Comparative Analysis of Machine Learning and Deep Learning Models for Predicting Student Selection Outcomes Using Examination Data

Felinda Aprilia Rahma[1,*], Siti Zayyana Ulfah[2]

[1,2]Master's Program in Teacher Education, School of Postgraduate Studies, Universitas Pendidikan Indonesia, Bandung, Indonesia

ABSTRACT

This study presents a comparative analysis of machine learning and deep learning models for predicting student selection outcomes using examination data. The models evaluated include Logistic Regression, Random Forest, Extra Trees, Gradient Boosting, and a Multilayer Perceptron. The dataset consists of structured features such as examination scores and demographic attributes, with preprocessing steps including handling missing values, feature engineering, encoding categorical variables, and addressing class imbalance through class merging. Model performance was evaluated using accuracy, precision, recall, and F1-score, along with 5-fold cross-validation to assess robustness. The results show that Gradient Boosting achieved the best overall performance, followed closely by Random Forest and the Multilayer Perceptron, while Logistic Regression performed significantly worse. Feature importance analysis indicates that academic performance variables, particularly the total score and its components, are the most influential predictors, whereas demographic features have minimal impact. These findings suggest that ensemble-based methods are highly effective for structured tabular data and that deep learning does not necessarily outperform traditional machine learning models in this context. This study highlights the importance of selecting appropriate models based on data characteristics and provides insights for developing AI-based student selection systems.

Keywords Student Selection Prediction, Gradient Boosting, Multilayer Perceptron, Tabular Data Classification, Educational Data Mining

Introduction

Student selection is a critical process in educational institutions, particularly in competitive admission systems where decisions must be made based on large volumes of applicant data. Traditionally, selection outcomes are determined using rule-based approaches or simple statistical methods that rely heavily on examination scores. While these approaches provide a straightforward evaluation mechanism, they often fail to fully utilize the richness of available data and may not capture complex relationships between multiple factors influencing student performance. With the increasing availability of educational data, there is a growing need for more advanced and data-driven approaches to improve the accuracy, efficiency, and fairness of the selection process [1].

In recent years, machine learning techniques have been widely applied in the field of educational data mining to predict student outcomes, including academic performance, dropout risk, and admission success [2]. Models such as Logistic Regression, Decision Trees, and Random Forest have demonstrated strong performance in handling structured tabular data due to their ability to model non-

Submitted 1 June 2026
Accepted 15 August 2026
Published 17 November 2026

*Corresponding author
Felinda Aprilia Rahma,
felindaaprilia16@upi.edu

Additional Information and Declarations can be found on page 249

DOI: 10.63913/ail.v2i4.72
Copyright
2026 Rahma and Ulfah

Distributed under
Creative Commons CC-BY 4.0

How to cite this article: F. A. Rahma and S. Z. Ulfah, "Comparative Analysis of Machine Learning and Deep Learning Models for Predicting Student Selection Outcomes Using Examination Data," *Artif. Intell. Learn.*, vol. 2, no. 4, pp. 237-251, 2026.

linear relationships and interactions between features. Ensemble methods, in particular, have been recognized as state-of-the-art approaches for classification tasks, as they combine multiple models to improve predictive accuracy and robustness [3]. At the same time, deep learning models, especially neural networks such as the Multilayer Perceptron, have gained popularity due to their success in complex domains such as image recognition and natural language processing [4].

Despite these advancements, there is still ongoing debate regarding the effectiveness of deep learning compared to traditional machine learning methods when applied to structured tabular data. Several studies suggest that tree-based ensemble methods often outperform deep learning models in such contexts, especially when the dataset size is moderate and features are well-defined. However, other studies highlight the potential of deep learning models when properly tuned and applied to sufficiently large datasets. This inconsistency in findings indicates that the superiority of one approach over another is not yet conclusive and depends heavily on the characteristics of the data and the problem domain [5].

Based on this context, a significant research gap can be identified. First, there is a lack of comprehensive comparative studies that evaluate both machine learning and deep learning models specifically in the context of student selection prediction using examination-based data. Second, many existing studies focus on either performance comparison or feature analysis, but rarely integrate both aspects within a single framework. Third, limited attention has been given to understanding the relative importance of academic and demographic features in influencing selection outcomes, which is essential for ensuring fairness and transparency in decision-making systems.

