Engineering Students' Performance Prediction on Board Examination Using Classification Algorithms

Jayson A. Batoon[1,*] and Sarah Jane L. Cabral[2]

[1]College of Information and Communications Technology, Faculty, Bulacan State University, Malolos, Philippines
[2]Information Technology Department, College of Engineering Eastern Visayas State University, Tacloban City Leyte, Philippines
Email: jasyon.batoon@bulsu.edu.ph (J.A.B.); sarahjane.cabral@evsu.edu.ph (S.J.L.C.)
*Corresponding author

Manuscript received March 12, 2025; revised April 24, 2025; accepted May 15, 2025; published September 11, 2025

***Abstract*—Board examinations are critical assessments that determine the academic and professional readiness of engineering students. Accurately predicting board exam outcomes can support timely interventions, helping institutions and educators enhance student preparedness. This study developed a predictive model using machine learning classification algorithms, specifically logistic regression, decision trees, random forest, and Naïve Bayes, to forecast the board examination performance of engineering students based on academic and preparatory indicators such as general weighted average, pre-board scores, and review center participation. Among the models tested, logistic regression achieved the highest accuracy (66.7%), closely followed by Naïve Bayes (66.1%). The findings emphasize the predictive value of pre-board performance and institutional review programs. This research highlights how predictive analytics can improve educational strategies and support systems, ultimately aiming to raise board exam success rates. Future research is encouraged to integrate additional variables, including psychological and behavioral factors, to further enhance model accuracy.**

***Keywords*—board exam prediction, machine learning, classification algorithm**

I. Introduction

In education, assessing and foreseeing student achievements play a pivotal role in shaping their academic trajectories. This significance is notably pronounced in engineering education, given the demanding academic hurdles and pivotal assessments such as board examinations that dictate their career opportunities and professional licenses [1]. Understanding the factors influencing students' performance in these examinations offers critical perspectives for educators, policymakers, and the students themselves [2].

In recent years, the emergence of machine learning and data analysis techniques has revolutionized the handling and examination of educational data [3]. Educational Data Mining (EDM) and learning analytics have enabled institutions to extract meaningful patterns from large datasets to predict academic success, support decision-making, and design targeted interventions [4]. In particular, the application of classification algorithms to forecast students' board examination outcomes has shown promising results, offering early identification of at-risk students and facilitating proactive academic support.

Several studies have demonstrated the power of predictive models in educational contexts. For instance, Albreiki *et al.* [5] explored the use of machine learning algorithms for predicting student academic performance, achieving notable predictive accuracies. Similarly, Zawacki-Richter *et al.* [6] emphasized the increasing role of artificial intelligence applications in enhancing educational outcomes in higher education. Muchuchuti *et al.* [7] also highlighted the effectiveness of feature selection and classification models in accurately predicting student performance. These developments affirm the importance and relevance of predictive analytics in education, particularly in high-stakes scenarios such as licensure examinations [8].

The focus of this study is to develop a predictive model using classification algorithms to forecast the board examination performance of engineering students. Board exams, also known as licensure exams, are standardized assessments that evaluate a student's competence and eligibility to practice professionally in the engineering field [9]. By analyzing various indicators such as prior academic achievements, study habits, preparation through review centers, and pre-board examination results, this study aims to build an accurate prediction model capable of identifying students who may need additional academic support. To provide a clearer direction for the study, this research is structured around three guiding questions. First, it seeks to determine which academic and preparatory factors such as General Weighted Average (GWA), pre-board examination scores, review center participation, and student enrollment status most significantly influence engineering students' success in licensure examinations. Second, it aims to identify which among the applied machine learning algorithms Logistic Regression, Decision Tree, Random Forest, and Naïve Bayes offers the most reliable and accurate predictions of board exam outcomes. Lastly, the study explores how the predictive insights generated by these models can be translated into practical strategies and interventions that educational institutions can implement to support at-risk students and enhance overall board exam performance. By addressing these questions, the study positions itself not only as a technical evaluation of predictive models but also as a tool for informing evidence-based educational practices.

This research offers several significant contributions. Firstly, it provides students with deeper insights into their academic standing, allowing them to adjust their study strategies effectively. Secondly, educational institutions can utilize the predictive model to detect at-risk students early and implement focused interventions to improve board exam outcomes [10]. Furthermore, policymakers and academic administrators can leverage the findings to refine educational policies and enhance institutional support mechanisms.

To accomplish these objectives, the study employed various classification algorithms, including logistic regression, decision trees, naïve Bayes, and random forests. These algorithms were trained and evaluated using a curated dataset of engineering graduates' academic and examination records. Rigorous data preprocessing ensured data quality and model integrity throughout the process [11].

Through this work, the study seeks to contribute to the growing body of knowledge in educational data analytics by demonstrating how machine learning techniques can be leveraged to forecast academic success, particularly in the context of professional licensure examinations. The findings could have a substantial impact on improving student outcomes, informing institutional strategies, and supporting students in achieving their academic and career goals.

II. Research Methodology

In this study, various classification algorithms were utilized to predict the performance of engineering students in their board examinations, and a comparative analysis of the results was conducted. The methodology encompassed key stages shown in Fig. 1. The first stage encompassed data preprocessing, where the dataset was meticulously prepared, consolidated, and cleaned to ensure the accuracy and reliability required for subsequent stages. This step was critical for eliminating inconsistencies and addressing missing or irrelevant data. The second stage centered on Exploratory Data Analysis (EDA), allowing researchers to delve into the structure of the dataset, comprehend the types of variables present, and uncover relationships between these variables. EDA also provided valuable insights into potential trends or patterns that may influence the prediction models. The third and final stage revolved around evaluating and optimizing the classification performance of each machine learning algorithm employed in the study. Classification algorithms such as decision trees and Bayesian networks can be applied to educational data to predict a student's exam success [12]. By assessing the performance of commonly utilized algorithms, the researchers were able to identify the model that yielded the highest accuracy and reliability in predicting student performance in board examinations [13].

