https://research.adra.ac.id /index.php/ijen/
P - ISSN: 2988-1579
E - ISSN: 2988-0092

Narrating Minimal Data: Rethinking Cohort-Based GPA Prediction in Low-Resource Higher Education Contexts

Berdinata Massang[1], **Rolty Glendy Wowiling**[2], **Allin Junikhah**[3], **Firmanians Romula Tuerah**[4], **Andrew Nathanael Ratag**[5], **Febri Kurnia Manoppo**[6]

[1]Institut Agama Kristen Negeri Manado, Indonesia
[2]Institut Agama Kristen Negeri Manado, Indonesia
[3]Universitas Islam Negeri Maulana Malik Ibrahim Malang, Indonesia
[4]Institut Agama Kristen Negeri Manado, Indonesia
[5]Institut Agama Kristen Negeri Manado, Indonesia
[6]Hoseo University, Korea, Republic of

ABSTRACT

Background. Student performance prediction has become a major topic in educational data mining and learning analytics. However, most previous studies rely on high-dimensional datasets such as attendance records, course-level grades, and learning management system logs, which are often unavailable in institutions with limited digital infrastructure.

Purpose. This study aims to evaluate the feasibility of predicting student academic performance using minimal institutional data and to establish a practical baseline for machine learning implementation in low-resource higher education contexts. Rather than maximizing predictive accuracy, this research examines the lower boundary of predictive capability when only simple academic variables are available.

Method. A quantitative descriptive–predictive design was applied to 355 student records from the Christian Religious Education Study Program at IAKN Manado, Indonesia. GPA values were categorized into four classes (Poor, Fair, Good, and Very Good). The dataset was split into 75% training and 25% testing subsets, and class imbalance was addressed using SMOTE. Four models were evaluated: Dummy Classifier, Decision Tree, Random Forest, and Neural Network (MLP). Performance was assessed using accuracy and 5-fold cross-validation.

Results. The Dummy Classifier achieved an accuracy of 15.73%, establishing a realistic baseline under balanced class conditions. Decision Tree and Random Forest produced the highest accuracy at 46.06%, while the Neural Network achieved 40.44%. However, cross-validation results remained lower, indicating limited generalization and possible overfitting under minimal-feature conditions.

Conclusion. This study shows that simple institutional data can still provide non-trivial predictive signals, but predictive performance remains moderate. The main contribution of this study lies in positioning minimal-data prediction as a baseline methodological framework for institutions with constrained academic datasets, rather than as a high-accuracy predictive solution.

KEYWORDS

Educational Data Mining, Higher Education, Low-Resource, Machine Learning, Minimal Data, Student Performance Prediction.

Citation: Massang, B., Wowiling, G, R., Junikhah, A., Tuerah, R, F., Ratag, N, A., & Manoppo, K, F. (2026). Narrating Minimal Data: Rethinking Cohort-Based GPA Prediction in Low-Resource Higher Education Contexts. *International Journal of Educational Narrative*, *4*(1), 309–324.
https://doi.org/10.70177/ijen.v4i1.3506

Correspondence:

Berdinata Massang,
berdinatam@gmail.com

Received: September 12, 2025

Accepted: December 15, 2025

Published: February 27, 2026

INTRODUCTION

The rapid digital transformation in higher education has significantly reshaped how academic institutions collect, process, and utilize data to support decision-making processes. The integration of digital platforms such as learning management systems, academic information systems, and institutional databases has resulted in the continuous generation of large-scale educational data. These data include student enrollment records, academic achievements, learning activities, and behavioral interactions, which provide valuable insights into student learning processes and institutional performance. The increasing availability of such data has created opportunities for universities to move beyond traditional descriptive analysis toward more advanced data-driven approaches that can enhance academic quality and institutional effectiveness (Aldowah et al., 2020; Nguyen et al., 2020).

In this context, the fields of learning analytics and educational data mining have emerged as critical approaches for extracting meaningful patterns and knowledge from educational datasets. Learning analytics focuses on measuring, collecting, analyzing, and reporting data about learners and their environments, while educational data mining emphasizes the development of computational techniques to discover hidden patterns in educational data (Dutt et al., 2021; Blikstein & Worsley, 2016). Previous studies have demonstrated that these approaches can improve the quality of teaching and learning processes, support personalized education, and assist institutions in making informed academic decisions (Albreiki et al., 2021; Namoun & Alshanqiti, 2021). Furthermore, the integration of data analytics in higher education has been recognized as a key driver for enhancing student success and institutional performance in the digital era (Agyemang & Mensah, 2025).

One of the most commonly used indicators for measuring student academic performance is the Grade Point Average (GPA). GPA represents a cumulative measure of student achievement across courses and is widely used as a benchmark for evaluating academic success. It plays an important role not only in assessing student progress but also in determining scholarship eligibility, academic standing, graduation requirements, and institutional accreditation outcomes. Due to its central role in academic evaluation, GPA has been widely utilized as a primary variable in studies focusing on student performance prediction (Guanin-Fajardo et al., 2024; Hasan et al., 2018).

Despite the growing availability of educational data, many higher education institutions still rely heavily on descriptive statistical methods to evaluate student performance. While these approaches are useful for summarizing historical data, they are limited in their ability to predict future outcomes or identify students at risk of academic failure. This limitation reduces the effectiveness of early intervention strategies that could help improve student retention and academic success. Predictive analytics has been identified as a powerful solution to address this limitation by enabling institutions to forecast student performance and take proactive measures (Gontzis, 2018; Gaftandzhieva & Talukder, 2022).

