International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V73I7P133
Volume 73 Issue 7, 443-457, July 2025

Original Article

Development of a Web-Based Student Academic Performance Prediction Using Machine Learning for Higher Education Institutions

Michael Marvin P. Cruz[1], Ramil G. Lumauag[2]

[1]*Graduate School, State University of Northern Negros, Sagay City, Negros Occidental, Philippines.*
[2]*Computer Department, Iloilo Science and Technology University, La Paz, Iloilo City, Philippines.*

[1]*Corresponding Author : michaelmarvincruz@cpsu.edu.ph*

Received: 08 April 2025 Revised: 01 July 2025 Accepted: 04 July 2025 Published: 30 July 2025

Abstract - *The rapid rise of predictive analytics and machine learning is changing higher education institutions' student academic performance monitoring. This study introduces HEAPS, a practical and deployable machine learning approach designed to predict at-risk students. This research adopts a comprehensive approach by developing, evaluating, and integrating the selected model into a real-time web application, in contrast to most studies that focus exclusively on identifying the best-performing model. A comparison was conducted between a singular model, Enhanced Random Forest (ERF), and a stacking ensemble model that integrates Random Forest and XGBoost as both base and meta-classifiers. Although the ensemble model exhibited superior prediction accuracy across key evaluation metrics, the ERF model achieved a significantly faster training and inference time. This difference emphasizes a critical trade-off between model complexity and computational efficiency. The results provide practical guidance for researchers and practitioners in model selection, emphasizing accuracy and real-world applicability, thereby ensuring that implemented systems are both efficient and responsive. Usability evaluation showed that HEAPS is an accessible and effective tool for academic intervention, connecting algorithmic research with educational implementation.*

Keywords - *Academic intervention, Early prediction system, Higher education, Machine Learning, Student academic performance.*

1. Introduction

Education is a key component of both individual and societal development, impacting everything from economic prosperity to personal development [1]. Individual achievement is significantly impacted by a person's educational background and capacity to apply what they have learned in the classroom [2]. Higher education, in particular, equips students with the knowledge and skills needed to excel in their chosen fields, making it a critical stage in academic and professional advancement. However, success in higher education is not guaranteed, as many students face academic challenges that may lead to poor performance, course shifting, or even dropout. Recognizing this, many institutions implement student success intervention programs, primarily targeting first-year students [3-5], as they are more vulnerable to academic struggles and transition difficulties. [6] contends that for a number of reasons, first-year students frequently find it difficult to transition from high school to a higher education setting. The majority of Higher Education Institutions (HEIs) use academic interventions as essential teaching and learning strategies [7]. Identifying struggling students early and providing timely intervention is essential to ensuring student success and institutional retention [8]. This strategy is crucial in preventing academic failure, allowing institutions to offer personalized support through academic advising, mentoring, and skill development programs before students reach a critical point in their studies. Without proactive intervention strategies, at-risk students may continue to underperform, leading to long-term consequences such as lower retention rates and higher dropout rates.

Globally, universities face challenges in retaining students [9]. This issue has become increasingly alarming in the Philippines, based on recent data from the Commission on Higher Education (CHED). Attrition rates in colleges and universities were recorded at 15.90% during the school Year 2020–2021, but sharply increased to 37.79% in 2021–2022, then peaked at 40.98% in 2022–2023 before declining to 29.4% in 2024. This significant fluctuation highlights the growing concern over students' ability to sustain their academic journey. In a 2024 press conference, CHED emphasized that while more students are entering higher education, many are unable to finish due to various barriers. Among these, academic difficulty ranks as the 5th most common reason why students discontinue their studies. A 2024 article from Rappler highlights a similar concern, noting that 4 in 10 Filipino college students discontinue their studies despite the enactment of the Free Tuition Law. The Second Congressional Commission on Education (EDCOM 2) raised concerns regarding the national higher education attrition rate, highlighting that higher dropout rates continue despite the government's implementation of free tuition in State Universities and Colleges (SUCs). The data indicate that financial support alone is not sufficient to ensure student retention and success.

Retention policies in higher education, particularly in board exam courses, necessitate that students maintain a minimum General Weighted Average (GWA) to remain in their program. Students who fall below the threshold are recommended to transition to an alternative program due to academic difficulties. This approach guarantees that only academically prepared students continue in rigorous programs, highlighting the necessity for early identification and support to assist students in improving prior to the transition point. Course shifting impacts program-specific retention rates, distinguishing it from student retention, which emphasizes maintaining overall enrollment in the institution without resulting in complete dropout. This is crucial for deans and academic administrators responsible for overseeing student success and ensuring that students remain in their selected programs rather than being compelled to transfer or exit the university altogether. CHED encourages SUCs to adopt intervention strategies to decrease student dropout rates. In the absence of an effective system to track student performance and deliver timely interventions, institutions face the risk of attrition from high-demand programs. This situation adversely impacts graduation rates and undermines the objectives of equitable and accessible higher education, ultimately affecting overall institutional performance. The absence of a proactive system hinders early intervention, resulting in numerous students, particularly freshmen, lacking the necessary support to enhance their performance before it becomes critical.

The growing volume of student data from school records has led to the emergence of Educational Data Mining (EDM) and Machine Learning (ML) as effective solutions to address this challenge. In recent years, these techniques have emerged as powerful tools for analyzing student data and predicting academic performance [10, 11]. While existing studies [12, 13] have applied these techniques to student performance prediction, many are limited in scope. Most studies either focus only on algorithm comparison without practical deployment or lack system-level implementation that integrates predictive models into real-world educational tools. This gap highlights the need for a complete, end-to-end solution that not only evaluates model performance but also translates these models into a usable, accessible platform for stakeholders in educational settings.