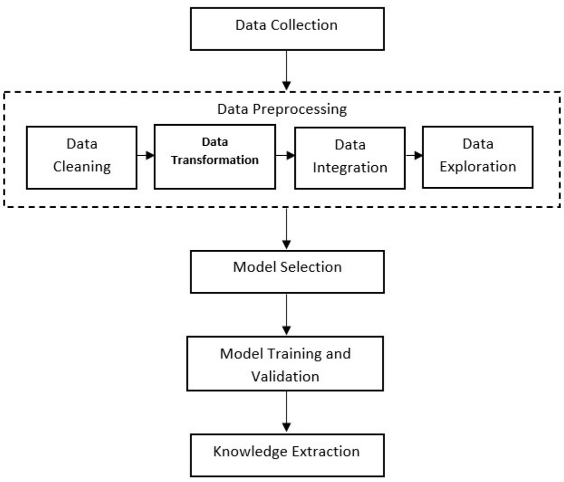

Fig. 1. Project methodology stages.

A. Data Preparation and Selection

This study employed a structured methodology encompassing data preprocessing, Exploratory Data Analysis (EDA), and model development to predict engineering students' board examination performance. In the data preprocessing stage, missing values in numerical features such as GWA and pre-board scores were addressed using mean imputation, while missing categorical values, including review center participation and enrollment status, were filled using mode imputation. Records with excessive missing data were excluded to maintain dataset integrity. Categorical variables were encoded in binary form for example, 0 for "No" and 1 for "Yes" and numerical features were normalized using min-max scaling to ensure consistent value ranges across all input variables. The dataset was then split into training and testing subsets in a 70:30 ratio, and stratified 10-fold cross-validation was employed to evaluate model stability while preserving the proportion of pass/fail outcomes in each fold.

Exploratory data analysis followed to uncover relationships among variables and identify trends relevant to student performance. Descriptive statistics, correlation matrices, and visual tools such as scatter plots, violin plots, and bar charts revealed that pre-board scores and participation in review programs were positively correlated with board exam success, while being a retaker or having a lower GWA showed weaker or negative associations. These insights informed the selection of classification algorithms used for model development.

Table 1 outlines the predictive attributes used to evaluate student performance, particularly in the context of their potential success in board examinations. Each attribute captures a critical factor that may influence a student's outcome. Here's a breakdown and discussion of each attribute:

GWA (General Weighted Average): This measures a student's overall academic performance. It is categorized into intervals from 1.0 to 5.0, representing the range from excellent to below-average performance. A lower GWA indicates better academic standing, which is often predictive of a student's ability to pass exams. Academic consistency, as reflected in a student's GWA, is a strong indicator of foundational knowledge and learning habits [14].

Review Center Admission: Whether a student attended a review center plays an important role in exam preparedness. Review centers are often designed to offer focused and exam-specific training, which can enhance a student's chances of passing. Hence, students who enrolled in review centers are expected to perform better [15].

Scholarship: Students who secured scholarships during college tend to have a history of high performance, as scholarships are usually awarded to those with exceptional academic or extracurricular achievements [16]. This attribute provides insight into a student's academic motivation and capability.

Pre-board Exam Result: The pre-board exam result serves as a direct predictor of the final board exam outcome [17]. A higher pre-board score typically indicates better preparedness and a higher likelihood of passing the board exam [18].

Regular Student: This attribute distinguishes between regular students and those with irregular enrollment patterns (e.g., those who may have taken a leave of absence or repeated subjects). Regular students generally follow the standard academic progression, which could positively impact their board exam performance due to continuous learning without significant interruptions.

Retaker: Students who are not taking the board exam for the first time are labeled as retakers. Previous failure in the exam may indicate gaps in knowledge or exam strategies, which could affect their chances of passing on subsequent attempts.

Board Exam Result: This is the outcome variable, which categorizes students based on whether they passed or failed the board exam. It is the dependent variable that the study aims to predict using the above attributes.

Table 1. Students prediction attributes

Attribute	Description	Category/Interval	Possible Value
GWA	Student's General Weighted Average	1.00–1.24	100, 99, 98, 97
		1.25–1.49	96, 95, 94
		1.50–1.74	93, 92, 91
		1.75–1.99	90, 89, 88
		2.00–2.24	87, 86, 85
		2.25–2.49	84, 83, 82
		2.50–2.74	81, 80, 79
		2.75–2.99	78, 77, 76
		3.00	75
		5.00	74 and below
Review Center Admission	Student's enrollment in a review center	-	0—No, 1—Yes
Scholarship	Student's scholarship status during college	-	0—No, 1—Yes
Pre-board Exam Result	Student's pre-board examination score	-	Score out of 100
Regular Student	Student's enrollment status during college	-	0—No (Irregular), 1—Yes (Regular)
Retaker	Whether the student is a first-time taker or not	-	0—No, 1—Yes
Board Exam Result	Final board examination outcome	-	0—Failed, 1—Passed

In this research, various classification algorithms were utilized to forecast the academic achievement of engineering students in their board examinations. The algorithms selected, including decision trees, logistic regression, random forest, and Naïve Bayes, were harnessed to capitalize on their unique strengths, enhancing the precision of predictions [19]. This segment delves into the significance of each algorithm in the data extraction procedure and underscores the computational instruments utilized for their execution.

1) Decision tree

When it comes to classification algorithms, two popular choices are decision trees and logistic regression. Decision trees are known for their simplicity and interpretability, as they create a tree-like structure by partitioning the dataset based on input features [20]. Each branch of the tree represents a decision path leading to a final prediction. This method is useful for handling both categorical and numerical data, making it valuable for interpreting relationships between variables affecting student performance in board exams.