To address these gaps, this study conducts a comparative analysis of multiple machine learning models, including Logistic Regression, Random Forest, Extra Trees, and Gradient Boosting, alongside a deep learning model based on the Multilayer Perceptron. The models are evaluated using standard performance metrics and cross-validation techniques to ensure robustness. In addition, feature importance analysis is performed to identify the most influential factors in predicting student selection outcomes. The main contributions of this study are threefold. First, it provides a systematic comparison between machine learning and deep learning models on structured educational data. Second, it offers empirical evidence on the effectiveness of ensemble methods for this specific prediction task. Third, it provides insights into the role of academic and demographic features, which can support the development of more transparent and data-driven student selection systems.

Literature Review and Related Works

The application of machine learning techniques in educational data mining has gained significant attention in recent years, particularly for predicting student performance, dropout risk, and admission outcomes. Traditional approaches such as Logistic Regression and Decision Trees have been widely used due to their simplicity and interpretability [6], [7]. However, these models often struggle to capture complex relationships in high-dimensional data, leading to the adoption of more advanced techniques.

Ensemble learning methods, including Random Forest, Gradient Boosting, and Extra Trees, have emerged as state-of-the-art approaches for classification

tasks involving structured tabular data [8], [9]. These models improve predictive performance by combining multiple learners and reducing overfitting. Several studies have demonstrated that ensemble methods outperform single models in educational prediction tasks, particularly when dealing with heterogeneous features and non-linear relationships [10], [11].

In parallel, deep learning models have been increasingly explored in the education domain due to their success in other fields such as computer vision and natural language processing [12]. Multilayer Perceptron models, in particular, have been applied to predict student outcomes by learning complex feature representations [13], [14]. Despite their flexibility, deep learning models often require large datasets and careful tuning to achieve optimal performance, which can limit their effectiveness in smaller or structured datasets [15].

Comparative studies between machine learning and deep learning models have produced mixed results. Some research indicates that deep learning models can outperform traditional methods when sufficient data and computational resources are available [16]. However, other studies report that tree-based ensemble models consistently achieve better performance on tabular datasets due to their ability to handle feature interactions and missing values more effectively [17], [18]. This suggests that the choice of model should be guided by the nature of the data rather than the complexity of the algorithm.

Feature importance and interpretability have also been key topics in educational data mining. Several studies highlight that academic performance indicators, such as examination scores, are the most significant predictors of student outcomes [19]. In contrast, demographic features often have limited influence, although they may still provide contextual information [20]. Understanding feature contributions is essential for building transparent and fair decision-making systems in education.

Overall, existing research demonstrates the effectiveness of both machine learning and deep learning approaches for educational prediction tasks. However, there remains a need for comprehensive studies that systematically compare multiple models within the same framework while also analyzing feature importance. This study aims to address this gap by providing an integrated evaluation of machine learning and deep learning models for predicting student selection outcomes.

Methodology

Research Framework

The overall research framework adopted in this study is illustrated in figure 1. The workflow consists of several sequential steps, including data collection, preprocessing, feature engineering, model development, evaluation, and analysis. The process begins with data cleaning and handling missing values, followed by feature transformation and encoding. Subsequently, multiple machine learning and deep learning models are trained using a unified pipeline. The models are then evaluated using standard performance metrics and cross-validation techniques. Finally, feature importance analysis is conducted to interpret the contribution of each variable to the prediction results.

