-902
DOI: 10.35631/IJMOE.724064

optimize resource allocation, and develop effective intervention plans. Additionally, the application of feature selection optimizes educational data analysis, allowing institutions to emphasize on the most influential factors impacting student outcomes.

Conclusion

The study aims to demonstrate the potential of EL techniques in educational data analysis for enhancing the accuracy of student performance predictions. The study explores the effectiveness of ensemble methods, which combine the strengths of multiple ML models, in achieving higher accuracy and robustness in student performance predictions. Additionally, the study investigates the impact of feature selection techniques on improving model performance by identifying and focusing on relevant attributes, thereby reducing complexity in educational data analysis.

The findings show that ensemble methods consistently outperform single models, with some achieving high prediction accuracies that match the ensemble model's performance. These methods allow educational institutions to identify at-risk students early, provide personalized learning support, and improve decision-making processes. Meanwhile, feature selection techniques further improve the efficiency and scalability of predictive models, making these techniques appropriate for analyzing diverse educational datasets.

To advance the field, particularly in modern education, this paper proposes integrating ensemble methods with emerging technologies like deep learning and explainable AI. These advancements can enhance predictive accuracy, transparency, and scalability, providing educators with actionable insights and enabling more effective interventions to improve student outcomes. Ensemble learning, therefore, stands as a transformative tool for addressing challenges in educational data mining and fostering better academic achievements.

Acknowledegment

The authors would like to express their sincere gratitude to Global Academic Excellence (M) Sdn Bhd for the generous granting the Publication Grant Scheme and extend their deepest appreciation to the research community for their invaluable contributions to the success of this study.

References

Abiodun, O. J., & Wreford, A. I. (2024). Student's Performance Evaluation Using Ensemble Machine Learning Algorithms. *Engineering and Technology Journal*, *09*(08). https://doi.org/10.47191/etj/v9i08.23

Alam, A., & Mohanty, A. (2023). Predicting Students' Performance Employing Educational Data Mining Techniques, Machine Learning, and Learning Analytics. *Communications in Computer and Information Science*. https://doi.org/10.1007/978-3-031-43140-1_15

Alija, S., Beqiri, E., Gaafar, A. S., & Hamoud, A. K. (2023). Predicting Students Performance Using Supervised Machine Learning Based on Imbalanced Dataset and Wrapper Feature Selection. *Informatica*, *47*(1). https://doi.org/10.31449/inf.v47i1.4519

Arif, M. A., Jahan, A., Mau, M. I., & Tummarzia, R. (2021). An Improved Prediction System of Students' Performance Using Classification model and Feature Selection Algorithm. *International Journal of Advances in Soft Computing and Its Applications*, *13*(1), 162–177.

International Journal of Modern Education IJMOE
EISSN: 2637-0905
Volume 7 Issue 24 (March 2025) PP. 887-902
DOI: 10.35631/IJMOE.724064
Babu, I., MathuSoothana, R., & Kumar, S. (2023). Evolutionary Algorithm Based Feature Subset Selection for Students Academic Performance Analysis. *Intelligent Automation & Soft Computing*, *36*(3), 3621–3636. https://doi.org/10.32604/iasc.2023.033791

Begum, S., & Padmannavar, S. S. (2022). Student performance prediction with BPSO feature selection and CNN classifier. *International Journal of Advanced and Applied Sciences*, *9*(11), 84–92. https://doi.org/10.21833/ijaas.2022.11.010

Butt, N. A., Mahmood, Z., Shakeel, K., Alfarhood, S., Safran, M., & Ashraf, I. (2023). Performance Prediction of Students in Higher Education Using Multi-Model Ensemble Approach. *IEEE Access*, *11*, 136091–136108. https://doi.org/10.1109/ACCESS.2023.3336987

Chu, Y.-W., Hosseinalipour, S., Tenorio, E., Cruz, L., Douglas, K., Lan, A., & Brinton, C. (2022). Mitigating Biases in Student Performance Prediction via Attention-Based Personalized Federated Learning. *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, *14*(8), 3033–3042. https://doi.org/10.1145/3511808.3557108