2) Logistic regression

On the other hand, logistic regression is a widely used statistical algorithm specifically designed for binary classification problems [21]. By modeling the probability of a specific outcome occurring based on input features, logistic regression transforms the data into a probability score between 0 and 1. This makes it suitable for predicting whether a student will succeed or fail in their board exams. Logistic regression is particularly effective when assessing the impact of multiple predictor variables on a specific outcome. Both decision trees and logistic regression have their strengths and can be utilized depending on the nature of the problem at hand.

3) Random forest

Random forest is an ensemble method that combines the output of multiple decision trees to enhance prediction accuracy [22]. By constructing a "forest" of decision trees through a process called bootstrap sampling and random feature selection, random forest minimizes the variance often observed in individual decision trees. This results in a more robust model, less prone to overfitting. The random forest algorithm proved to be highly effective for this study due to its ability to handle complex datasets, particularly in scenarios where numerous features or variables were involved.

4) Naïve bayes

Naïve Bayes is a probabilistic algorithm based on Bayes' theorem, characterized by the assumption that all features are independent. Despite this "naive" assumption, the algorithm has demonstrated strong performance in various classification tasks, especially when dealing with high-dimensional data. In the context of predicting student performance, Naïve Bayes offered simplicity and scalability, making it suitable for the dataset used in this study [23]. Its efficiency in handling large amounts of data contributed to its inclusion among the classification techniques employed.

B. Combining Algorithmic Strengths

The author utilized four classification algorithms to investigate various angles in forecasting student outcomes. Each algorithm presented specific benefits:

- Decision Trees yielded transparent and understandable models.
- Logistic Regression provided insights into probabilities.
- Random Forest employed ensemble learning to enhance accuracy and reliability.
- Naïve Bayes excelled in scalability and managing high-dimensional data.

This thorough strategy facilitated the consideration of multiple variables in predicting the academic performance of engineering students during their board examinations. By integrating these algorithms, the research could leverage the distinct advantages of each model, thereby improving the overall predictive capability [24]. The selection of classification algorithms was guided by both the nature of the dataset and established precedents in related educational data mining research. Logistic Regression was chosen for its effectiveness in binary classification problems and its interpretability, which is important when communicating findings to educational stakeholders. It is widely used in academic risk prediction due to its ability to model the probability of outcomes based on linear relationships between independent variables and the target.

Decision Trees were selected for their intuitive, rule-based structure, which is highly interpretable and capable of handling both numerical and categorical variables without requiring extensive preprocessing. This makes them ideal for uncovering relationships between student characteristics and exam outcomes.

Random Forest, as an ensemble method, was included due to its robustness and ability to reduce overfitting a common issue with single decision trees by aggregating multiple tree predictions. Its proven performance in handling heterogeneous data makes it particularly well-suited to complex educational datasets with mixed variable types.

Lastly, Naïve Bayes was chosen for its computational efficiency and strong performance on high-dimensional data. Despite its assumption of feature independence, it has been shown in prior studies to yield competitive results in academic performance prediction tasks, especially when features are weakly correlated, as observed in this study.

By incorporating these four algorithms, the study aimed to balance interpretability, computational efficiency, and predictive performance, while drawing on best practices from similar predictive modeling research in education.

C. Data Mining Tool

In order to implement the classification algorithms, this study utilized Google Colab, a cloud-based Jupyter notebook environment. Google Colab offers a free platform for developing and executing machine learning models, eliminating the need for local hardware resources. By utilizing Google Colab's cloud infrastructure, the author was able to access high-performance computational power, facilitating the efficient execution of complex algorithms. Moreover, the flexibility of Google Colab allowed the author to work from any device with internet access, providing convenience and enabling focused algorithmic development. The utilization of Google Colab significantly minimized computational challenges and ensured the smooth execution of machine learning code. When combined with the selected algorithms, this platform provided a powerful framework for accurately predicting student outcomes in engineering board exams.

III. Result and Discussion

A. Dataset Exploration

Exploratory Data Analysis (EDA) serves as the phase in the data analysis journey, offering valuable insights into the underlying patterns, characteristics, and relationships within a dataset. This exploration aims to understand the data's structure, uncover hidden trends, and identify potential issues or anomalies. EDA involves techniques such as data visualization, summary statistics, and data transformation. By visualizing data through plots, charts, and graphs, researchers can gain a holistic view of the variables and their distributions. The prediction of students' academic success is a major topic of interest for both educational institutions and students, underlining the importance of doing comprehensive study to find elements impacting performance through data mining. Key factors influencing student results include demographic background, socioeconomic status, parental education levels, involvement in extracurricular activities, instructional quality, and student learning behaviors [7].

The author intends to utilize pie chart to visualize the distribution of passers across different programs. This graphical representation displays the percentage of students who passed the board exam in each program, allowing for a clear understanding of the relative success rates among the various programs.

In Fig. 2, a set of pie charts is presented to illustrate the distribution of pass and fail rates within different engineering disciplines. These charts use color coding, with yellow indicating the percentage of students who passed the board examination, and red indicating those who failed. Fig. 2 encompasses seven significant engineering fields, namely Electrical Engineering, Industrial Engineering, Chemical Engineering, Geodetic Engineering, Civil Engineering, Electronics Engineering, and Mechanical Engineering.

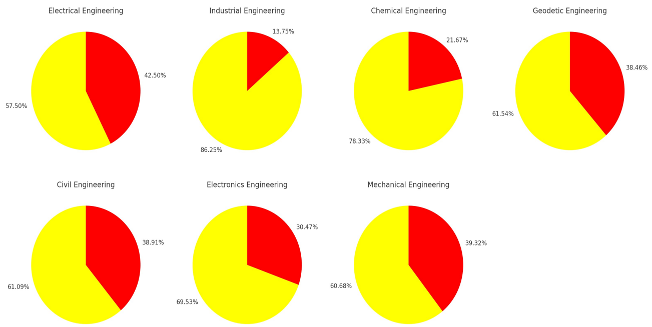

Fig. 2. Pass-fail distribution across different engineering disciplines.