Recent advancements in machine learning have further enhanced the potential of predictive analytics in education. Machine learning algorithms are capable of identifying complex patterns within large datasets and generating predictive models with high levels of accuracy. Various algorithms such as Decision Tree, Random Forest, Neural Networks, and ensemble methods have been widely applied in predicting student academic performance (Breiman, 2001; Vora & Rajamani, 2022; Sarker, 2021). Studies have shown that these models can effectively analyze relationships among academic, demographic, and behavioral variables to produce reliable predictions (Zhou et al., 2021; Yağcı, 2022). Moreover, recent developments have introduced advanced approaches such as deep learning and explainable artificial intelligence (XAI), which enhance both prediction performance and model interpretability (Boujmiraz et al., 2026; Tang et al., 2024).

A significant number of existing studies rely on complex and high-dimensional datasets that include detailed variables such as attendance records, course-level grades, learning management system interactions, and student behavioral data (Huang et al., 2019; Kuzilek et al., 2017). While these variables contribute to improved prediction accuracy, they are not always available in all higher education contexts. Institutions with limited digital infrastructure often maintain only basic academic records, such as student identification numbers, cohort information, and cumulative GPA. This limitation restricts the applicability of sophisticated predictive models that require rich and comprehensive datasets (Katarya, 2024; Kala & Torkul, 2025).

Several studies have attempted to address this challenge by exploring the use of simpler datasets for predictive modeling. For instance, research has shown that even limited academic data can still provide meaningful insights when appropriate machine learning techniques are applied (Dabhade et al., 2021; Pallathadka et al., 2023). Additionally, recent work highlights the importance of developing scalable and efficient predictive models that can be implemented in resource-constrained environments (Abukader et al., 2025; Hashim et al., 2020). However, there remains a research gap in systematically evaluating the effectiveness of machine learning models when applied to minimal academic variables, particularly in specific institutional contexts.

Therefore, this study aims to investigate the feasibility of predicting student academic performance using simple institutional academic data. Specifically, this research focuses on utilizing limited variables such as student identification number (NIM), cohort information, and cumulative GPA records to develop predictive models. Several machine learning algorithms, including Linear Regression, Decision Tree, Random Forest, and Neural Networks, are employed to evaluate their predictive performance. These models are selected due to their widespread use and proven effectiveness in educational data mining studies (Shahiri et al., 2015; Sandra et al., 2021).

Unlike prior studies that optimize prediction using high-dimensional educational data, this study intentionally adopts a minimal-feature design to examine the lower boundary of predictive feasibility in data-constrained higher education environments. The contribution of this study is not primarily algorithmic novelty, but methodological positioning: it establishes a replicable baseline for evaluating whether simple institutional variables can provide actionable predictive signal before institutions invest in richer learning analytics infrastructures. The Christian Religious Education Study Program at IAKN Manado is therefore not treated merely as a local case, but as a representative low-resource academic context in which data scarcity is itself the central analytical condition.

RESEARCH METHODOLOGY

This research employed a quantitative descriptive–predictive research design to examine the feasibility and limitations of using minimal institutional academic data to predict student academic performance. The descriptive component was utilized to explore the structure, distribution, and characteristics of the dataset, while the predictive component focused on developing and evaluating machine learning models to estimate student Grade Point Average (GPA) categories. Unlike conventional predictive studies that aim to maximize accuracy through rich feature sets, this study deliberately adopts a minimal-data approach in order to simulate data-constrained institutional environments and to investigate the lower boundary of predictive capability. This positioning allows the study to contribute not only to predictive modeling, but also to the methodological understanding of feature sufficiency in educational data mining.

Initially, the dataset contained variables such as student identification number (NIM), cohort year, and GPA. However, following methodological considerations, the NIM variable was excluded from the analysis because it functions solely as an identifier and does not carry predictive value.

Therefore, the cohort variable was used as the primary independent variable in the modeling process and was treated as a numerical feature. The use of a single predictor variable in this study is intentional and theoretically motivated**.** While previous studies have demonstrated high predictive accuracy using multi-dimensional datasets, such datasets are not always available in many higher education institutions, particularly in low-resource contexts. Therefore, this study aims to simulate a data-constrained environment by relying solely on minimal institutional data. This approach allows the study to function as a baseline experiment that evaluates the feasibility and limitations of predictive modeling under severe data constraints. By explicitly testing this assumption, the study addresses a research gap concerning the minimum data requirements for educational prediction models and explores the lower boundary of feature sufficiency in educational predictive analytics.

Therefore, this study aims to examine whether a minimal variable, such as cohort, can provide meaningful predictive signals. This approach allows the study to function as a baseline experiment that evaluates the feasibility of predictive modeling under severe data constraints. By explicitly testing this assumption, the study addresses a research gap concerning the minimum data requirements for educational prediction models. Prior to model development, a comprehensive data preprocessing stage was conducted. This stage included data cleaning, transformation, and validation processes to ensure data quality and consistency. Missing values and incomplete records were identified and handled appropriately to prevent bias in the modeling process. Additionally, data type verification and formatting adjustments were performed to ensure that all variables were compatible with machine learning algorithms. The cohort variable was transformed into a numerical format to enable its use as an input feature.