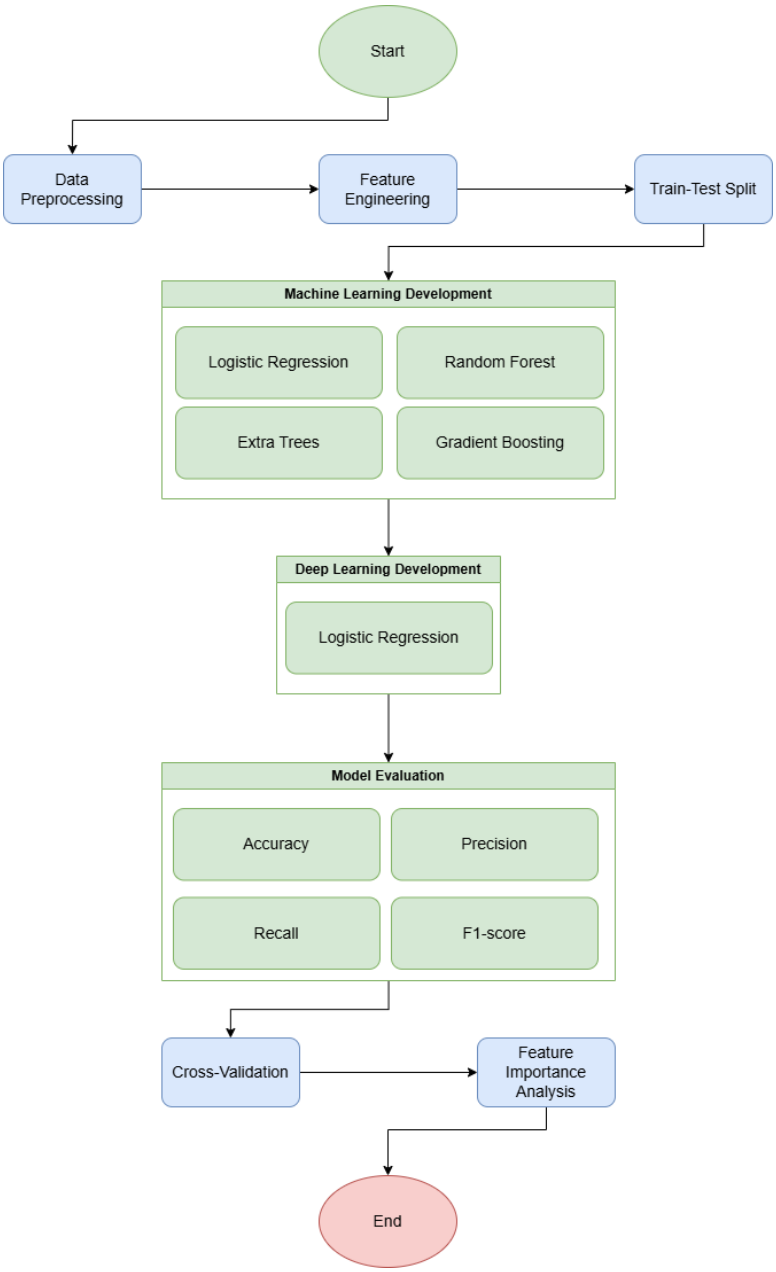


Figure 1 **Research workflow**

Dataset Description

The dataset consists of student selection examination results with both numerical and categorical features. Numerical features include skd.twk, skd.tiu, and skd.tkp, while categorical features include gender, origin, and education level. The target variable represents the selection outcome.

To ensure data quality, records with missing target values were removed. Additionally, due to severe class imbalance, the PA/L class was merged into P/L to improve model stability.

Feature Engineering

A new feature, skd.total, was constructed to represent the overall score:

$$skd.total = skd.twk + skd.tiu + skd.tkp \quad (1)$$

This feature provides a more comprehensive representation of student performance.

Data Preprocessing

Numerical features were processed using median imputation and standardization:

$$x' = \frac{x - \mu}{\sigma} \quad (2)$$

Categorical features were transformed using One-Hot Encoding:

$$x_{cat} \rightarrow [0,1,0,\dots] \quad (3)$$

These transformations were implemented using a ColumnTransformer within a pipeline to ensure consistency.

Model Development

Logistic Regression models the probability using the sigmoid function:

$$P(y = 1 \mid x) = \frac{1}{1 + e^{-(w^T x + b)}} \quad (4)$$

Random Forest combines multiple decision trees:

$$\hat{y} = \frac{1}{T}\sum_{t=1}^{T} h_t(x) \quad (5)$$

$h_t(x)$is the prediction of the t-th tree.

Gradient Boosting builds models sequentially:

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x) \quad (6)$$

$h_m(x)$is the weak learner and $\gamma_m$is the learning rate.

The MLP computes outputs through layered transformations:

$$a^{(l)} = f(W^{(l)} a^{(l-1)} + b^{(l)}) \quad (7)$$

$f$is the ReLU activation function:

$$f(x) = \max(0, x) \quad (8)$$

Evaluation Metrics

The models were evaluated using:

Accuracy

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (9)$$

Precision

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

Recall

$$Recall = \frac{TP}{TP + FN} \tag{11}$$

F1-Score

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{12}$$

Macro averaging was used to ensure equal importance across classes.

Cross-Validation

A 5-fold stratified cross-validation was applied:

$$CV = \frac{1}{k}\sum_{i=1}^{k} Scor\, e_i \tag{13}$$

This approach ensures robust evaluation across different subsets of the data.

Feature Importance

Feature importance in Random Forest is computed based on impurity reduction:

$$Importance(j) = \sum_{t} \Delta\, I_t(j) \tag{14}$$

$\Delta I_t(j)$represents the decrease in impurity from feature j.