- 61.09% of the graduates of Civil Engineering program passed the board examination and 38.91% have not.
- 60.68% of the graduates of Mechanical Engineering program passed the board examination and 39.32% have not.
- 61.54% of the graduates of Geodetic Engineering program passed the board examination and 38.46% have not.
- 69.53% of the graduates of Electronics Engineering program passed the board examination and 30.47% have not.
- 57.50% of the graduates of Electrical Engineering program passed the board examination and 42.50% have not.
- 86.25% of the graduates of Industrial Engineering program passed the board examination and 13.75% have not.

Fig. 3 provides a clear comparison of the number of students who passed versus those who failed the board examination. The bar chart illustrates two categories: students who passed (represented in blue) and students who failed (represented in red). A bar graph was employed to present a visual representation of the total count of passers and failures in the board exam. Fig. 3 shows that out of 1,041 students who have taken the board exam, 461 passed and 580 failed.

Fig. 4 shows that based on the observation from the years 2018 to 2023, it is notable that the year 2022 exhibited a substantial number of board exam passers.

The correlation matrix in Fig. 5 offers valuable insights into the variables influencing performance on board exams. The correlation values range from −1 to 1, where positive values indicate a positive relationship, and negative values suggest an inverse relationship between variables. It highlights that preboard exam scores and engagement in review programs exhibit notably positive correlations with board exam achievement, underscoring their predictive significance and potential for enhancing student outcomes. Conversely, the limited or negative correlations associated with GWA and exam retakes imply that academic performance in college and reattempting the exam may not ensure success in the board exams. These results suggest that institutions can benefit from prioritizing the enhancement of review programs and preboard evaluations to elevate pass rates in board examinations.

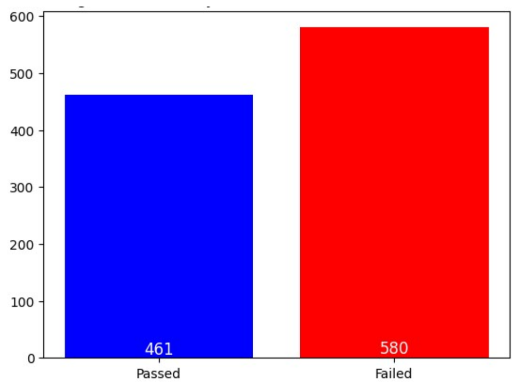

Fig. 3. Pass and fail count in board examination.

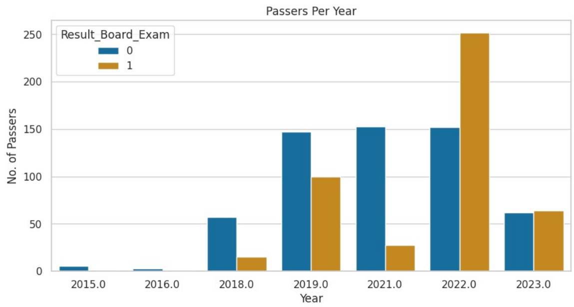

Fig. 4. Board passers per year.

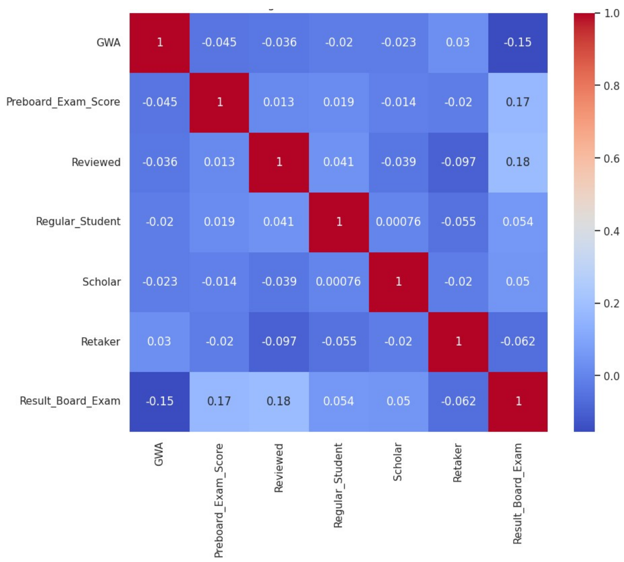

Fig. 5. Correlation matrix.

One of the main findings of the study reveals a significant positive correlation (0.17) between preboard exam scores and board exam results. This suggests that students who excel in their preboard exams are more likely to succeed in the board exam, indicating the predictive value of preboard exams in determining student performance. Moreover, there is also a positive correlation (0.18) between participating in review programs and board exam results. This implies that students who undergo review programs have a higher probability of passing the board exam, underscoring the efficacy of review sessions in adequately preparing students for the actual examination.

Interestingly, the General Weighted Average (GWA) exhibits a weak negative correlation (–0.15) with board exam outcomes, suggesting that students with higher academic averages may not necessarily excel in the board exam. This observation hints at a potential discrepancy between academic achievement in college and the competencies or knowledge essential for passing the board examination.

Moreover, the correlation for retakers manifests a negative trend (–0.062), indicating that students retaking the exam might encounter greater difficulty in passing compared to first-time test-takers. This underscores the necessity for tailored support or interventions specifically designed for students retaking the exam. Conversely, the correlation between being a regular student and board exam performance (0.054) is weakly positive, signifying that regular students may have a slight edge, although it is not a definitive predictor of success. Similarly, scholarship status demonstrates a very slight positive correlation (0.05) with exam results, implying that being a scholar has minimal influence on passing the board exam.

The correlation matrix provides valuable insights, with preboard exam scores and review programs emerging as key factors that positively influence board exam success. In contrast, academic performance (GWA) and retaking the exam show weaker or negative correlations, suggesting that these factors do not significantly predict exam outcomes. These insights can guide institutions in enhancing student support, particularly in preboard preparation and review programs, to improve board exam pass rates.

These findings emphasize that targeted academic preparation particularly through mock exams and structured review programs plays a crucial role in board examination success. While strong college academic performance (GWA) helps, specialized review interventions are equally, if not more, critical.