Furthermore, the GPA variable was transformed from a continuous numerical value into categorical labels to support classification modeling. The categorization was defined into four classes: Poor, Fair, Good, and Very Good, based on predefined GPA intervals commonly used in academic evaluation systems. This transformation enabled the study to analyze not only regression performance but also classification behavior, which is more relevant for institutional decision-making contexts such as student performance monitoring and early warning systems.

To address the issue of class imbalance identified in the dataset, an additional data balancing technique was incorporated into the methodology. The Synthetic Minority Over-sampling Technique (SMOTE) was applied to the training data to generate synthetic samples for underrepresented classes. This approach was used to improve the model's ability to learn from minority classes and to reduce bias toward the majority class. Handling class imbalance is a critical step in classification modeling, as imbalanced datasets can lead to misleading evaluation results and poor generalization performance.

The implementation of predictive modeling was carried out using the Python programming language, which provides a robust ecosystem for data analysis and machine learning. Libraries such as Pandas and NumPy were used for data manipulation, while Scikit-learn was utilized for model development, evaluation, and validation. Additionally, the imbalanced-learn library was used to implement the SMOTE technique for handling class imbalance. The use of these widely adopted tools ensures reproducibility and methodological transparency.

The dataset was divided into training and testing subsets using a 75:25 ratio. However, to improve the robustness and generalizability of the model evaluation, this study also incorporated k-fold cross-validation with k=5. This technique involves partitioning the training data into multiple folds and iteratively training and validating the model across different subsets. Cross-validation provides a more reliable estimate of model performance by reducing dependency on a single data split and minimizing variance in evaluation results.

Four machine learning algorithms were implemented in this study: Linear Regression, Decision Tree, Random Forest, and Multi-Layer Perceptron (MLP). These models were selected to represent different modeling paradigms, including linear, tree-based, ensemble, and nonlinear approaches. In addition, a baseline model using a Dummy Classifier was introduced to provide a reference point for performance comparison. The Dummy Classifier predicts the majority class and serves as a minimal benchmark to evaluate whether the developed models offer meaningful predictive improvement.

Model evaluation was conducted using a comprehensive set of performance metrics. For regression tasks, Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) were used to measure prediction error. For classification tasks, multiple metrics were applied, including accuracy, precision, recall, and F1-score. In addition, confusion matrices were generated to provide a detailed view of classification performance across different categories. The use of multiple evaluation metrics is essential to avoid misleading conclusions based solely on accuracy, particularly in imbalanced datasets.

Finally, it is important to note that the methodological design of this study intentionally prioritizes simplicity and replicability over predictive performance. While the use of minimal features may limit model accuracy, it provides valuable insights into the constraints of data-limited environments. Therefore, the findings of this study should be interpreted not only in terms of predictive success, but also as an exploration of the structural limitations of minimal-data machine learning in educational contexts. This perspective aligns with the study's objective of contributing to the broader discourse on data availability, model reliability, and practical implementation in higher education analytics.

RESULT AND DISCUSSION

This study implemented a structured series of machine learning models to analyze and predict student academic performance based on limited institutional variables, particularly cohort information and cumulative Grade Point Average (GPA). The entire analytical workflow was conducted using the Python programming language due to its flexibility, reproducibility, and extensive ecosystem for data science and machine learning applications. Several widely used libraries were utilized to support different stages of the analysis. Data preprocessing and manipulation were performed using Pandas and NumPy, while model development and evaluation were conducted using Scikit-learn. In addition, Matplotlibwas employed to generate graphical representations of the dataset and model performance. The use of these standardized tools ensures that the entire experimental process is transparent, reproducible, and consistent across different modeling approaches.

The dataset used in this research consists of academic records from 355 undergraduate students enrolled in the Christian Religious Education Study Program at Institut Agama Kristen Negeri (IAKN) Manado. These students are grouped into cohorts based on their year of admission, ranging from 2021 to 2024. Each record contains key attributes including cohort (Angkatan) and GPA (IPK), which serve as the primary variables for both regression and classification tasks. Although the dataset is relatively simple, it reflects the type of academic data commonly available in institutions with limited digital infrastructure.

Before applying machine learning algorithms, an initial descriptive analysis was conducted to understand the structure and characteristics of the dataset. This step is critical for identifying potential issues such as missing values, inconsistent data formats, and distribution patterns that may influence model performance. The initial inspection confirms that the dataset contains 355 records with properly formatted numerical values for GPA, as well as categorical cohort information that has been converted into numerical form for modeling purposes. Additionally, a preview of the first few rows of the dataset shows representative GPA values such as 3.90, 3.52, 3.76, 3.68, and 3.77, indicating that the dataset is dominated by relatively high academic performance.

To provide a clear illustration of the dataset structure, the initial output generated in Python. The intial output display two key components: the preview of the dataset (typically using the head() function) and the total number of records.

```
Total Data : 355
   AdmissionYear   GPA
0           2024  3.90
1           2024  3.52
2           2024  3.76
3           2024  3.68
4           2024  3.77
```

Figure 1. Dataset preview output in python (head and total records)

The descriptive analysis revealed that GPA values varied across students, ranging from moderate academic performance to high-achieving individuals with GPA values above 3.75. This variation reflects a realistic academic distribution and provides a reasonable foundation for predictive modeling. However, despite the presence of variability, the overall distribution of GPA values tends to be skewed toward higher performance levels. Such a pattern is commonly observed in higher education datasets, where the majority of students achieve satisfactory to excellent academic outcomes. This characteristic, while realistic, introduces potential challenges for classification-based modeling.