Dong, X., Yu, Z., Cao, W., Shi, Y., & Ma, Q. (2020). A survey on ensemble learning. *Frontiers of Computer Science*, *14*(2), 241–258. https://doi.org/10.1007/s11704-019-8208-z

Evangelista, E. (2021). A Hybrid Machine Learning Framework for Predicting Students' Performance in Virtual Learning Environment. *International Journal of Emerging Technologies in Learning (IJET)*, *16*(24), 255–272. https://doi.org/10.3991/ijet.v16i24.26151

Evangelista, E., & Sy, B. (2022). An approach for improved students' performance prediction using homogeneous and heterogeneous ensemble methods. *International Journal of Electrical and Computer Engineering (IJECE)*, *12*(5), 5226. https://doi.org/10.11591/ijece.v12i5.pp5226-5235

Farissi, A., Mohamed Dahlan, H., & Samsuryadi. (2020). Genetic Algorithm Based Feature Selection With Ensemble Methods For Student Academic Performance Prediction. *Journal of Physics: Conference Series*, *1500*(1), 012110. https://doi.org/10.1088/1742-6596/1500/1/012110

Fida, S., Masood, N., Tariq, N., & Qayyum, F. (2022). A Novel Hybrid Ensemble Clustering Technique for Student Performance Prediction. *JUCS - Journal of Universal Computer Science*, *28*(8), 777–798. https://doi.org/10.3897/jucs.73427

Hamad, Z. O. (2023). Review of feature selection methods using optimization algorithm. *Polytechnic Journal*, *12*(2), 203–214. https://doi.org/10.25156/ptj.v12n2y2022.pp203-214

Hilbert, S., Coors, S., Kraus, E., Bischl, B., Lindl, A., Frei, M., Wild, J., Krauss, S., Goretzko, D., & Stachl, C. (2021). Machine learning for the educational sciences. *Review of Education*, *9*(3). https://doi.org/10.1002/rev3.3310

Karakose, T., Kocabas, I., Yirci, R., Papadakis, S., Ozdemir, T. Y., & Demirkol, M. (2022). The Development and Evolution of Digital Leadership: A Bibliometric Mapping Approach-Based Study. *Sustainability*, *14*(23), 16171. https://doi.org/10.3390/su142316171

Kaspi, S., & Venkatraman, S. (2023). Data-Driven Decision-Making (DDDM) for Higher Education Assessments: A Case Study. *Systems*, *11*(6), 306. https://doi.org/10.3390/systems11060306

Keser, S. B., & Aghalarova, S. (2022). HELA: A novel hybrid ensemble learning algorithm for predicting academic performance of students. *Education and Information Technologies*. https://doi.org/10.1007/s10639-021-10780-0

Volume 7 Issue 24 (March 2025) PP. 887-902
DOI: 10.35631/IJMOE.724064
Kurniawati, G., & Maulidevi, N. U. (2022). Multivariate Sequential Modelling for Student Performance and Graduation Prediction. *2022 9th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)*, 293–298. https://doi.org/10.1109/ICITACEE55701.2022.9923971

Kushagra, Kumar, R., & Jain, S. (2023). Feature Selection for Medical Diagnosis Using Machine Learning: A Review. In *Lecture Notes in Electrical Engineering* (pp. 179–189). https://doi.org/10.1007/978-981-19-8493-8_14

Latif, G., Abdelhamid, S. E., Fawagreh, K. S., Brahim, G. Ben, & Alghazo, R. (2023). Machine Learning in Higher Education: Students' Performance Assessment Considering Online Activity Logs. *IEEE Access*, *11*, 69586–69600. https://doi.org/10.1109/ACCESS.2023.3287972

Lavidas, K., Voulgari, I., Papadakis, S., Athanassopoulos, S., Anastasiou, A., Filippidi, A., Komis, V., & Karacapilidis, N. (2024). Determinants of Humanities and Social Sciences Students' Intentions to Use Artificial Intelligence Applications for Academic Purposes. *Information*, *15*(6), 314. https://doi.org/10.3390/info15060314