This study aims to fill gaps by developing a web-based prediction tool, Higher Education Academic Prediction System (HEAPS). This tool incorporates machine learning outcomes into a practical, user-centered application designed to predict students' academic performance and identify students at risk. By analyzing current and historical academic data, particularly first-year students' GWA, the system will enable early intervention and data-driven decision-making. The novelty of this study lies in its comprehensive lifecycle approach: from comparative experimentation with hybrid algorithms to real-time deployment in a web interface. Unlike previous works, the study does not treat algorithm comparison as the end goal, but rather as a means to develop a web prediction system that is both computationally efficient and highly usable. Aligned with the university's retention policy, the developed system will enable educators to monitor student progress and implement necessary support strategies. The goal is to enhance academic success and student retention through an automated, technology-driven solution that assists both students and administrators in maintaining academic standards.

2. Related Works

2.1. Machine Learning in Education

ML is increasingly recognized for its transformative potential in education, enhancing personalized learning experiences and improving educational outcomes. ML is significantly impacting the education sector, which believes it has the potential to enhance aspects of education and learning that are currently tiresome and difficult to manage [14]. In today's trends, big data and learning analytics have taken center stage in the field. Despite machine learning still being in its early stages, its ability to analyze and interpret data is well recognized [15]. Machine learning-based predictive analytics enables educational institutions to analyze extensive student data. Analytics-derived information assists HEIS administrators in understanding the diverse aspects influencing decision-making. This process is facilitated by machine learning, which provides a range of methods and methodologies suitable for different data kinds and prediction types needed [15]. Machine learning techniques are becoming more common in education, aiding in university admissions, predicting student attrition, and supporting the growth of Massive Open Online Courses (MOOCs) [16]. This suggests that machine learning has significant potential in higher education, enabling the automation of decision-making processes and the analysis of relevant data.

2.2. Predicting Student Academic Performance

The growing availability of student data in HEIs has facilitated the adoption of data-driven approaches to improve academic performance. Traditional methods of evaluating student performance often rely on manual assessments and historical analysis, limiting the ability to provide timely interventions. This allows educational institutions to utilize predictive analytics for identifying at-risk students and evaluating academic success. Several studies have explored the application of machine learning techniques in predicting student performance. The majority of studies on applying machine learning techniques for precision education concentrated on predicting learning performance or dropouts in a divergent source of data [17].

A literature review of [13] also revealed that the predominant technique for predicting student behavior is supervised learning, due to its accuracy and reliability. Researchers have utilized models such as Random Forest, Naïve Bayes, Logistic Regression, Support Vector Machines (SVM), and Decision Trees to analyze student-related factors like academic achievement grades, attendance, study habits, and socioeconomic background [18, 19]. Supervised learning through Random Forest demonstrates potential in predicting students' academic performance. In the [20] study, Decision Tree, Random Forest, and Naïve-Bayes classifiers are employed to evaluate the performance of their predictive model based on accuracy metrics. Among the three algorithms, Random Forest achieved the highest accuracy, resulting in an accuracy rate of 81%.

A paper from [21] reveals that their findings illustrate the effective identification of over 80% of students at low-performance risk using the Random Forest model at the end of the second semester, in contrast to Decision Tree (DT) and Gradient Boosting Decision Tree (GBDT). Random Forest also shows its predictive capabilities in the study of [22], where their experimental results for both SVM and RF algorithms applied to both datasets demonstrated that the accuracy for binary classification attained a remarkable prediction rate of 93% in forecasting student academic performance, thereby facilitating improved academic outcomes for students. [23] employed Random Forest, Voting, Gradient Boosting, XGBoosting algorithm, SVM, and CART to predict students' academic achievement, where Random Forest attained the highest overall accuracy, with a result of 92%, an F1-score of 91%, and superior recall and precision compared to other models. A comparative analysis of these techniques has shown that models, notably Random Forest, tend to yield higher accuracy, especially in classification.

2.3. Early Prediction Systems in Education

Early prediction systems play a crucial role in identifying students at risk before they face academic failure. Several research studies have focused on developing predictive models that use student demographics, engagement levels, and academic records to forecast future performance. According to [24], early prediction systems in education can prevent student dropouts by using the predictive model to identify at-risk students early in the course and intervene promptly to keep them on track.

Additionally, studies have highlighted the importance of incorporating historical academic data from high school into prediction models to improve accuracy. Research by [25] shows that students' prior secondary school GPA was considered a stronger predictor of retention for both university education and higher professional education. Predicting a student's performance based on their prior grades is a prominent use of educational data mining, therefore serving as a significant source of information utilized for many purposes [26, 27].

The paper by [28] indicates a correlation between students' socio-economic background and their entrance examination scores in predicting CGPA. By integrating students' historical data, socio-economic background, entrance examination results, and other academic indicators such as CGPA, early prediction systems can enhance the precision of risk assessment models and provide valuable insights for educators and policymakers.

Even though several studies have proposed early prediction systems using popular machine learning algorithms for higher education, there remains a critical need for more advanced and comprehensive tools that can effectively identify at-risk students. These tools should facilitate timely interventions, enabling institutions to provide targeted support and improve student outcomes. To bridge this gap, this study proposes the development of the HEAPS. HEAPS aims to predict student performance in higher education institutions by integrating historical academic data and real-time academic engagement metrics. By leveraging machine learning techniques, this system aims to enhance institutional efforts in the early identification of at-risk students and provide data-driven recommendations for academic support.