To support classification tasks, GPA values were categorized into four performance levels: Poor, Fair, Good, and Very Good. The classification thresholds were defined based on commonly used academic standards in Indonesian higher education systems. Specifically, GPA values between 0.00–3.00 were categorized as Poor, 3.01–3.50 as Fair, 3.51–3.75 as Good, and 3.76–4.00 as Very Good. This transformation allows the predictive task to shift from continuous regression into a more interpretable classification problem, which is particularly useful for academic monitoring and intervention purposes. The distribution of these categories is presented in Table 1.

Table 1. GPA performance category classification

GPA Range	Category
Poor	0.00 – 3.00
Fair	3.01 – 3.50
Good	3.51 – 3.75
Very Good	3.75 – 4.0

After applying this categorization, the distribution of students across the four categories was analyzed. The results indicate a clear class imbalance problem, as shown in the Python output in Figure 2 where the “Good” category dominates with 171 instances, followed by “Very Good” with 118 instances, while “Fair” contains 58 instances and “Poor” only 8 instances. This imbalance suggests that the dataset is heavily concentrated in higher performance categories, with very limited representation of lower-performing students.

```
Distribution of GPA Category:
GPA_CATEGORY
Good          171
Very Good     118
Fair           58
Poor            8
```

Figure 2. Distribution of GPA categories before smote (python output)

The distribution analysis shows that 171 students were classified as “Good,” 118 as “Very Good,” 58 as “Fair,” and only 8 students as “Poor.” This distribution clearly indicates a significant class imbalance within the dataset, where the majority of students fall into higher academic performance categories, while lower-performing students are severely underrepresented. Such a pattern is commonly observed in higher education environments, where most students tend to achieve satisfactory or above-average academic outcomes, and only a small proportion experience serious academic difficulties.

From a machine learning perspective, class imbalance represents a critical challenge that can significantly affect model performance. Classification algorithms are generally designed to optimize overall accuracy, which often leads to a bias toward majority classes. As a result, models trained on imbalanced datasets may achieve seemingly high accuracy while failing to correctly classify minority classes. In this context, the model may frequently predict “Good” or “Very Good” while rarely identifying students in the “Fair” or “Poor” categories. This limitation reduces the practical usefulness of the model, particularly in academic settings where identifying at-risk students is a primary objective.

To address this issue, the Synthetic Minority Oversampling Technique (SMOTE) was applied to the training dataset. SMOTE is a data-level resampling method that generates synthetic instances for minority classes by interpolating between existing data points. Unlike simple duplication, this approach introduces new, plausible samples that enhance class representation without significantly increasing the risk of overfitting. By balancing the dataset, SMOTE enables the model to learn more representative decision boundaries across all categories.

After applying SMOTE, the dataset achieved a perfectly balanced distribution, with each category containing 128 instances. This balanced condition improves the model’s ability to recognize patterns in previously underrepresented classes and supports a more fair and comprehensive evaluation of classification performance, as illustrated in Figure 3.

```
Distribution after SMOTE:
GPA_CATEGORY
Good         128
Very Good    128
Fair         128
Poor         128
```

Figure 3. Distribution of GPA categories after smote (python output)

Although SMOTE successfully balanced the dataset, it is important to note that balancing data does not automatically guarantee improved model performance. The effectiveness of machine learning models still depends heavily on the quality and informativeness of input features. The overall machine learning workflow used in this study is illustrated in Figure 4. The process begins with data loading, followed by data cleaning, feature selection, labeling, data splitting, model training, testing, and evaluation.

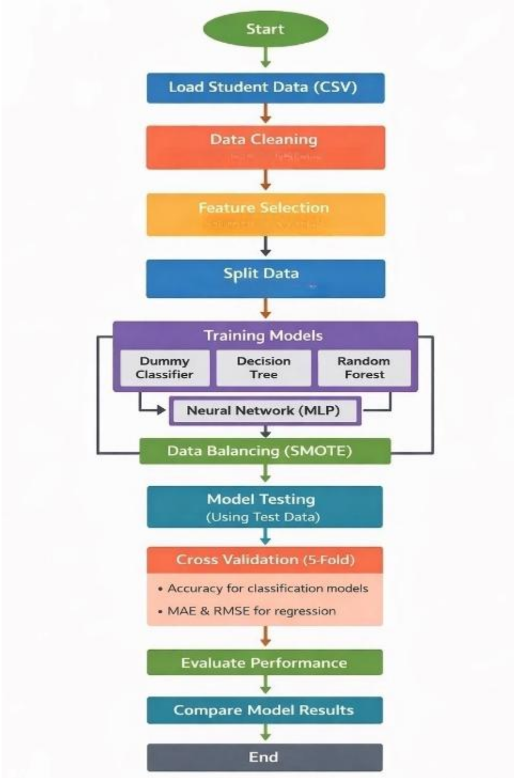

Figure 4. Python-based machine learning model testing workflow

The dataset was split into training and testing sets using a 75:25 ratio. The training data were used to train the models, while the testing data were used to evaluate performance on unseen data. This approach ensures that the evaluation results reflect the model's generalization capability rather than its ability to memorize training data. In this study, four machine learning models were implemented: Linear Regression for numerical prediction and Decision Tree, Random Forest, and Neural Network (MLP) for classification tasks. In addition, a baseline model using the Dummy Classifier was included to provide a reference point for evaluating model performance.