Li, X., & Li, C. (2024). Research on Construction of Student Academic Early Warning Model Based on Ensemble Learning. *Communications in Computer and Information Science*. https://doi.org/10.1007/978-981-99-9499-1_18

Mangat, P. K., & Saini, K. S. (2020). Predictive Analytics for Students Performance Prediction. *International Journal of Recent Technology and Engineering (IJRTE)*, *9*(3), 300–305. https://doi.org/10.35940/ijrte.C4417.099320

Manoharan, R., Stalin, M. S., Loganathan, G. B., S, D., & K, V. (2023). *Ensemble Model for Educational Data Mining Based on Synthetic Minority Oversampling Technique* (pp. 1–14). https://doi.org/10.21203/rs.3.rs-2946501/v1

Mashagba, E., Al-Saqqar, F., & Al-Shatnawi, A. (2023). Using Gradient Boosting Algorithms in Predicting Student Academic Performance. *2023 International Conference on Business Analytics for Technology and Security (ICBATS)*, 1–7. https://doi.org/10.1109/ICBATS57792.2023.10111325

Muhammad Ricky Perdana Putra, & Ema Utami. (2024). Comparative Analysis of Hybrid Model Performance Using Stacking and Blending Techniques for Student Drop Out Prediction In MOOC. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *8*(3), 346–354. https://doi.org/10.29207/resti.v8i3.5760

Munir, H., Vogel, B., & Jacobsson, A. (2022). Artificial Intelligence and Machine Learning Approaches in Digital Education: A Systematic Revision. *Information*, *13*(4), 203. https://doi.org/10.3390/info13040203

Munshi, A. A., & Alhindi, A. (2021). Big Data Platform for Educational Analytics. *IEEE Access*. https://doi.org/10.1109/ACCESS.2021.3070737

Papadakis, S., Kiv, A. E., Kravtsov, H. M., Osadchyi, V. V., Marienko, M. V., Pinchuk, O. P., Shyshkina, M. P., Sokolyuk, O. M., Mintii, I. S., Vakaliuk, T. A., Azarova, L. E., Kolgatina, L. S., Amelina, S. M., Volkova, N. P., Velychko, V. Y., Striuk, A. M., & Semerikov, S. O. (2023). Unlocking the power of synergy: the joint force of cloud technologies and augmented reality in education. *CEUR Workshop Proceedings*. https://doi.org/10.31812/123456789/7399

Papadakis, S., Kiv, A. E., Kravtsov, H. M., Osadchyi, V. V., Marienko, M. V., Pinchuk, O. P., Shyshkina, M. P., Sokolyuk, O. M., Mintii, I. S., Vakaliuk, T. A., Striuk, A. M., & Semerikov, S. O. (2023). Revolutionizing education: using computer simulation and cloud-based smart technology to facilitate successful open learning. *CEUR Workshop Proceedings*. https://doi.org/10.31812/123456789/7375

International Journal of Modern Education
IJMOE
EISSN: 2637-0905
Volume 7 Issue 24 (March 2025) PP. 887-902
DOI: 10.35631/IJMOE.724064

Pawar, A., Manjula Shenoy, K., Prabhu, S., & Guruprasad Rai, D. (2023). Performance analysis of machine learning algorithms: Single Model VS Ensemble Model. *Journal of Physics: Conference Series*. https://doi.org/10.1088/1742-6596/2571/1/012007

Pek, R. Z., Ozyer, S. T., Elhage, T., Ozyer, T., & Alhajj, R. (2023). The Role of Machine Learning in Identifying Students At-Risk and Minimizing Failure. *IEEE Access*, *11*(January), 1224–1243.https://doi.org/10.1109/ACCESS.2022.3232984

Putri, N. L., Nugroho, R. A., & Herteno, R. (2021). Intrusion Detection System Berbasis Seleksi Fitur Dengan Kombinasi Filter Information Gain Ratio Dan Correlation. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, *8*(3), 457. https://doi.org/10.25126/jtiik.0813154