3. Materials and Methods

3.1. Research Design

This study follows an experimental-developmental-quantitative research design, integrating machine learning techniques into a web-based application to predict student academic performance. Training, testing, and evaluating different machine learning models are involved in the experimental process to determine their effectiveness in predicting student success. It is developmental, focusing on designing and implementing a web-based system that enables educators and administrators to make data-driven decisions. Additionally, it is quantitative since the study depends on numerical data, statistical analysis, and machine learning assessment metrics to assess model performance.

3.2. Proposed Software Development Model

The proposed development model of HEAPS, illustrated in Figure 1, follows a systematic and iterative methodology that integrates the concepts and phases of the CRISP-DM framework [29] and the principles and processes of the Evolutionary Prototyping Model [30].

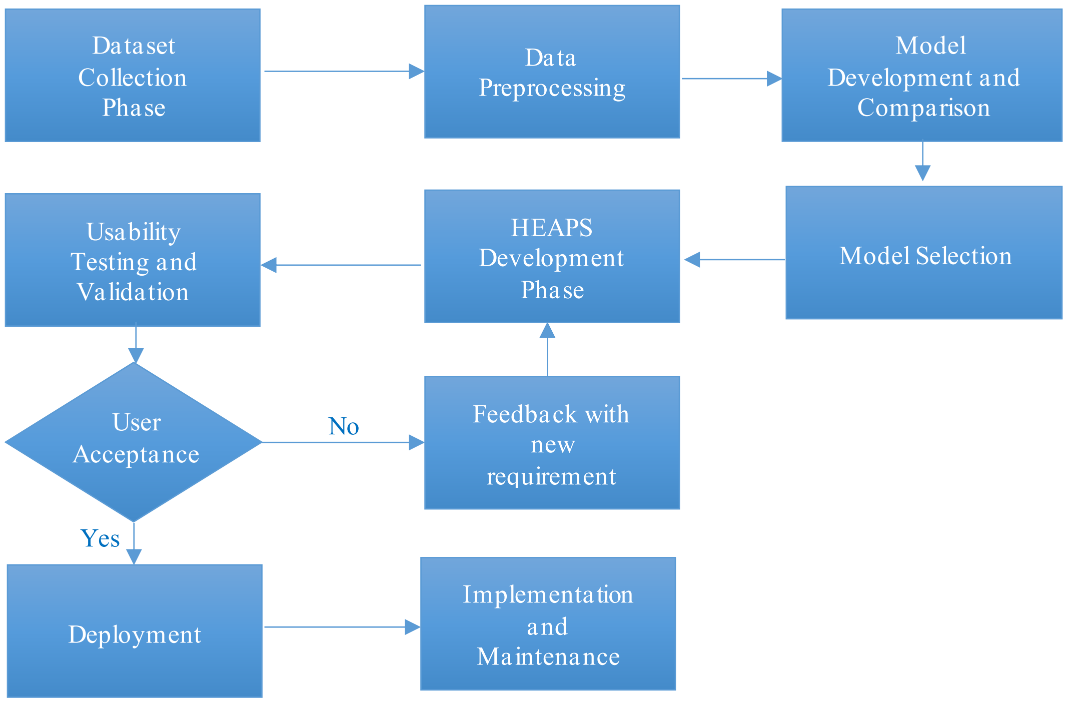


Fig. 1 Proposed software development model of HEAPS

The CRISP-DM framework provided a systematic methodology for managing data-centric activities, including data collection, preprocessing, modeling, and evaluation, ensuring that the data mining process remained organized and goal-oriented. Using evolutionary prototyping helps enable adaptable, user-centric web application development. Developed gradually, the system was improved under constant user feedback instead of adhering strictly to a linear approach. With usability and performance assessed at every stage, this iterative process allowed the application to go through many iterations. When initial prototypes did not meet user expectations, the gathered feedback affected design improvements, thereby promoting continuous development. This development cycle ensured that the final product met user expectations as well as technical requirements. By combining evolutionary prototyping with CRISP-DM, a development approach that was both flexible and data-driven was established, therefore enabling the effective application of HEAPS as a consistent academic risk prediction tool in HEIs.

3.3. Data Collection and Sources

The results of this study were gathered using a validated survey questionnaire, carefully designed based on the knowledge from the literature review, input from faculty members at educational institutions, and contributions from researchers in the field of education. The questionnaire was evaluated for relevance and validity by five experts, including a statistician and the Dean of Instruction at the institution, both of whom had significant knowledge in data collection and academic research techniques. The survey was conducted at the Central Philippines State University (CPSU) campuses in Hinigaran and San Carlos, targeting freshmen students from all courses. This study assessed freshmen due to their transitional phase from high school to college, during which academic adjustment greatly influences their performance. Identifying at-risk students at this stage facilitates early intervention, which can significantly enhance retention rates and overall academic performance. A total of 690 students participated in the survey, which aimed to collect data regarding their initial semester in college. GWAs were obtained from each campus's respective Office of the Registrar to supplement the survey responses. A formal letter was sent to key university officials, including the university's president, requesting authorization for data collection through surveys. The survey questionnaire was conducted using Google Forms to facilitate distribution and data management.

However, a hard copy version was also provided for students who had limited internet access to ensure inclusivity. To facilitate better understanding and accurate responses, the questionnaire was administered room by room, with the researchers explaining each question before students proceeded with answering. The questionnaire consisted of 18 questions, structured into five sections: Demographic Information, Pre-University Academic Background, Study Habits & Academic Behavior, Financial and Personal Factors, and College Academic Performance. The dataset is anonymized to comply with data privacy regulations and ethical considerations in educational research.

3.4. Data Preprocessing

[31] highlighted various challenges in applying machine learning, notably the susceptibility to errors when models are trained on inadequate or low-quality data. Limited or

imbalanced datasets can result in biased and overly generalized predictions. A thorough and well-executed data preprocessing is essential to ensure the data is clean, consistent, and properly structured to support optimal model performance.