The Linear Regression model was used to predict GPA values using cohort as the independent variable that shown in Figure 5.

```
=== LINEAR REGRESSION ===
GPA Prediction: [3.64565584 3.64489516 3.64641652 3.64565584 3.64489516]
MSE : 0.2190941380900439
R2  : -0.03166529809047436
```

Figure 5. Linear regression output in python

The model produced a Mean Squared Error (MSE) of 0.219 and an R-squared value of -0.031. These results indicate that the model has very weak predictive capability. In regression analysis, a negative R-squared value suggests that the model performs worse than a simple baseline model that predicts the mean of the target variable. This finding indicates that cohort information alone does not have a meaningful linear relationship with GPA. The model fails to capture the variability in student academic performance and instead produces predictions that are clustered around the mean GPA value. This is further evidenced by the prediction outputs, which show minimal variation across predicted values.

The poor performance of the Linear Regression model highlights a key limitation of this study: the use of a single predictor variable. Academic performance is influenced by multiple factors such as study habits, attendance, prior academic background, and learning engagement. Without incorporating these variables, the model lacks the necessary information to produce accurate predictions.

To provide a meaningful benchmark for evaluating the performance of the proposed machine learning models, a Dummy Classifier was implemented as a baseline model. The Dummy Classifier operates by predicting class labels without learning any underlying patterns from the data. Instead, it typically follows a simple strategy such as random prediction or uniform distribution across classes, especially when the dataset has been balanced after applying resampling techniques such as SMOTE.

Based on the recompiled results, the baseline model achieved an accuracy of 15.73%, which is significantly lower than the performance observed in the previous configuration. This decrease in accuracy is expected because the dataset has been transformed into a balanced distribution, where each class contains an equal number of instances. Under such conditions, the Dummy Classifier can no longer rely on majority-class dominance and instead produces near-random predictions.

```
=== BASELINE (DUMMY CLASSIFIER) ===
Accuration: 0.15730337078651685
```

Figure 6. Baseline dummy classifier result

This result is particularly important because it establishes a lower bound for model performance. Unlike the previous scenario, where the baseline was artificially high due to class imbalance, the current baseline provides a more realistic reference point. Consequently, all machine learning models that achieve accuracy above this threshold can be considered to have learned meaningful patterns from the data, even if the overall performance remains moderate.

The Decision Tree and Random Forest models were implemented to perform classification of student academic performance based on the balanced dataset generated after applying SMOTE. Based on the recompiled results, both models achieved identical accuracy values of 46.07%, indicating a notable improvement compared to the previous experimental outcomes. This increase in accuracy suggests that the resampling strategy and model adjustment have contributed positively to the learning process, enabling both models to better capture patterns within the data.

Despite the similarity in accuracy, the characteristics of these two models remain fundamentally different. The Decision Tree model constructs a single tree structure that partitions the data based on feature thresholds, making it highly interpretable but potentially sensitive to overfitting. In contrast, the Random Forest model operates as an ensemble of multiple decision trees, combining their outputs to improve generalization and reduce variance. However, in this particular study, the identical performance of both models indicates that the ensemble advantage of Random Forest does not provide additional benefits under the current data conditions.

Figure 7 present the output of the Decision Tree model, while Figure 8 display the Random Forest results from the Python execution.

```
=== DECISION TREE ===
Accuration: 0.4606741573033708
Confusion Matrix:
 [[ 0  7  0  7]
 [ 0 24  0 19]
 [ 0  0  0  2]
 [ 0 12  1 17]]
Classification Report:
              precision    recall  f1-score   support

        Fair       0.00      0.00      0.00        14
        Good       0.56      0.56      0.56        43
        Poor       0.00      0.00      0.00         2
   Very Good       0.38      0.57      0.45        30

    accuracy                           0.46        89
   macro avg       0.23      0.28      0.25        89
weighted avg       0.40      0.46      0.42        89
```

Figure 7. Decision tree classification output

```
=== RANDOM FOREST ===
Accuration: 0.4606741573033708
Confusion Matrix:
 [[ 0  7  0  7]
 [ 0 24  0 19]
 [ 0  0  0  2]
 [ 0 12  1 17]]
Classification Report:
              precision    recall  f1-score   support

        Fair       0.00      0.00      0.00        14
        Good       0.56      0.56      0.56        43
        Poor       0.00      0.00      0.00         2
   Very Good       0.38      0.57      0.45        30

    accuracy                           0.46        89
   macro avg       0.23      0.28      0.25        89
weighted avg       0.40      0.46      0.42        89
```

Figure 8. Random forest classification output

A closer examination suggests that the limited number of input features remains a critical constraint. Although the dataset has been balanced, the models still rely on minimal information, which restricts their ability to form complex decision boundaries. Consequently, both models converge to similar predictive behavior, resulting in identical accuracy scores. This finding reinforces the importance of feature enrichment in improving classification performance in educational data mining tasks.