Algorithm 1: Student Selection Prediction Framework

Input: Dataset D, features X, target label y

Output: Best prediction model M^*, evaluation metrics, feature importance

Process:

Start

Load dataset D

Remove samples with missing target label:

$$D' = \{(x_i, y_i) \in D \mid y_i \neq \emptyset\}$$

Merge minority class:

if $y_i = PA/L$then

$$y_i = P/L$$

end if

Create total score feature:

$$skd.total = skd.twk + skd.tiu + skd.tkp$$

Define feature set:

$$X = \{jenis.kelamin, kode.asal, pendidikan, skd.twk, skd.tiu, skd.tkp, skd.total\}$$

Encode target label $y$using Label Encoder

Split dataset into training and testing data:

$$(X_{train}, X_{test}, y_{train}, y_{test}) = Split(X, y, 80{:}\,20)$$

Apply preprocessing:

Impute numerical features using median value

Standardize numerical features:

$$x' = \frac{x - \mu}{\sigma}$$

Impute categorical features using most frequent value

Encode categorical features using One-Hot Encoding

Define model set:

$$\mathcal{M} = \{LR, RF, ET, GB, MLP\}$$

For each model $M \in \mathcal{M}$:

Train Musing X_{train}, y_{train}

Predict test label: $\hat{y} = M(X_{test})$

Calculate Accuracy, Precision, Recall, and F1-score

Store evaluation results

end for

Select best model:

$$M^* = \arg\max_{M \in \mathcal{M}} F1_{macro}(M)$$

Perform 5-fold cross-validation:

$$CV = \frac{1}{5}\sum_{i=1}^{5} F\,1_i$$

Calculate feature importance using Random Forest:

$$Importance(j) = \sum_{t=1}^{T} \Delta\, I_t(j)$$

Return M^*, evaluation metrics, and feature importance

End

Results

Class Distribution

The distribution of the target classes after preprocessing is shown in figure 2. Initially, the dataset contained a severe class imbalance, with the PA/L category having only 9 instances compared to thousands of samples in other classes. Such an extreme imbalance can negatively affect model training by causing bias toward majority classes and reducing the model's ability to learn meaningful patterns for minority classes. To mitigate this issue, the PA/L class was merged into the P/L category. This step increases the representation of the minority group and helps stabilize the learning process, especially for algorithms that are sensitive to class distribution.

After preprocessing, the class distribution becomes more balanced, although some imbalance still remains. The P class is the most dominant with 7,744 samples, followed by TL with 5,086 samples. In contrast, P/L and TH remain minority classes with 1,711 and 1,565 samples respectively. This distribution indicates that while the dataset is still skewed, it is within a manageable range for machine learning models, particularly when combined with techniques such as class weighting. The improved balance allows the models to better capture patterns across all classes while reducing the risk of overfitting to the majority class.



Figure 2 Distribution of target classes after preprocessing.

Correlation Analysis

The correlation matrix of numerical features is presented in figure 3 and provides insight into the relationships between individual test components and the overall score. A very strong positive correlation is observed between skd.tiu and skd.total (r = 0.89), indicating that the TIU score contributes significantly to the total score. Similarly, skd.twk shows a strong correlation with skd.total (r = 0.72), while skd.tkp demonstrates a moderate correlation (r = 0.63). These patterns are expected since the total score is derived from the sum of these components, but the varying correlation strengths suggest that TIU has the greatest influence among the three. Additionally, the correlations among the individual components themselves appear relatively lower, indicating that each test measures different aspects of student ability. Overall, these results confirm that the total score is strongly driven by its constituent features and plays a central role in the prediction process.