Fig. 6 presents two violin plots that illustrate the distribution of board exam results for students based on two different factors: (1) participation in review programs and (2) whether the student is a first-time taker or a retaker of the exam.

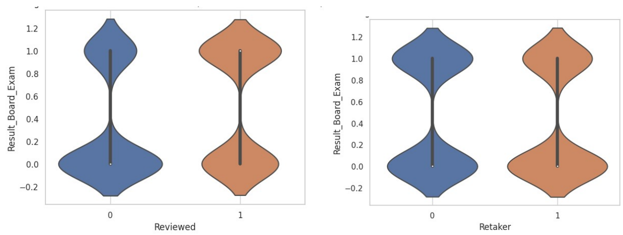

Fig. 6. Student score distribution based on review and retaking status.

Violin plots (Figs. 6 and 7) offered a deeper look into performance distributions:

Students who attended review centers consistently achieved higher scores compared to those who did not.

First-time takers outperformed retakers, confirming that exam freshness and continuity of learning enhance outcomes. Regular students (those who completed their program without interruptions) also scored higher, although the advantage was less pronounced.

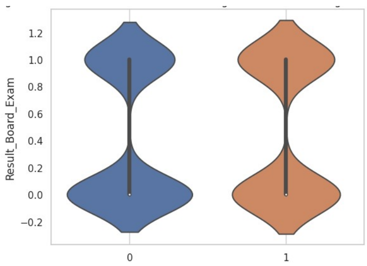

Fig. 7. Student score distribution based on regular vs irregular.

These visualizations reinforce the importance of continuous, uninterrupted education and structured preparation for board examinations.

B. Student Score Distribution: Reviewed vs. Not Reviewed (Left Plot)

This plot depicts the comparison of board exam scores between students who engaged in a review program (labeled as 1) and those who did not (labeled as 0) The violin plot illustrates the density of the score distribution, enabling a visual assessment of the central tendency and variability of the outcomes.

Reviewed Students (1): The distribution appears broader, exhibiting a concentrated range of higher scores. It implies that students who underwent the review program tend to achieve higher scores, given the higher density of scores towards the upper end of the scale.

Not Reviewed Students (0): In contrast, the distribution for students who did not review seems more dispersed, with a lower central tendency. This suggests that students who did not review generally attained lower scores than those who participated in the review. The narrower spread and density concentrated around the central point indicate that non-reviewed students are less likely to excel in the board exam.

The comparison between these two groups highlights the effectiveness of review programs in improving students' board exam performance, as reviewed students have a higher likelihood of achieving better scores.

C. Student Score Distribution: Retakers vs. First-Time Takers (Right Plot)

This plot compares the distribution of board exam results between first-time takers (marked as "0") and retakers (marked as "1"). The violin plot reveals the differences in score distribution between these two groups.

- *First-Time Takers (0):* The distribution for first-time takers is wider and more centered towards higher scores, suggesting that students taking the board exam for the first time are more likely to score higher. The density around the higher score range is more concentrated, indicating better performance overall.
- *Retakers (1):* The distribution for retakers shows a narrower spread and lower central tendency. This implies that students retaking the exam tend to score lower compared to first-time takers, with fewer students achieving higher scores. The distribution indicates that retakers generally face more difficulty in improving their performance, which is reflected in the lower scores.

Regular Students (1): The distribution is more concentrated towards higher scores, indicating that regular students tend to perform better and more consistently on the board exam.

Irregular Students (0): Their distribution is wider, with more variance, suggesting that irregular students are more likely to struggle, with a larger portion scoring lower.

Fig. 8 displays a scatter plot that analyzes the relationship between GWA (General Weighted Average) and Preboard Exam Scores, with the color coding representing board exam results. Orange dots (1) indicate students who passed the board exam, while blue dots (0) represent students who failed.

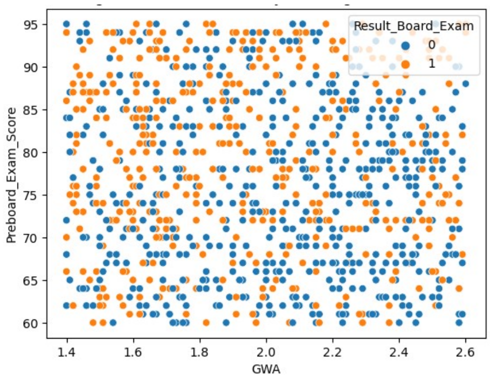

Fig. 8. Bivariate analysis (using scatter plot).

The scatter plot in Fig. 8 depicted the relationship between GWA and pre-board exam scores, colored by board exam results. Students with high pre-board scores (80 and above) were more likely to pass, regardless of slight variations in GWA.

This suggests that pre-board examination performance serves as a more immediate and practical predictor of board exam success compared to cumulative academic averages.

D. Key Observations

The scatter plot shows no strong linear relationship between GWA and Preboard Exam Scores. Students with various GWAs (ranging from 1.4 to 2.6) and Preboard Exam Scores (ranging from 60 to 95) are scattered throughout, with pass and fail outcomes intermixed across the chart.

Higher preboard exam scores (above 80) tend to have a higher concentration of students who passed (orange dots), particularly among those with better GWAs. However, both passing and failing students are present across the range of GWAs and Preboard Exam Scores.

E. Model Building and Evaluation

The data were gathered and then treated in response to the objective presented in this study. It is found out that Logistic regression classification model is the suitable algorithm in prediction of student performance in engineering licensure examinations with 66.7% accuracy and followed by Naïve Bayes Gaussian with 66.1% accuracy using the 70% percentage split.

Table 2 depicts the algorithm of the board examination performance results with the criteria of accuracy, precision, recall, and F1-Score.