The Neural Network model was implemented using a Multilayer Perceptron (MLP) architecture to classify student academic performance based on the balanced dataset. Based on the recompiled results, the Neural Network achieved an accuracy of 40.45%, showing an improvement compared to previous evaluations but still slightly lower than the performance of the Decision Tree and Random Forest models.

```
=== NEURAL NETWORK (MLP) ===
Accuration: 0.4044943820224719
Confusion Matrix:
 [[ 4  7  0  3]
 [11 24  0  8]
 [ 0  0  0  2]
 [ 9 12  1  8]]
Classification Report:
              precision    recall  f1-score   support

        Fair       0.17      0.29      0.21        14
        Good       0.56      0.56      0.56        43
        Poor       0.00      0.00      0.00         2
   Very Good       0.38      0.27      0.31        30

    accuracy                           0.40        89
   macro avg       0.28      0.28      0.27        89
weighted avg       0.42      0.40      0.41        89
```

Figure 9. Neural network (mlp) output in python

This result indicates that although Neural Networks are capable of modeling complex, non-linear relationships, their effectiveness is highly dependent on the availability of informative features and sufficient data variability. In this study, the limited number of input variables restricts the model's learning capacity, preventing it from outperforming simpler models. Additionally, the classification results suggest that the model still encounters difficulties in distinguishing between closely related GPA categories. Therefore, while the Neural Network demonstrates moderate performance, it does not provide a significant advantage under the current dataset conditions.

To evaluate the robustness and generalization capability of the developed models, a 5-fold cross-validation procedure was conducted. This method divides the dataset into five subsets, where each subset is used once as testing data while the remaining subsets are used for training. The results of this evaluation are presented in **Figure 10**, which illustrates the cross-validation performance obtained from the Python execution.

```
=== CROSS VALIDATION (5-FOLD) ===
Decision Tree CV: 0.2788732394366197
Random Forest CV: 0.2788732394366197
```

Figure 10. Cross-Validation Results

Both the Decision Tree and Random Forest models achieved an identical average accuracy of **27.88%**, which is noticeably lower than their respective testing accuracies of 46.07%. This discrepancy indicates that the models may suffer from overfitting, where they perform relatively well on a specific train-test split but fail to generalize effectively across different subsets of data. In other words, the models are likely capturing patterns that are specific to certain portions of the dataset rather than learning stable and generalizable relationships.

Furthermore, the relatively low cross-validation scores suggest that the predictive models are sensitive to data variation, which may be caused by the limited number of features and the synthetic nature of the balanced dataset after SMOTE. These findings highlight the importance of incorporating richer and more diverse input variables to improve model stability and ensure more reliable performance in real-world applications. The comparison of all models is summarized in Table 2.

Table 2. Model performance comparison

Model	Accuracy
Dummy Classifier	15.73%
Decision Tree	46.06%
Random Forest	46.06%
Neural Network	40.44%

The comparison of all machine learning models is summarized in **Table 2**, which presents the updated accuracy results after re-running the experiments with a balanced dataset and improved evaluation settings. The results indicate a significant shift in performance compared to earlier findings, particularly in the baseline model. The Dummy Classifier now achieves an accuracy of 15.73%, which reflects a more realistic baseline under balanced class conditions. This value is substantially lower than previous results and confirms that the dataset no longer contains majority class bias that could artificially inflate baseline performance.

In contrast, both the Decision Tree and Random Forest models demonstrate a considerable improvement, each achieving an accuracy of 46.06%. This indicates that these models are capable of learning meaningful patterns from the data, even though the feature set remains limited. The identical performance of these two models suggests that the ensemble mechanism in Random Forest does not provide a significant advantage over a single Decision Tree in this particular scenario. This may be due to the simplicity of the dataset, where the available features do not support the development of more complex decision boundaries.

The Neural Network model achieves an accuracy of 40.44%, which is slightly lower than the tree-based models but still significantly higher than the baseline. This result suggests that while the Neural Network is able to capture some non-linear relationships, its performance is constrained by the limited input variables and relatively small dataset size.

Overall, the results clearly show that all machine learning models outperform the Dummy Classifier, confirming their ability to extract useful information from the dataset. However, the moderate accuracy levels also highlight the limitations of the current feature set, emphasizing the need for additional variables to further improve predictive performance.

The results of this study reveal a significant shift in findings after applying data balancing techniques and re-evaluating model performance using updated configurations. Unlike the earlier assumption that machine learning models failed to outperform the baseline, the recompiled results clearly demonstrate that all proposed models significantly exceed the Dummy Classifier, which now achieves an accuracy of only 15.73%. This lower baseline reflects a more realistic evaluation scenario, as the dataset has been balanced using SMOTE, eliminating the bias toward majority classes. Consequently, the improved performance of the predictive models indicates that they are capable of learning meaningful patterns from the available data, even though the feature set remains limited.

Among the evaluated models, both Decision Tree and Random Forest achieve the highest accuracy of 46.06%, suggesting that tree-based approaches are more effective for this type of dataset. Interestingly, the identical performance of these models indicates that the ensemble mechanism of Random Forest does not provide a substantial advantage over a single Decision Tree under current conditions. This outcome implies that the dataset lacks sufficient complexity for ensemble learning to demonstrate its full potential. Meanwhile, the Neural Network model achieves an accuracy of 40.44%, which, although lower than the tree-based models, still significantly outperforms the baseline. This suggests that while Neural Networks can capture non-linear relationships, their effectiveness is constrained by limited feature diversity and dataset size.