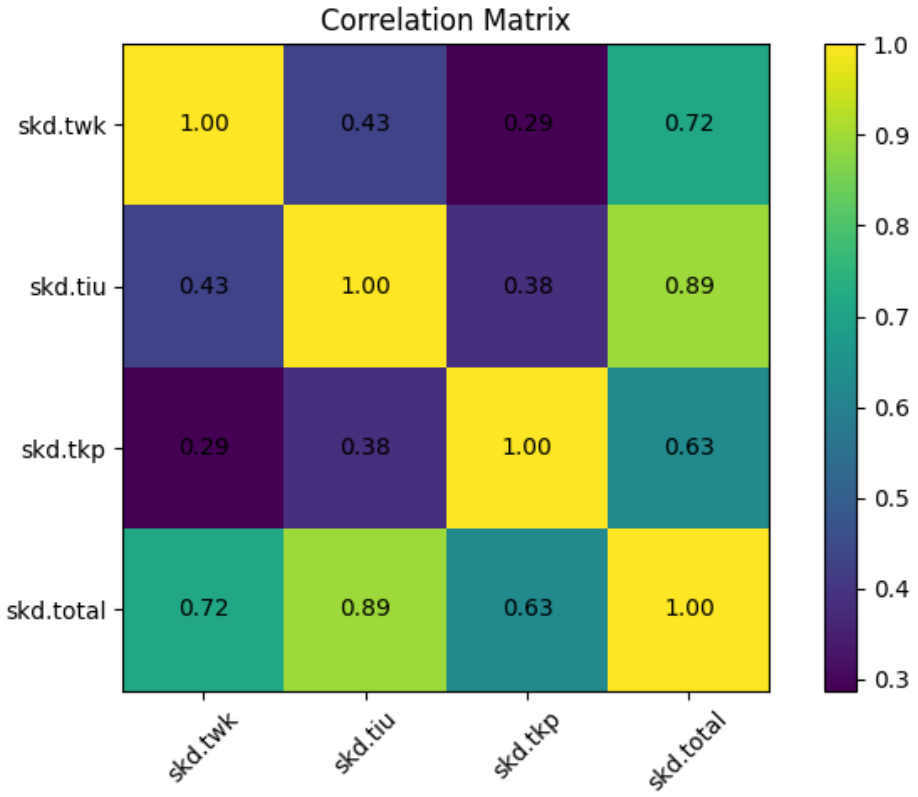


Figure 3 **Correlation matrix of numerical features.**

Model Performance

The performance comparison of all models is summarized in table 1, which presents evaluation metrics including Accuracy, Precision, Recall, and F1-score. Among the evaluated models, Gradient Boosting achieved the highest overall performance, with an accuracy of 0.949 and an F1-score (macro) of 0.920. It also demonstrated the highest precision among all models, indicating strong capability in correctly classifying instances across different classes. Random Forest and MLP showed very competitive results, with F1-scores of 0.917 and 0.918 respectively, suggesting that both ensemble learning and neural network approaches are effective for this classification task. Extra Trees also performed well, although slightly lower than the top three models.

In contrast, Logistic Regression exhibited significantly lower performance, with an accuracy of 0.735 and an F1-score (macro) of 0.720. Although it achieved relatively high recall, its precision was notably lower, indicating a tendency to misclassify samples across classes. This gap in performance highlights the limitation of linear models in capturing complex patterns within the dataset. Overall, the results demonstrate that non-linear models, particularly ensemble-based methods, are better suited for predicting student selection outcomes due to their ability to model complex relationships in the data.

Table 1 **Performance comparison of machine learning and deep learning models**

Model	Accuracy	Precision (Macro)	Recall (Macro)	F1-score (Macro)
Logistic	0.735	0.694	0.835	0.720

Regression				
Random Forest	0.946	0.929	0.908	0.917
Extra Trees	0.937	0.917	0.901	0.908
Gradient Boosting	0.949	0.941	0.905	0.920
MLP (Deep Learning)	0.944	0.924	0.912	0.918

Cross-Validation Results

The robustness of each model was evaluated using 5-fold cross-validation, where the dataset was split into five subsets and each model was trained and tested iteratively across different folds. The results, presented in table 2, include the mean F1-score (macro) and standard deviation for each model. Gradient Boosting achieved the highest mean F1-score of 0.928, followed by Random Forest at 0.921 and MLP at 0.920. Extra Trees also demonstrated competitive performance with a mean F1-score of 0.914, while Logistic Regression showed the lowest performance with a mean of 0.689. These results are consistent with the single train-test evaluation, reinforcing the reliability of the findings.

In addition to performance, the standard deviation values provide insight into model stability across different data splits. All top-performing models, including Gradient Boosting, Random Forest, and MLP, exhibit low standard deviation values around 0.007 to 0.008, indicating consistent performance and good generalization ability. Extra Trees shows even lower variability, although with slightly reduced performance. In contrast, Logistic Regression has a much higher standard deviation of 0.036, suggesting less stable behavior across folds. Overall, the cross-validation results confirm that ensemble-based models, particularly Gradient Boosting, not only achieve high accuracy but also maintain stable performance across different subsets of the data.

Table 2 Cross-validation results (F1 Macro)

Model	Mean F1 Macro	Std Dev
Gradient Boosting	0.928	0.007
Random Forest	0.921	0.008
MLP	0.920	0.008
Extra Trees	0.914	0.005
Logistic Regression	0.689	0.036

Confusion Matrix

The confusion matrix of the best-performing model, Gradient Boosting, is presented in figure 4 and provides a detailed view of the classification results across all classes. The model demonstrates excellent performance in correctly classifying instances of the TH and TL classes, with nearly all samples accurately predicted, indicating that these classes are well separated in the feature space. For the P class, the model also achieves a high number of correct predictions, although a portion of instances are misclassified. The most notable errors occur between the P and P/L classes, where a significant number of P/L instances are incorrectly predicted as P. This pattern suggests that these two classes share similar characteristics, making them more difficult to distinguish. Overall, the confusion matrix confirms that the model performs very well on clearly defined classes, while classification challenges remain for classes with

overlapping feature distributions.