Raw data often contains missing values, inconsistencies, and redundant information that can negatively impact predictive accuracy. By preprocessing the dataset, data quality will be enhanced, biases will be reduced, and the reliability of model predictions will be improved. For this study, each phase was conducted using the Python programming language. The dataset initially consisted of 690 student records. However, 41 entries with missing or incomplete data were removed to ensure data integrity, reducing the dataset to 649 complete entries.

This step was necessary, as missing values could introduce biases or inaccuracies in model predictions. Then, data transformation was performed on several attributes to ensure uniformity and compatibility with machine learning models. Each categorical attribute was encoded into numerical values for better model interpretability. Table 1 presents a summary of the transformations applied to the dataset.

Table 1. Features of the dataset

Feature	Categories	Numerical Encoding	Variable Type
Sex	Male, Female	Male = 0, Female = 1	Nominal (Binary)
Scholarship	Yes, No	No = 0, Yes = 1	Nominal (Binary)
Type of High School	Public, Private	Public = 0, Private = 1	Nominal (Binary)
SHS GWA	90 and above (With Honor), 89 and below (Average)	With Honor = 0, Average = 1	Ordinal
Entrance Exam Result	1st Qualifier, 2nd Qualifier	1st Qualifier = 1, 2nd Qualifier = 0	Ordinal
Study Hours	Less than 1 hour, 2-3 hours, 4-5 hours, More than 5 hours	<1hr = 0, 2-3hrs = 1, 4-5hrs = 2, >5hrs = 3	Ordinal
Submission Activities	Never, Rarely, Sometimes, Always	Never = 0, Rarely = 1, Sometimes = 2, Always = 3	Ordinal
Consultation	Never, Rarely, Sometimes, Always	Never = 0, Rarely = 1, Sometimes = 2, Always = 3	Ordinal
Attendance	Below 30 days (<50%), 30-40 days (50-69%), 45-59 days (70-89%), 60-70 days (90-100%)	<50% = 0, 50-69% = 1, 70-89% = 2, 90-100% = 3	Ordinal
Part-Time Job	Yes, No	No = 0, Yes = 1	Nominal (Binary)
Devices	Yes, No	No = 0, Yes = 1	Nominal (Binary)
Internet Access	Yes, No	No = 0, Yes = 1	Nominal (Binary)
Daily Allowance	Below Php 50, Php 50-100, Php 101-200, More than Php 200	<50 = 0, 50-100 = 1, 101-200 = 2, >200 = 3	Ordinal

The GWA was binned into two categories based on the university's retention policy:

- Low risk (0) – Students with a GWA between 1.1 and 2.5 are considered academically stable.
- At risk (1) – Students with a GWA of 2.6 and above, indicating the need for academic intervention.

This binning process aligns with university policies for monitoring student performance and allows the model to predict students at risk of academic failure effectively. Through these preprocessing steps, the dataset was refined and structured for machine learning applications, ensuring consistency and improved model performance.

3.5. Machine Learning Models

This research examines the subsequent models:

3.5.1. Random Forest

Random Forest (RF) has consistently emerged as one of the top-performing algorithms for academic prediction tasks. Scholarly studies [32, 33] have shown that RF consistently delivers strong and reliable results in similar educational contexts, further supporting its inclusion in this research and identifying it as a potential candidate for deployment in the HEAPS web application.

This study employs the Synthetic Minority Over-sampling Technique (SMOTE) to address the significant issue of class imbalance, characterized by the underrepresentation of "at-risk" students. SMOTE creates synthetic instances of the minority class to improve model sensitivity and mitigate bias towards the majority class. Research of [34, 35] validates the effectiveness of combining SMOTE with RF for improved classification performance in imbalanced settings.

The Enhanced Random Forest (ERF) in this study refers to Random Forest applied with SMOTE for balancing, 10-fold cross-validation for model validation, and a 70-30 train-test split for consistent evaluation.

3.5.2. Hybrid Random Forest + XGBoost (RF+XGB)

This study implements a stacked ensemble model that combines Random Forest (RF) and Extreme Gradient Boosting (XGB) to investigate the potential performance enhancement of a hybrid ensemble approach. XGB was selected as the model to be combined with Random Forest because of its established scalability and enhanced predictive accuracy, particularly regarding student academic performance [12].

The combination of RF and XGB seeks to utilize the advantages inherent in both models. This hybrid approach aims to utilize the advantages of both algorithms, as stacking ensemble methods frequently exceed the performance of individual learners [36].

Like the ERF model, the hybrid model incorporates SMOTE, employs 10-fold cross-validation, and utilizes a 70-30 data split, facilitating a consistent and equitable comparison across all evaluation metrics. This study compares the two models to identify the most effective and generalizable option for deployment in the HEAPS web application.

3.6. Model Evaluation and Validation

To identify the most suitable model for integration with the web application, several evaluation metrics are used:

Accuracy: Quantifies the ratio of accurately classified instances to the overall number of instances. It is calculated as:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

Precision: Calculates the true positive prediction ratio relative to the model's overall positive predictions. It is given by:

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

Recall: Evaluates the model's ability to identify all real positive instances accurately. It is expressed as:

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

F1-Score: Offers a fair balance between recall and accuracy, especially beneficial for unbalanced data.. It is computed as:

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

ROC-AUC: Evaluates the model's ability to distinguish between classes at different threshold values. It is computed as:

$$F1\ AUC = \int_0^1 TPR\ (FPR)\, dFPR \tag{5}$$

Training and Inference Time: Measures the duration necessary for model training and the interval required for making predictions. This metric is critical for determining the feasibility of deploying the model in a real-time web-based system. The Python time library is used to track execution time during training and inference.