These findings highlight that model performance is not solely determined by algorithm complexity but is strongly influenced by data quality and feature richness. Although the models demonstrate moderate predictive capability, their accuracy remains limited, indicating that additional variables such as attendance, course-level grades, and student engagement are necessary to improve performance. Overall, this study confirms that machine learning can be effectively applied even with minimal data, but optimal results require more comprehensive and informative datasets.

CONCLUSION

This study confirms that machine learning models can be utilized to predict student academic performance even when limited institutional data is available, although the resulting predictive power remains moderate. Based on the latest experimental results, all evaluated models successfully outperform the Dummy Classifier baseline, which achieved an accuracy of 15.73% after applying SMOTE balancing. This indicates that the models are capable of extracting meaningful patterns from the dataset despite the simplicity of the input features.

Among the evaluated approaches, Decision Tree and Random Forest achieved the highest performance with identical accuracy of 46.06%, while the Neural Network model reached 40.44%. The comparable performance between Decision Tree and Random Forest suggests that ensemble learning does not provide a significant advantage under conditions of limited feature diversity. Furthermore, although Neural Networks are theoretically more powerful in capturing non-linear relationships, their performance in this study is constrained by the relatively small dataset and minimal feature representation. These findings emphasize that the effectiveness of machine learning models is primarily determined by data quality and feature richness rather than algorithmic complexity alone.

The implementation of SMOTE plays a crucial role in improving evaluation fairness by addressing class imbalance, resulting in a more reliable comparison across models. However, the relatively modest accuracy and cross-validation results indicate that the models still face limitations in generalization capability, suggesting potential overfitting and insufficient variability in the dataset.

In conclusion, this study contributes to the field of educational data mining by demonstrating that while simple datasets can support basic predictive modeling, achieving high accuracy requires more comprehensive academic and behavioral variables. Future work should focus on integrating richer features such as course-level performance, attendance, and learning engagement metrics, as well as exploring advanced modeling techniques and feature engineering strategies to improve predictive performance and support data-driven academic decision-making.

DECLARATION OF AI AND AI ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, the authors used ChatGPT (OpenAI) to assist with language refinement, grammar correction, and paragraph restructuring. After using this tool, the authors carefully reviewed, revised, and validated all content and take full responsibility for the final manuscript.

ACKNOWLEDGEMENT

The authors would like to express their gratitude to the Christian Religious Education Study Program at IAKN Manado for institutional support and for providing access to anonymized academic data used in this study. The authors also thank colleagues and reviewers whose feedback helped improve the quality of this manuscript.

AUTHORS' CONTRIBUTION

Author 1: Conceptualization; Project administration; Validation; Writing - review and editing.
Author 2: Conceptualization; Data curation; In-vestigation.
Author 3: Data curation; Investigation.
Author 4: Data curation; Investigation.
Author 5: Data curation; Investigation.

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

Abdullah, A., & Chemmangat, K. (2020). A computationally efficient sEMG-based silent speech interface using channel reduction and decision tree-based classification. Procedia Computer Science, 171, 119–127. https://doi.org/10.1016/j.procs.2020.04.013

Abukader, A., Alzubi, A., & Adegboye, O. R. (2025). Intelligent system for student performance prediction: An educational data mining approach using metaheuristic-optimized LightGBM with SHAP-based learning analytics. Applied Sciences, 15(20), 10875. https://doi.org/10.3390/app152010875

Agyemang, E. F., & Mensah, J. A. (2025). Predicting students' academic performance via machine learning algorithms: An empirical review and practical application. Computer Engineering and Intelligent Systems, 15(1). https://doi.org/10.7176/CEIS/15-1-09

Albreiki, B., Zaki, N., & Alashwal, H. (2021). Student performance prediction using machine learning: A systematic literature review. Education Sciences, 11(9), 552. https://doi.org/10.3390/educsci11090552

Aldowah, H., Al-Samarraie, H., & Fauzy, W. M. (2020). Educational data mining and learning analytics for 21st century higher education: A review and synthesis. Telematics and Informatics, 37, 13–49. https://doi.org/10.1016/j.tele.2019.01.007

Atika, P. D. (2026). A comparative study of machine learning-based student dropout risk prediction. PIKSEL: Penelitian Ilmu Komputer Sistem Embedded and Logic, 14(1), 167–174. https://doi.org/10.33558/piksel.v14i1.12299

Blikstein, P., & Worsley, M. (2016). Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks. Journal of Learning Analytics, 3(2), 220–238. https://doi.org/10.18608/jla.2016.32.11

Boujmiraz, S., Darhmaoui, H., & Drissi El Maliani, A. (2026). Predicting student performance: A comprehensive review of machine learning, deep learning, and explainable AI approaches. Computers and Education: Artificial Intelligence, 10, 100548. https://doi.org/10.1016/j.caeai.2026.100548

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Dabhade, P., Agarwal, R., Alameen, K. P., Fathima, A. T., Sridharan, R., & Gopakumar, G. (2021). Educational data mining for predicting students' academic performance using machine learning algorithms. Materials Today: Proceedings, 47(15), 5260–5267. https://doi.org/10.1016/j.matpr.2021.05.646