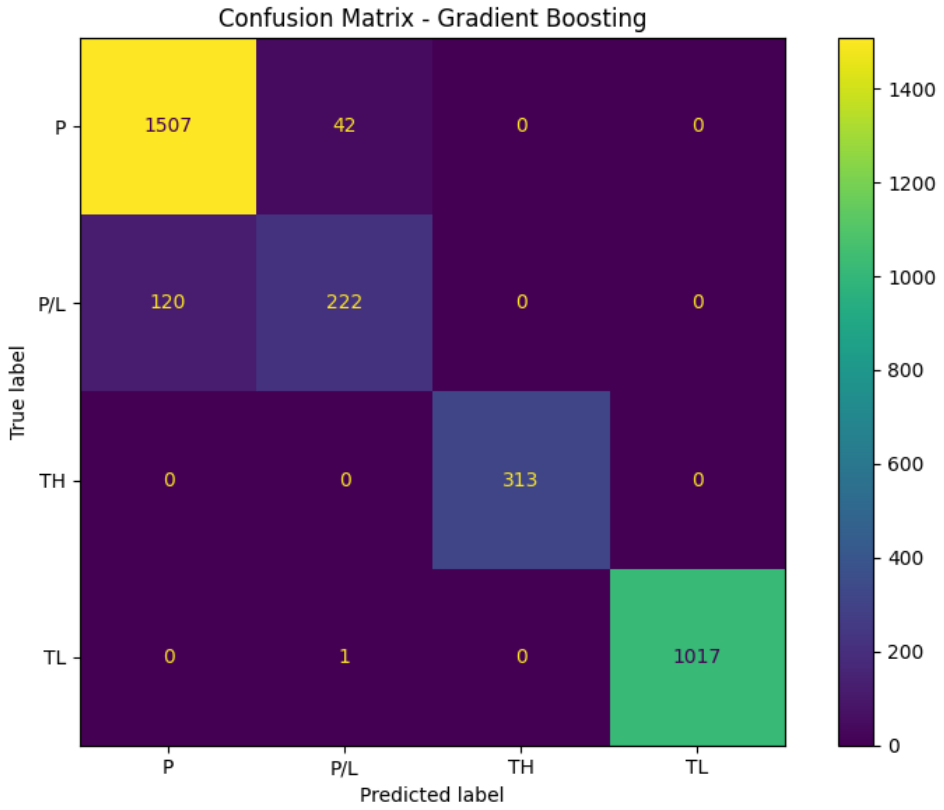


Figure 4 Confusion matrix of Gradient Boosting model.

Feature Importance

Feature importance derived from the Random Forest model is presented in figure 5, highlighting the relative contribution of each feature in predicting student selection outcomes. The most influential feature is skd.total, which has the highest importance score, indicating that the overall test score plays a dominant role in the classification process. This is followed by individual test components, namely skd.twk, skd.tiu, and skd.tkp, which also contribute significantly to the model's predictions. In contrast, demographic features such as gender, educational background, and regional origin exhibit substantially lower importance values, suggesting that they have minimal impact on the prediction results. This distribution of feature importance indicates that the model primarily relies on academic performance indicators rather than personal or demographic characteristics when determining selection outcomes.