4. Experimental Results and Discussion

This section presents the performance evaluation of the predictive models, including ERF and the hybrid model with different meta-models: Hybrid RF+XGB (XGB as meta-classifier) and Hybrid RF+XGB (RF as meta- classifier), which comprehensively discusses the obtained results, justifying the trade-offs among the models and selecting the most suitable approach for integration into the web application. Table 2 presents the comparative evaluation of the models based on the considered evaluation metrics:

4.1. Accuracy and F1-Score

Table 2. Accuracy and F1-score comparison of all models

Model	Accuracy	F1-Score
ERF	95.38	96.12
Hybrid RF+XGB (XGB)	96.92	96.55
Hybrid RF+XGB (RF)	96.92	96.92

Among the tested models based on Table 2, Hybrid RF + XGB (RF) and Hybrid RF + XGB (XGB) achieved the highest accuracy of 96.92%. The Hybrid RF + XGB (RF) also attained the highest F1-Score of 96.92%, indicating a strong balance between precision and recall. ERF, on the other hand, achieved 95.38% accuracy and an F1-score of 96.12%, which is slightly lower than the stacking models. Although the difference in accuracy is marginal, the stacking models demonstrated better overall classification consistency.

4.2. Precision and Recall

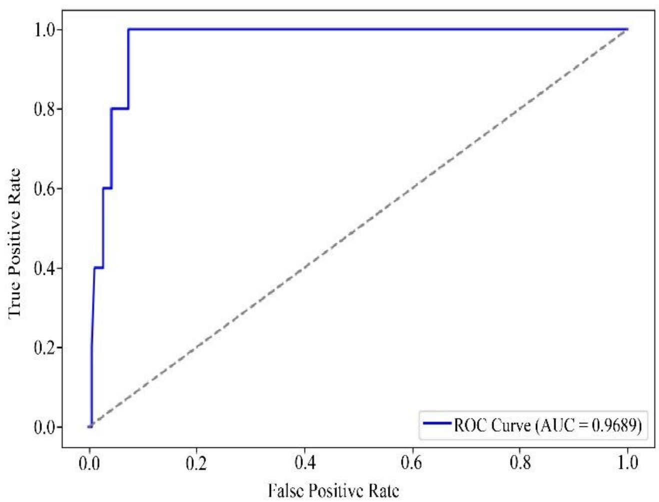
Precision measures the proportion of true positive predictions among all positive predictions. High precision across models indicates that they effectively minimize false positives (students identified as needing intervention but who do not).

Table 3. Precision and recall scores of all models

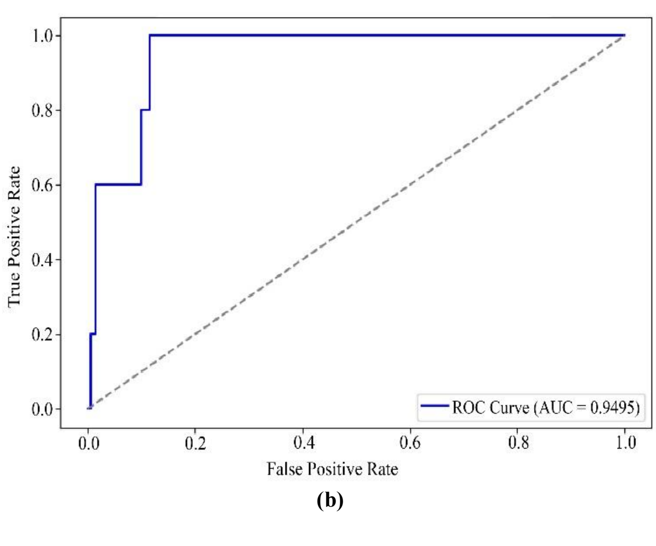
Model	Precision	Recall
ERF	97.15	95.38
Hybrid RF+XGB (XGB)	96.26	96.92
Hybrid RF+XGB (RF)	96.92	96.92

The precision and recall metrics shown in Table 3 further provide insight into model effectiveness. The Hybrid RF+XGB (RF) demonstrated a balanced precision and recall of 96.92%, which implies an optimized balance between correctly identifying at-risk students and minimizing false positives. ERF exhibited the highest precision (97.15%), making it preferable in scenarios where minimizing false positives is crucial. However, it had a slightly lower recall (95.38%), which indicates that while it is more precise, it might miss some at-risk students. The Hybrid RF+XGB (XGB) showed the lowest precision (96.26%) but had a slightly higher recall (96.92%), meaning it prioritized identifying more at-risk students at the cost of more false positives.

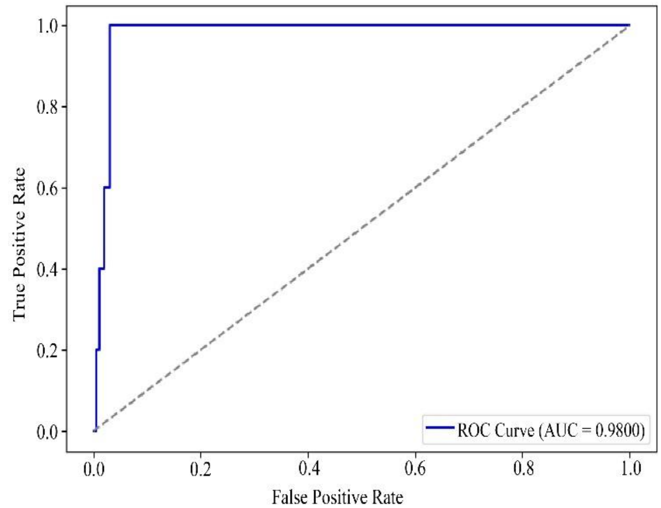
4.3. ROC-AUC Score



(a)



(b)



(c)

Fig. 2 ROC Curve of (a) ERF, (b) Hybrid RF+ XGB (XGB), and (c) Hybrid RF+ XGB (RF).