Table 2. Model comparison summary

Method	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.667732	0.581967	0.572581	0.577236
Randon Forest	0.645367	0.544218	0.645161	0.590406
Decision Trees	0.648562	0.559322	0.532258	0.545455
Naïve Bayes	0.661342	0.570312	0.58871	0.579365

Accuracy measures the proportion of correctly classified instances out of the total number of instances [12]. It is calculated as the ratio of the number of true predictions (true positives and true negatives) to the total number of predictions. Accuracy provides an overall measure of how well the model predicts the correct class labels.

Precision quantifies the model's ability to correctly identify positive instances among the instances predicted as positive. It is calculated as the ratio of true positives to the sum of true positives and false positives.

Recall measures the ability of the model to identify all positive instances. It is calculated as the ratio of true positives to the sum of true positives and false negatives.

The F1-Score is the harmonic mean of precision and recall. It provides a single metric that combines both precision and recall, giving equal importance to both. The F1-Score is useful to balance precision and recall, and to focus solely on either metric.

Furthermore, Logistic Regression gains the highest score in accuracy, precision, and F1-Score, but not in recall, it generally indicates good overall performance with low false positives.

Cross-validation is a technique used in this study to assess the performance and generalization ability of the model. It involves partitioning the available dataset into multiple subsets or folds, training the model on some folds, and evaluating it on the remaining fold(s) [25, 26]. Cross-validation provides a more robust estimate of model performance by averaging the results across multiple iterations [27]. It helps in assessing how well the model is likely to perform on unseen data, selecting the best model, and tuning hyperparameters. Cross-validation is effective for overall model evaluation and selection. The study employs the stratified k-fold where n_split is 10.

Table 3, furthermore, shows the average cross-validation score of each model. Naïve Bayes outperforms the other model with 62.9% using the 10-fold cross-validation.

Table 3. 10-fold cross-validation score

Fold	Logistic Regression	Random Forest	Decision Trees	Naïve Bayes
Fold-1	0.61904762	0.59047619	0.57142857	0.59047619
Fold-2	0.63461538	0.67307692	0.64423077	0.65384615
Fold-3	0.60576923	0.50961538	0.64423077	0.625
Fold-4	0.52884615	0.5	0.54807692	0.51923077
Fold-5	0.63461538	0.63461538	0.66346154	0.65384615
Fold-6	0.57692308	0.51923077	0.50961538	0.57692308
Fold-7	0.63461538	0.625	0.64423077	0.64423077
Fold-8	0.76923077	0.69230769	0.71153846	0.76923077
Fold-9	0.55769231	0.51923077	0.51923077	0.56730769
Fold-10	0.65384615	0.58653846	0.65384615	0.69230769
CV Average	**0.621520145**	**0.58500915**	**0.61098901**	**0.62923992**

Another evaluation metric used in this study to evaluate the classification models is the area under the receiver operating characteristic curve (ROC AUC). The ROC curve plots the true positive rate (sensitivity) against the false positive rate (1—specificity) for different classification thresholds. It helps visualize the trade-off between sensitivity and specificity and allows for threshold selection based on the desired balance between true positive and false positive rates. The Area Under the ROC Curve (AUC-ROC) is often used as a summary metric, where higher values indicate better classifier performance.

Fig. 9 presents Receiver Operating Characteristic (ROC) curves for four classification algorithms: Naïve Bayes, Random Forest, Decision Tree, and Logistic Regression. The ROC curve plots the True Positive Rate (TPR) against the False Positive Rate (FPR), with the Area under the Curve (AUC) indicating the model's performance. A higher AUC suggests better classification performance, with a value of 1 representing a perfect model.

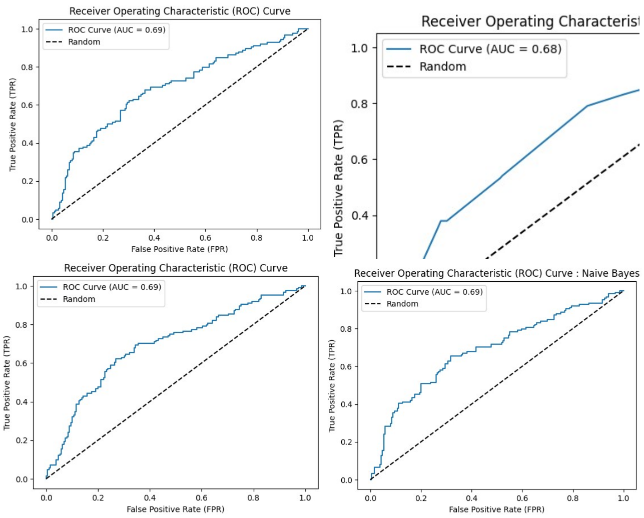

Fig. 9. ROC curves for Naïve Bayes, random forest, decision tree, and logistic regression models.

According to general machine learning benchmarks [28], an AUC above 0.7 is considered "acceptable" for classification tasks, while 0.6–0.7 falls into the "moderate" category. These thresholds, while commonly used, are somewhat arbitrary and lack a definitive scientific basis. They are often employed to provide a qualitative interpretation of model performance, transforming numerical AUC values into categories like "acceptable", "fair", "good", or "excellent".

Naïve Bayes ROC Curve: AUC: 0.69. The Naïve Bayes model performs fairly well, with an AUC of 0.69, indicating that the model has a moderate ability to distinguish between students who pass and fail. The curve rises above the diagonal line, suggesting that the model performs better than random guessing.

Random Forest ROC Curve: AUC: 0.68. The Random Forest model demonstrates similar performance to Naïve Bayes, with an AUC of 0.68. The curve shows a gradual rise in the true positive rate as the false positive rate increases, reflecting a moderate classification capability.

Decision Tree ROC Curve: AUC: 0.68. The Decision Tree model also has an AUC of 0.68, similar to the Random Forest model. Its performance is comparable to that of the Naïve Bayes and Random Forest models, indicating moderate classification performance.

Logistic Regression ROC Curve: AUC: 0.68. The

Logistic Regression model produces an AUC of 0.68, demonstrating performance that is consistent with the other models. The ROC curve rises steadily, indicating that the model moderately differentiates between passing and failing students.