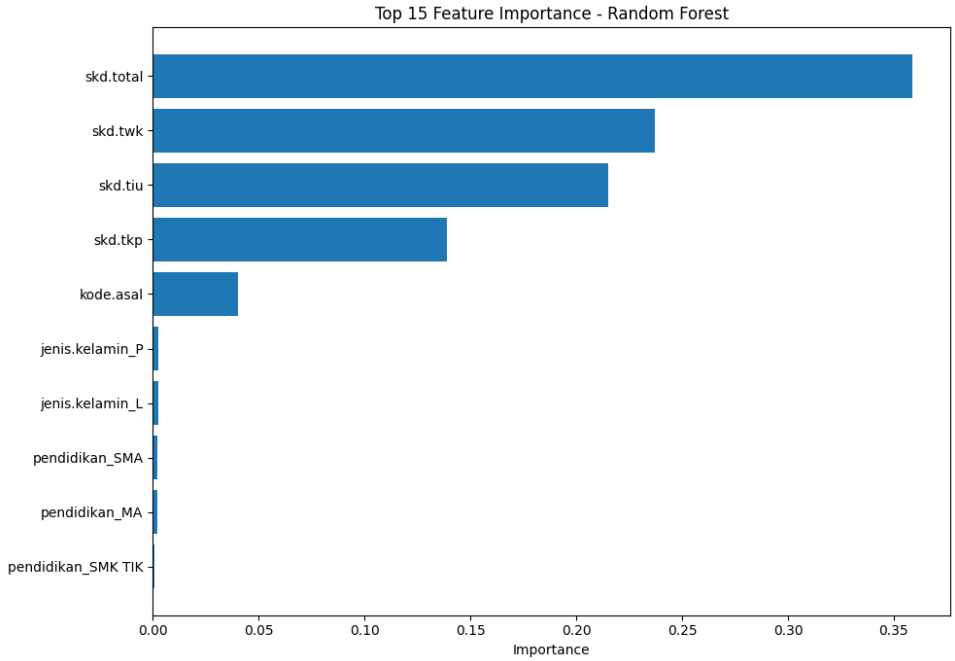


Figure 5 Feature importance based on Random Forest.

Discussion

The experimental results clearly show that ensemble-based models, particularly Gradient Boosting, provide the highest performance in predicting student selection outcomes. This can be explained by the ability of ensemble methods to model complex and non-linear relationships between input features and target classes. By iteratively combining multiple weak learners, Gradient Boosting improves prediction accuracy while reducing both bias and variance. The consistency of its performance across both the test set and cross-validation further demonstrates its robustness and reliability. Random Forest and Extra Trees also perform strongly, confirming that tree-based ensemble approaches are highly effective for this type of structured tabular data. This finding is consistent with previous studies showing that ensemble learning methods often achieve strong predictive performance in educational data mining tasks, particularly when handling heterogeneous academic and non-academic features [21], [22].

In contrast, the Multilayer Perceptron model achieves competitive but slightly lower performance compared to the best ensemble methods. This suggests that deep learning models do not necessarily outperform traditional machine learning techniques when applied to tabular datasets with limited size and well-defined features. Unlike tree-based models, neural networks often require larger datasets, more extensive hyperparameter tuning, and careful optimization to fully capture underlying patterns [23]. The significantly lower performance of Logistic Regression further emphasizes the importance of model selection, as linear methods struggle to capture the non-linear relationships present in the data. Similar findings have been reported in comparative studies where traditional machine learning and ensemble models performed competitively or even better than deep learning models in student performance prediction tasks [24]. Overall, these findings highlight that model complexity alone does not

guarantee better performance, and that ensemble learning remains a highly suitable approach for predictive tasks in educational data, especially when combined with proper evaluation and interpretability techniques [25].

Conclusion

This study conducts a comparative evaluation of machine learning and deep learning models for predicting student selection outcomes using examination data, with a focus on performance, robustness, and feature influence. The results demonstrate that ensemble-based models, particularly Gradient Boosting, consistently achieve the highest performance across evaluation metrics, followed closely by Random Forest and the Multilayer Perceptron model. These findings indicate that tree-based methods are highly effective for structured tabular data due to their ability to capture complex feature interactions without requiring extensive data preprocessing. In contrast, Logistic Regression shows substantially lower performance, highlighting the limitations of linear approaches in modeling non-linear relationships present in the dataset. Additionally, feature importance analysis reveals that academic performance indicators, especially the total score and its components, are the dominant factors influencing prediction outcomes, while demographic variables have minimal impact. This suggests that the selection process is largely driven by measurable examination results. Overall, the study confirms that selecting models appropriate to data characteristics is crucial, and that more complex approaches such as deep learning do not necessarily yield superior results in all contexts, particularly when dealing with structured and moderately sized datasets.

Declarations

Author Contributions

Conceptualization: F.A.R., S.Z.U.; Methodology: S.Z.U.; Software: F.A.R.; Validation: F.A.R., S.Z.U.; Formal Analysis: F.A.R., S.Z.U.; Investigation: F.A.R.; Resources: S.Z.U.; Data Curation: S.Z.U.; Writing Original Draft Preparation: F.A.R., S.Z.U.; Writing Review and Editing: S.Z.U., F.A.R.; Visualization: F.A.R.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, andor publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or

personal relationships that could have appeared to influence the work reported in this paper.

References

[1] C. Romero and S. Ventura, “Educational data mining: A review of the state of the art,” *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 40, no. 6, pp. 601–618, 2010, doi: 10.1109/TSMCC.2010.2053532.

[2] C. Romero, S. Ventura, and E. García, “Data mining in course management systems: Moodle case study and tutorial,” *Computers & Education*, vol. 51, no. 1, pp. 368–384, 2008, doi: 10.1016/j.compedu.2007.05.016.

[3] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[4] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015, doi: 10.1038/nature14539.

[5] J. H. Friedman, “Greedy function approximation: A gradient boosting machine,” *The Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001, doi: 10.1214/aos/1013203451.