Figure 2 reflects the model's capability to distinguish between at-risk and low-risk students. The Hybrid RF + XGB (RF) model, as shown in Figure 2(c), achieved the highest ROC-AUC score of 98.00%, reflecting an exceptional ability to differentiate between students who are at risk and those who are not.

This suggests that the ensemble structure effectively leverages the complementary strengths of both RF and XGB during base-level learning, while the RF meta-learner consolidates the outputs into a highly discriminative final prediction.

The nearly optimal AUC indicates minimal overlap in predicted probabilities between the two classes, leading to a reduction in false positives and false negatives.

Figure 2(a) demonstrates that the ERF achieved a score of 96.89%, reflecting robust and consistent efficacy in differentiating student risk categories. Although marginally inferior to the hybrid model, ERF continues to deliver highly reliable classification outcomes. The method's strength is attributed to its simplicity and computational efficiency, rendering it practical for deployment while maintaining high accuracy.

On the other hand, the Hybrid RF + XGB (XGB) model shown in Figure 2(b) recorded the lowest ROC-AUC score of 94.95% among the three. While this is still a commendable score, it indicates a relatively lower discriminative capacity than the other models.

One possible reason is that XGB, as a meta-classifier, may have introduced overfitting or failed to generalize the ensemble predictions optimally, especially in the presence of synthetic instances from SMOTE.

4.4. Model Performance Comparison

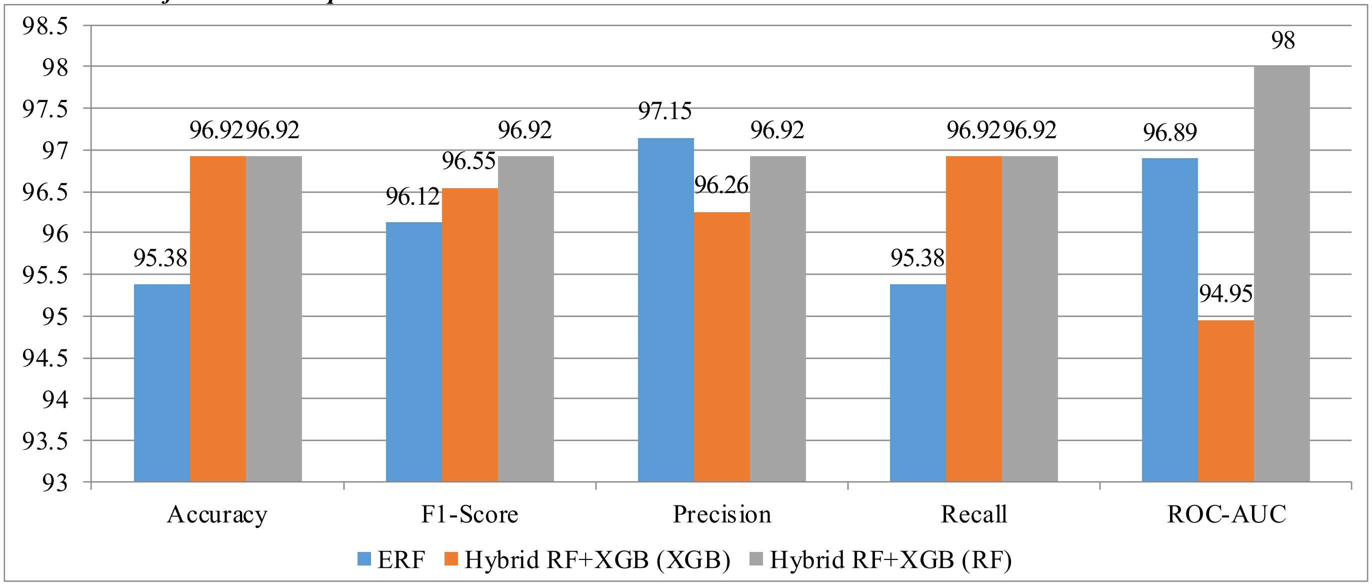


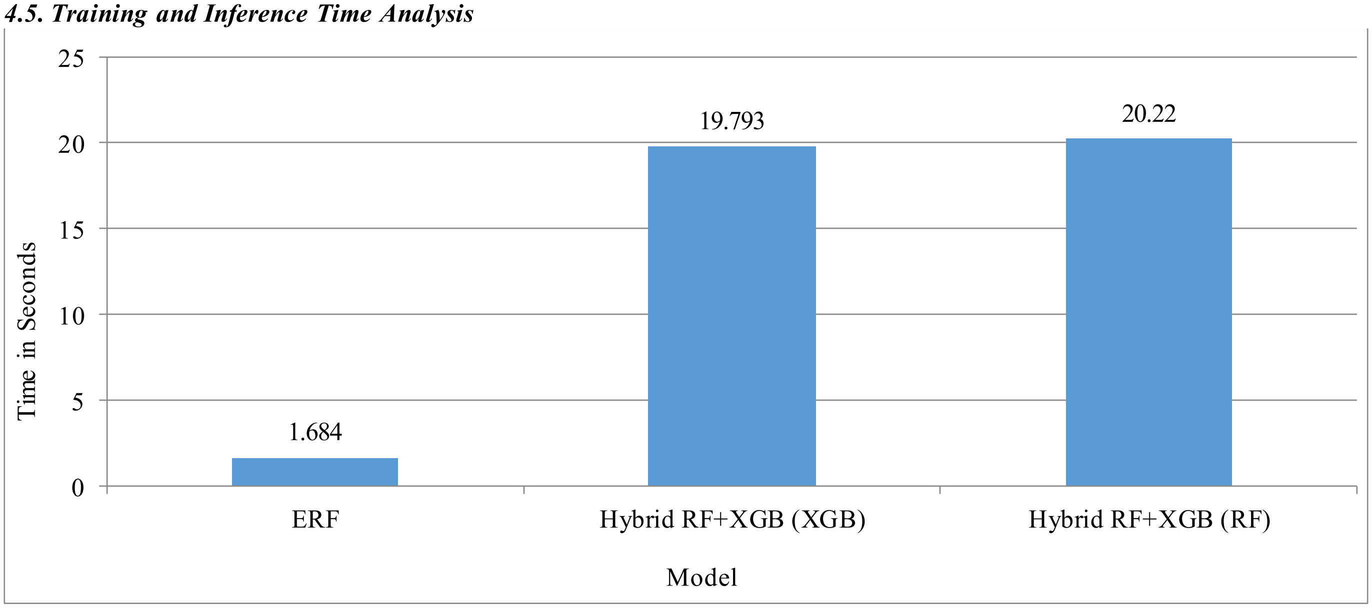
Fig. 3 Summary of model performance based on evaluation metrics