Rahul, & Katarya, R. (2023). Deep auto encoder based on a transient search capsule network for student performance prediction. *Multimedia Tools and Applications*, *82*(15), 23427–23451. https://doi.org/10.1007/s11042-022-14083-5

Sahlaoui, H., Alaoui, E. A. A., Nayyar, A., Agoujil, S., & Jaber, M. M. (2021). Predicting and Interpreting Student Performance Using Ensemble Models and Shapley Additive Explanations. *IEEE Access*, *9*, 152688–152703. https://doi.org/10.1109/ACCESS.2021.3124270

Saleem, F., Ullah, Z., Fakieh, B., & Kateb, F. (2021). Intelligent Decision Support System for Predicting Student's E-Learning Performance Using Ensemble Machine Learning. *Mathematics*, *9*(17), 2078. https://doi.org/10.3390/math9172078

Saluja, R., Rai, M., & Saluja, R. (2023). Designing new student performance prediction model using ensemble machine learning. *Journal of Autonomous Intelligence*, *6*(1), 583. https://doi.org/10.32629/jai.v6i1.583

Schildkamp, K. (2019). Data-based decision-making for school improvement: Research insights and gaps. *Educational Research*, *61*(3), 257–273. https://doi.org/10.1080/00131881.2019.1625716

Sengupta, S. (2023). Towards Finding a Minimal Set of Features for Predicting Students' Performance Using Educational Data Mining. *International Journal of Modern Education and Computer Science*, *15*(3), 44–54. https://doi.org/10.5815/ijmecs.2023.03.04

Shah, M., Gandhi, K., Patel, K. A., Kantawala, H., Patel, R., & Kothari, A. (2023). Theoretical Evaluation of Ensemble Machine Learning Techniques. *2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 829–837.https://doi.org/10.1109/ICSSIT55814.2023.10061139

Tong, T., & Li, Z. (2025). Predicting learning achievement using ensemble learning with result explanation. *PLOS ONE*, *20*(1), e0312124.https://doi.org/10.1371/journal.pone.0312124

Vanneschi, L., & Silva, S. (2023). Ensemble Methods. In *Natural Computing Series* (pp. 283–288). https://doi.org/10.1007/978-3-031-17922-8_11

Wu, B., & Zheng, C. (2021). An Analysis of the Effectiveness of Machine Learning Theory in the Evaluation of Education and Teaching. *Wireless Communications and Mobile Computing*, *2021*, 1–10. https://doi.org/10.1155/2021/4456222

Xiaoming, D., Ying, C., Xiaofang, Z., & Yu, G. (2022). Study on Feature Engineering and Ensemble Learning for Student Academic Performance Prediction. *International Journal of Advanced Computer Science and Applications*, *13*(5), 495–502. https://doi.org/10.14569/IJACSA.2022.0130558

Xuan Lam, P., Mai, P. Q. H., Nguyen, Q. H., Pham, T., Nguyen, T. H. H., & Nguyen, T. H. (2024). Enhancing educational evaluation through predictive student assessment

International Journal of Modern Education
IJMOE
EISSN: 2637-0905
Volume 7 Issue 24 (March 2025) PP. 887-902
DOI: 10.35631/IJMOE.724064
modeling. *Computers and Education: Artificial Intelligence*, *6*(October 2023), 100244. https://doi.org/10.1016/j.caeai.2024.100244

Yacoub, M. F., Maghawry, H. A., Helal, N. A., Gharib, T. F., & Ventura, S. (2022). An Enhanced Predictive Approach for Students' Performance. *International Journal of Advanced Computer Science and Applications*. https://doi.org/10.14569/IJACSA.2022.01304101

Yan, L., & Liu, Y. (2020). An Ensemble Prediction Model for Potential Student Recommendation Using Machine Learning. *Symmetry*, *12*(5), 728. https://doi.org/10.3390/sym12050728

Zeineddine, H., Braendle, U., & Farah, A. (2021). Enhancing prediction of student success: Automated machine learning approach. *Computers & Electrical Engineering*, *89*, 106903. https://doi.org/10.1016/j.compeleceng.2020.106903