All four models Naïve Bayes, Random Forest, Decision Tree, and Logistic Regression perform similarly, with AUC values ranging from 0.68 to 0.69. These AUC values indicate that the models provide moderate predictive power, but none achieve high classification accuracy. The results suggest that while the models can distinguish between students who pass and fail the board exam, further improvements or the use of additional features may be necessary to enhance their performance.

Table 4 ranks the importance of various features used in predicting student board exam results, along with their respective importance scores. GWA and Preboard Exam Scores are the dominant factors in predicting board exam results, while other factors like review participation, retaking status, scholarship, and regular student status contribute less to the prediction model [29, 30]. This indicates the primary importance of academic and preparatory performance in determining board exam success.

Table 4. Feature importance ranking

Features	Score
GWA	0.458417
Preboard Exam Score	0.384829
Reviewed	0.045242
Retaker	0.040101
Scholar	0.036295
Regular Student	0.035115

Fig. 10 plot highlights that GWA and Preboard Exam Scores are the dominant factors in predicting board exam results, with the remaining features playing relatively minor roles. These insights suggest that academic performance and preboard preparation are critical in determining board exam success, while other factors like reviewing and student status have less predictive power.

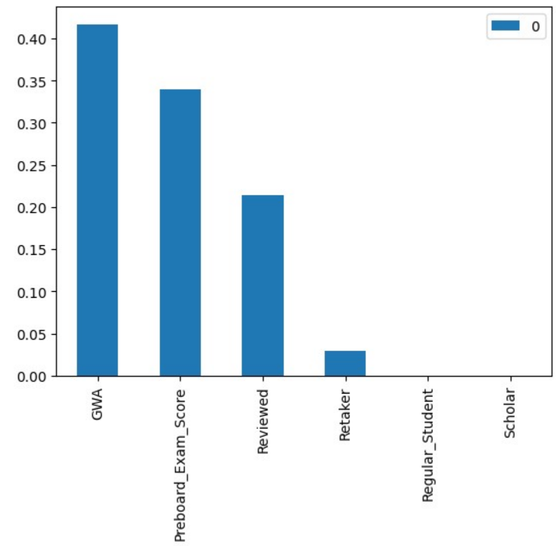

Fig. 10. Plot on feature importance.

Compared to Issah *et al.* [31], who achieved around 65% accuracy in board exam prediction using classification models, this study's models slightly outperformed previous efforts, reaching up to 66.7% accuracy [32]. Similarly, Tharwat [33] reported moderate AUC values (0.65–0.7) when predicting student success using feature-selected models. These parallels reinforce the credibility and feasibility of this study's approach while suggesting room for further optimization [28, 29].

IV. Conclusion

This study aimed to develop a predictive model for forecasting engineering students' board examination performance using machine learning classification algorithms. The results demonstrated that logistic regression achieved the highest predictive accuracy at 66.7%, closely followed by Naïve Bayes at 66.1%, indicating their effectiveness for binary classification tasks in academic performance prediction. Among the evaluated input variables, General Weighted Average (GWA) and pre-board examination scores emerged as the most influential predictors, as confirmed by both correlation analysis and feature importance rankings. These findings highlight the critical role of sustained academic performance and preparatory assessments in determining licensure examination outcomes.

The main contribution of this research lies in its demonstration that relatively simple classification models, when applied to well-preprocessed educational data, can offer actionable insights into student readiness for high-stakes exams. This supports educational institutions in early identification of at-risk students and targeted intervention planning.

Practically, the study recommends that the College of Engineering leverage predictive analytics to strengthen academic advising and review strategies. Special attention should be given to first-time takers and students with low pre-board scores, as these groups benefit the most from focused support. Moreover, the effectiveness of external review centers, as observed in this study, suggests the need to enhance and promote the institution's internal review programs to improve their appeal and efficacy.

For future research, it is recommended to expand the dataset to include behavioral and engagement variables, such as time spent in review sessions, class attendance, study habits, and psychological readiness, which may further improve prediction accuracy. Longitudinal tracking and the inclusion of qualitative feedback from students could also enrich the model's robustness. These enhancements will support the development of more holistic and accurate student performance prediction systems that align closely with real-world academic challenges.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Jayson A. Batoon conceptualized the study, developed the predictive models, and carried out data analysis and evaluation. Sarah Jana L. Cabral contributed to the research methodology, literature review, and manuscript editing. Both authors reviewed and approved the final version of the manuscript.

ACKNOWLEDGMENT

The authors wish to thank the College of Information and Communications Technology, Bulacan State University, Malolos, Philippines, and the Information Technology Department, College of Engineering, Eastern Visayas State University, for their support and collaboration throughout the research process. Their assistance in providing resources and institutional backing was invaluable to the completion of this study.

REFERENCES

[1] B. Minaei-Bidgoli, D. A. Kashy, G. Kortemeyer, and W. F. Punch, "Predicting student performance: An application of data mining methods with an educational web-based system," in *Proc. 33rd Annu. IEEE Front. Educ.*, vol. 1, Nov. 2003.

[2] K. Priya and A. Kumar, "Improving the student's performance using educational data mining," *Int. J. Adv. Netw. Appl.*, vol. 4, no. 4, pp. 1680–1685, 2013.

[3] C. Romero and S. Ventura, "Educational data mining: A survey from 1995 to 2005," *Expert Syst. Appl.*, vol. 33, no. 1, pp. 135–146, 2007. doi: 10.1016/j.eswa.2006.04.005

[4] R. A. Rustia *et al.*, "Predicting student's board examination performance using classification algorithms," in *Proc. ACM Int. Conf. Ser.*, 2018, pp. 233–237. doi: 10.1145/3185089.3185101

[5] B. Albreiki, D. Zeglat, and R. Alrawashdeh, "Digital transformation and its impact on firm performance: A systematic literature review," *International Journal of Innovation, Creativity and Change*, vol. 13, no. 12, pp. 1–14, 2020.