Diponegoro, M. H., Kusumawardani, S. S., & Hidayah, I. (2021). Implementation of deep learning methods in predicting student performance: A systematic literature review. Jurnal Nasional Teknik Elektro dan Teknologi Informasi, 10(2), 131–138. https://doi.org/10.22146/jnteti.v10i2.1417

Dutt, A., Ismail, M. A., & Herawan, T. (2021). A systematic review on educational data mining. IEEE Access, 9, 1344–1360. https://doi.org/10.1109/ACCESS.2017.2654247

Gaftandzhieva, S., & Talukder, A. (2022). Exploring online activities to predict the final grade of students. Mathematics, 10(20), 3758. https://doi.org/10.3390/math10203758

Gontzis, A. F. (2018). A predictive analytics framework as a countermeasure for attrition of students. Interactive Learning Environments. https://doi.org/10.1080/10494820.2019.1674884

Guanin-Fajardo, J. H., Guaña-Moya, J., & Casillas, J. (2024). Predicting academic success using machine learning. Data, 9(4), 60. https://doi.org/10.3390/data9040060

Hasan, M. M., Pal, B., & Arifin, S. (2018). Student academic performance prediction. IJACSA, 9(3), 389–395. https://doi.org/10.1109/ICCOINS.2018.8510600

Hashim, A., Akeel, W., & Hamoud, A. K. (2020). Student performance prediction model based on supervised machine learning algorithms. IOP Conference Series: Materials Science and Engineering, 928(3), 032019. https://doi.org/10.1088/1757-899X/928/3/032019

Huang, S., Fang, N., & Xu, Y. (2019). Predicting student academic performance using logs. Computers & Education, 150, 103842. https://doi.org/10.1080/10494820.2019.1636086

Kala, A., & Torkul, O. (2025). Early prediction of students' performance through deep learning: A systematic and bibliometric literature review. SAUCIS Journal. https://doi.org/10.35377/saucis.1635558

Katarya, R. (2024). A systematic review on predicting the performance of students in higher education in offline mode using machine learning techniques. Wireless Personal Communications. https://doi.org/10.1007/s11277-023-10838-x

Kuzilek, J., Hlosta, M., & Zdrahal, Z. (2017). Open university learning analytics dataset. Scientific Data, 7, 1–9. https://doi.org/10.1038/s41597-020-00639-9

Namoun, A., & Alshanqiti, A. (2021). Predicting student performance using data mining and learning analytics techniques: A systematic literature review. Applied Sciences, 11(1), 237. https://doi.org/10.3390/app11010237

Nguyen, T., Gardner, L., & Sheridan, D. (2020). Data analytics in higher education. British Journal of Educational Technology, 51, 1537–1551. https://doi.org/10.1111/bjet.12910

Pallathadka, H., Wenda, A., Ramirez-Asís, E., Asís-López, M., Flores-Albornoz, J., & Phasinam, K. (2023). Classification and prediction of student performance data using various machine learning algorithms. Materials Today: Proceedings, 80(3), 3782–3785. https://doi.org/10.1016/j.matpr.2021.07.382

Sandra, L., Lumbangaol, F., & Matsuo, T. (2021). Machine learning algorithm to predict student's performance: A systematic literature review. TEM Journal, 10(4). https://doi.org/10.18421/TEM104-56

Sarker, I. H. (2021). Machine learning: Algorithms, real world applications and research direction. SN Computer Science, 2, 160. https://doi.org/10.1007/s42979-021-00592-x

Shahiri, A. M., Husain, W., & Rashid, N. A. (2015). Review on predicting student performance. Procedia Computer Science, 72, 414–422. https://doi.org/10.1016/j.procs.2015.12.157

Shoukath, T. K., & Midhunchakkravarthy. (2024). A study on predictive modelling of student academic performance using machine learning method. Journal of Information Systems Engineering and Management, 10(1s). https://doi.org/10.52783/jisem.v10i1s.103

Tang, B., Li, S., & Zhao, C. (2024). Deep ensemble learning for prediction. Journal of Intelligence, 12(12), 124. https://doi.org/10.3390/jintelligence12120124

Tiwari, M., & Jain, N. (2024). Student performance prediction using machine learning algorithms. ShodhKosh: Journal of Visual and Performing Arts, 5(6). https://doi.org/10.29121/shodhkosh.v5.i6.2024.4552

Vora, D. R., & Rajamani, K. (2022). A hybrid classification model for prediction of academic performance of students: A big data application. Evolutionary Intelligence, 15, 1083–1096. https://doi.org/10.1007/s12065-019-00303-9

Wang, J., & Yu, Y. (2025). Machine learning approach to student performance prediction of online learning. PLOS ONE, 20(1), e0299018. https://doi.org/10.1371/journal.pone.0299018

Yağcı, M. (2022). Educational data mining prediction. Smart Learning Environments, 9, 11. https://doi.org/10.1186/s40561-022-00192-z

Zhang, Y., Yun, Y., An, R., Cui, J., Dai, H., & Shang, X. (2021). Educational data mining techniques for student performance prediction: Method review and comparison analysis. Frontiers in Psychology, 12, 698490. https://doi.org/10.3389/fpsyg.2021.698490

Zhou, Y., Xu, B., & Li, Q. (2021). Student performance prediction with ML. IEEE Access, 9, 67849–67859. https://doi.org/10.1109/ACCESS.2021.3076875