Figure 3 provides a comparative analysis of the performance of each predictive model, using key evaluation metrics. This study illustrates that stacking complex tree-based ensemble models in both base and meta layers can yield promising results across all key evaluation metrics, despite the common practice of using simple meta-learners to minimize computational overhead.

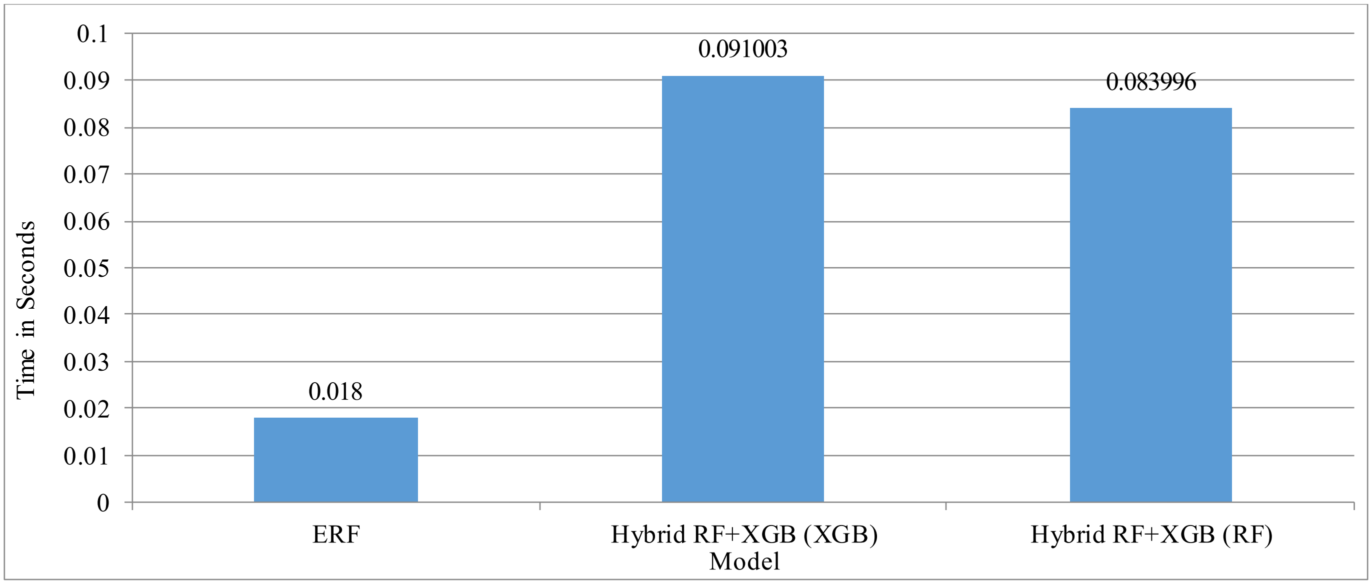
The Hybrid RF+XGB model, with RF serving as the meta-classifier, achieved superior overall metrics in accuracy, precision, recall, F1-score, and ROC-AUC, surpassing single-model alternatives such as the ERF.

This suggests that despite increased complexity, the stacking of such models can effectively learn deeper interactions and patterns in educational data that simpler combinations might overlook. However, a closer examination reveals that while the improvement in precision is marginal compared to the single ERF model, there is also a slight decline in ROC-AUC when XGB is used as the meta-classifier. This suggests that although stacking improves some metrics, its advantage is not uniformly distributed across all aspects of model performance, and more complex models may not always generalize better, particularly in sensitive imbalanced classification scenarios.

4.5. Training and Inference Time Analysis



(a)



(b)

Fig. 4 (a) Training, and (b) Inference time comparison of each model,

A significant consideration in model selection for real-time applications is computational efficiency, particularly in terms of training and inference times. Machine learning algorithms require adequate time to learn effectively, which can result in delays if not properly managed. Prediction techniques, in particular, demand both time and precision to ensure accurate results. Time complexity tends to increase with larger datasets, especially in education, where data can be wide-ranging and extensive, resulting in longer processing times as the volume of data grows [31]. The visualization in Figure 4 illustrates the differences in training and inference times, emphasizing the computational efficiency of each model. The ERF model demonstrated the fastest training time at 1.6840 seconds and the lowest inference time per batch at 0.018000 seconds, making it highly efficient for both development and deployment phases. On the other hand, the Hybrid RF + XGB (XGB) model exhibited a significantly longer training time of 19.7930 seconds and an inference time of 0.091003 seconds per batch, highlighting its computational cost. Meanwhile, the Hybrid RF + XGB (RF as meta-classifier) required the longest training time of 20.2200 seconds and an inference time of 0.083996 seconds per batch.