[6] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, "Same, same but different!? A systematic review of the literature on digital transformation," *Educational Technology & Society*, vol. 23, no. 1, pp. 1–14, 2020.

[7] S. Muchuchuti, L. Narasimhan, and F. Sidume, "Classification model for student performance amelioration," *Lect. Notes Netw. Syst.*, vol. 68, 2019.

[8] S. K. Yadav and S. Pal, "Data mining: A prediction for performance improvement of engineering students using classification," *World Comput. Sci. Inf. Technol. J.*, vol. 2, no. 2, pp. 51–56, 2012.

[9] A. Arif and A. Jahan, "An improved prediction system of students' performance using classification model and feature selection algorithm," *Int. J. Adv. Soft Comput. Appl.*, vol. 12, no. 1, pp. 162–177, 2021.

[10] F. M. Cagle Jr., "High-stakes testing: A study of the relevance and its contribution to career and college readiness," Ph.D. dissertation, Gardner-Webb Univ., Boiling Springs, NC, USA, 2023.

[11] M. P. Mohammed and M. P. Mohammed, "Licensure examination performance evaluation of the candidate engineers as basis for a proposed action plan," *Asia Pac. J. Multidiscip. Res.*, vol. 5, no. 2, pp. 51–57, 2017.

[12] M. Adnan *et al.*, "Predicting at-risk students at different percentages of course length for early intervention using machine learning models," *IEEE Access*, vol. 9, pp. 7519–7539, 2021.

[13] A. N. Prasad, "Data quality and preprocessing," *Introduction to Data Governance for Machine Learning Systems*, Berkeley, CA, USA: Apress, 2024, pp. 109–223.

[14] S. Batool *et al*., "Educational data mining to predict students' academic performance: A survey study," *Educ. Inf. Technol.*, vol. 28, no. 1, pp. 905–971, 2023.

[15] Y. A. Alsariera *et al*., "Assessment and evaluation of different machine learning algorithms for predicting student performance," *Comput. Intell. Neurosci.*, vol. 2022, 4151487, 2022.

[16] K. L. Maslang. (2023). Predictors of academic performance among pre-service teachers: The case of a private higher educational institution in Northern Philippines. *SSRN*. [Online]. Available: https://ssrn.com/abstract=4325530

[17] A. B. A. Lopez, "Perceptions of students to open and distance learning of review classes in a review center in the Philippines," *Int. J. Open Distance e-Learn.*, vol. 5, no. 1, 2019.

[18] R. Ahmed *et al*., "Impact of scholarships on student success," *Pak. Dev. Rev.*, vol. 61, no. 2, pp. 231–258, 2022.

[19] K. J. Flores, "Academic performance, pre-board examination results and dentistry licensure examination performance of a dental school in Laguna," in *Proc. 11th Int. Scholars Conf.*, vol. 11, no. 1, 2024, pp. 415–425.

[20] E. A. C. Balanga *et al*., "Perceived psychological preparedness of the new tertiary curriculum graduates in taking the professional board examination," *J. Educ. Res. Innov.*, vol. 12, no. 3, 2023.

[21] R. Kholwal, "Text-classify: A comprehensive comparative study of logistic regression, random forest, and KNN models for enhanced text classification performance," *Int. J. Adv. Eng. Technol.*, vol. 16, no. 5, pp. 415–433, 2023.

[22] Priyanka and D. Kumar, "Decision tree classifier: A detailed survey," *Int. J. Inf. Decis. Sci.*, vol. 12, no. 3, pp. 246–269, 2020.

[23] A. Zaidi and A. S. M. Al Luhayb, "Two statistical approaches to justify the use of the logistic function in binary logistic regression," *Math. Probl. Eng.*, vol. 2023, 5525675, 2023.

[24] H. A. Salman, A. Kalakech, and A. Steiti, "Random forest algorithm overview," *Babylonian J. Mach. Learn.*, vol. 2024, pp. 69–79, 2024.

[25] V. Nakhipova *et al*., "Use of the naive Bayes classifier algorithm in machine learning for student performance prediction," *Int. J. Inf. Educ. Technol.*, vol. 14, no. 1, pp. 92–98, 2024.

[26] J. A. Batoon *et al*., "THESISQUO: Research management system with plagiarism checker and prescriptive analytics," *Int. J. Comput. Sci. Issues*, vol. 11, no. 3, 2022.

[27] N. White, R. Parsons, G. Collins *et al*., "Evidence of questionable research practices in clinical prediction models," *BMC Med*, vol. 21, 339, 2023. https://doi.org/10.1186/s12916-023-03048-6

[28] R. T. Camañero, A. M. Berico, and D. A. Tan, "Predictors of licensure examination performance of bachelor's in education graduates of a community college in Bukidnon, Philippines," *Int. J. Res. Innov. Soc. Sci.*, vol. 8, no. 10, pp. 821–831, 2024. doi: 10.47772/IJRISS.2024.8100067.

[29] V. W. Lumumba *et al*., "Comparative analysis of cross-validation techniques: LOOCV, K-folds cross-validation, and repeated K-folds cross-validation in machine learning models," *Am. J. Theor. Appl. Stat.*, vol. 13, pp. 127–137, 2024.

[30] A. Vahdati *et al.*, "Enhancing data integrity in Electronic Health Records: Review of methods for handling missing data," *medRxiv*, 2024. doi: 10.1101/2024-05

[31] I. Issah *et al.*, "A systematic review of the literature on machine learning application of determining the attributes influencing academic performance," *Decis. Anal. J.*, vol. 7, 100204, 2023.

[32] S. R. Page, "Predictive factors of next generation National Council Licensure Examination success in a baccalaureate nursing program," Ph.D. dissertation, Brenau Univ., Gainesville, GA, USA, 2023.

[33] A. Tharwat, "Classification assessment methods," *Appl. Comput. Inform.*, vol. 17, no. 1, pp. 168–192, 2021.