[6] T. M. Mitchell, *Machine Learning*. New York, NY, USA: McGraw-Hill, 1997. [Online]. Available: https://www.cs.cmu.edu/~tom/mlbook.html

[7] S. B. Kotsiantis, “Supervised machine learning: A review of classification techniques,” *Informatica*, vol. 31, no. 3, pp. 249–268, 2007. [Online]. Available: https://informatica.si/index.php/informatica/article/view/148

[8] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York, NY, USA: Springer, 2009, doi: 10.1007/978-0-387-84858-7.

[9] P. Geurts, D. Ernst, and L. Wehenkel, “Extremely randomized trees,” *Machine Learning*, vol. 63, no. 1, pp. 3–42, 2006, doi: 10.1007/s10994-006-6226-1.

[10] C. Márquez-Vera, A. Cano, C. Romero, A. Y. M. Noaman, H. M. Fardoun, and S. Ventura, “Early dropout prediction using data mining: A case study with high school students,” *Expert Systems*, vol. 33, no. 1, pp. 107–124, 2016, doi: 10.1111/exsy.12135.

[11] A. Peña-Ayala, “Educational data mining: A survey and a data mining-based analysis of recent works,” *Expert Systems with Applications*, vol. 41, no. 4, pp. 1432–1462, 2014, doi: 10.1016/j.eswa.2013.08.042.

[12] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. [Online]. Available: https://www.deeplearningbook.org/

[13] S. R. K. Branavan, H. Chen, L. Zettlemoyer, and R. Barzilay, “Reinforcement learning for mapping instructions to actions,” in *Proceedings of the Joint Conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, Suntec, Singapore, Aug. 2009, vol. 2009, no. August, pp. 82–90. [Online]. Available: https://aclanthology.org/P09-1010/

[14] H. Abdi, “The method of least squares,” in *Encyclopedia of Measurement and Statistics*, N. J. Salkind, Ed. Thousand Oaks, CA, USA: SAGE, 2007, pp. 530–532. [Online]. Available: https://personal.utdallas.edu/~herve/Abdi-LeastSquares06-

pretty.pdf

[15] S. Shalev-Shwartz and S. Ben-David, *Understanding Machine Learning: From Theory to Algorithms*. Cambridge, U.K.: Cambridge University Press, 2014, doi: 10.1017/CBO9781107298019.

[16] Y. Bengio, "Learning deep architectures for AI," *Foundations and Trends in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009, doi: 10.1561/2200000006.

[17] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "LightGBM: A highly efficient gradient boosting decision tree," in *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 3146–3154.

[18] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, vol. 2016, no. August, pp. 785–794, doi: 10.1145/2939672.2939785.

[19] C. Gray and J. Perkins, "Utilizing early engagement and machine learning to predict student outcomes," *Computers & Education*, vol. 131, no. April, pp. 22–32, 2019, doi: 10.1016/j.compedu.2018.12.006.

[20] A. J. Bowers, R. Sprott, and S. A. Taff, "Do we know who will drop out? A review of the predictors of dropping out of high school: Precision, sensitivity, and specificity," *The High School Journal*, vol. 96, no. 2, pp. 77–100, 2013, doi: 10.1353/hsj.2013.0000.

[21] H. Farhood, I. Joudah, A. Beheshti, and S. Muller, "Evaluating and enhancing artificial intelligence models for predicting student learning outcomes," *Informatics*, vol. 11, no. 3, Art. no. 46, 2024, doi: 10.3390/informatics11030046.

[22] E. Al-Momani, A. Shatnawi, M. Almomani, A. Almomani, and M. Alauthman, "A classification model for predicting course outcomes using ensemble methods," *International Journal of Electrical and Computer Engineering*, vol. 14, no. 6, pp. 7090–7102, 2024, doi: 10.11591/ijece.v14i6.pp7090-7102.

[23] B. Tang, S. Li, and C. Zhao, "Predicting the performance of students using deep ensemble learning," *Journal of Intelligence*, vol. 12, no. 12, Art. no. 124, 2024, doi: 10.3390/jintelligence12120124.

[24] A. Kord, A. Aboelfetouh, and S. M. Shohieb, "Academic course planning recommendation and students' performance prediction multi-modal based on educational data mining techniques," *Journal of Computing in Higher Education*, vol. 38, no. 1, pp. 38–76, 2026, doi: 10.1007/s12528-024-09426-0.

[25] T. Tong and Z. Li, "Predicting learning achievement using ensemble learning with result explanation," *PLOS ONE*, vol. 20, no. 1, Art. no. e0312124, 2025, doi: 10.1371/journal.pone.0312124.