Although this model offered the highest performance metrics overall, the training and inference time costs must be weighed, particularly in applications requiring scalability or limited computational resources. This further highlights the computational complexity of the stacking approach. This shows that the Hybrid RF + XGB (RF) model excels in predictive performance, while ERF stands out in terms of computational efficiency.

4.6. Model Deployment Considerations

This study investigated stacking ensemble learning, specifically the combination of RF and XGB as both base and meta classifiers, to evaluate whether multi-layer ensemble models could enhance the predictive accuracy of academic performance classification. The findings indicated that the hybrid stacking model (Hybrid RF+XGB with RF as the meta-classifier) attained better results on the majority of evaluation metrics, achieving a ROC-AUC score of 98.00% along with balanced accuracy, F1-score, and precision-recall values.

These findings align with prior studies that reported stacking techniques often outperform single models in complex classification problems [36, 37]. However, while the stacked model delivered slightly better results, this study also carefully considered the computational trade-offs, particularly in training and inference times, which are essential in real-world applications. The ERF, though slightly behind in predictive accuracy, demonstrated significantly faster training (1.68s) and inference times (0.018s per batch), compared to over 20 seconds of training and 0.08s–0.09s per batch in stacked models. The findings of this study are consistent with those of [32, 33], whose work also demonstrated that combining RF with SMOTE significantly improves classification accuracy and the detection of at-risk students, which proves that enhanced RF could be a strong candidate for academic performance prediction in higher education settings.

This study sets its novelty through its approach, which goes beyond the usual comparison and evaluation of machine learning models. Many existing studies [18, 20, 22, 23] focus solely on identifying the most accurate algorithm for predicting student academic performance, in which their findings often remain in theoretical contexts. This study addresses the gap between algorithmic development and practical application by implementing the optimal model within a fully operational web-based system.

5. HEAPS System Development

The HEAPS was developed using Streamlit, an open-source Python framework designed for building interactive and user-friendly web applications. Streamlit is an optimal platform for this application as it facilitates the smooth integration of machine learning models developed in Python, while offering a user-friendly yet robust interface. Streamlit is a good choice for launching data-driven applications since it significantly reduces development time by removing the requirement for major front-end programming, unlike standard web frameworks. Moreover, HEAPS is carried out using Streamlit Community Cloud, a quick, easy and free deployment tool that allows immediate public access to the system without requiring a server configuration. HEAPS can be accessed by users straight from a shared web link, therefore saving local installations from necessity.

5.1. Key Features of the HEAPS Web Application

Easy and user-friendly design of the HEAPS web application guarantees accessibility for administrators and educators. Acting as the main page, the Home tab provides a summary of the system, its goals, and how it assists educational institutions in predicting student performance. The Home tab includes a brief introduction designed to help users easily utilize the system. Figure 5 gives an overview of the Home Tab.

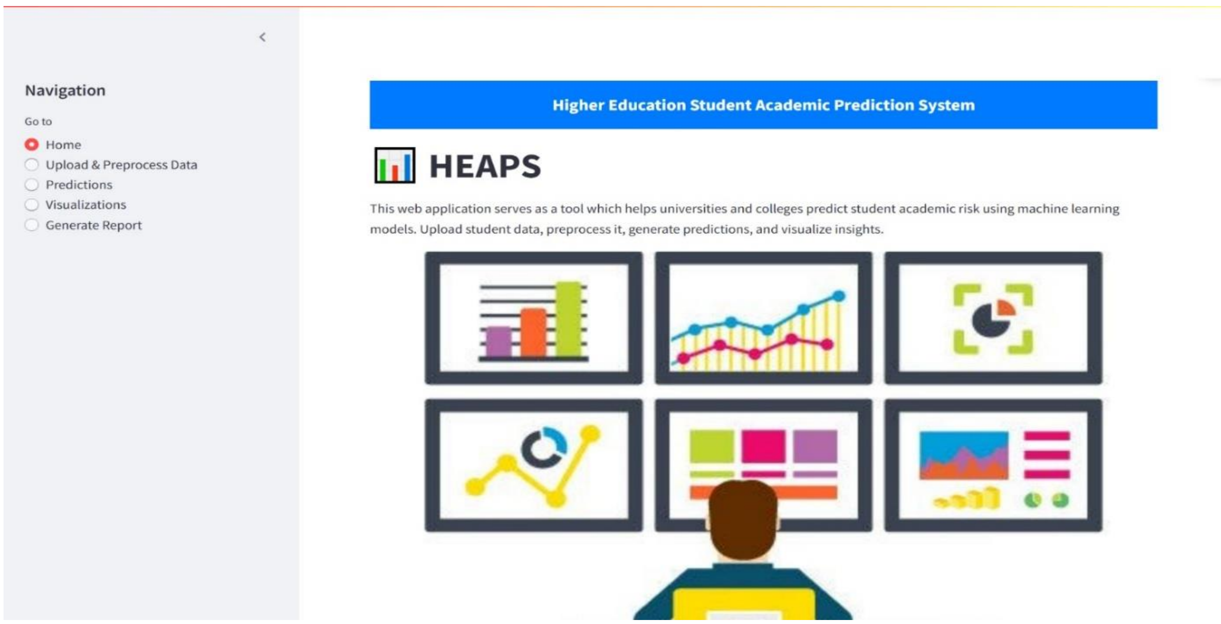


Fig. 5 HEAPS home interface

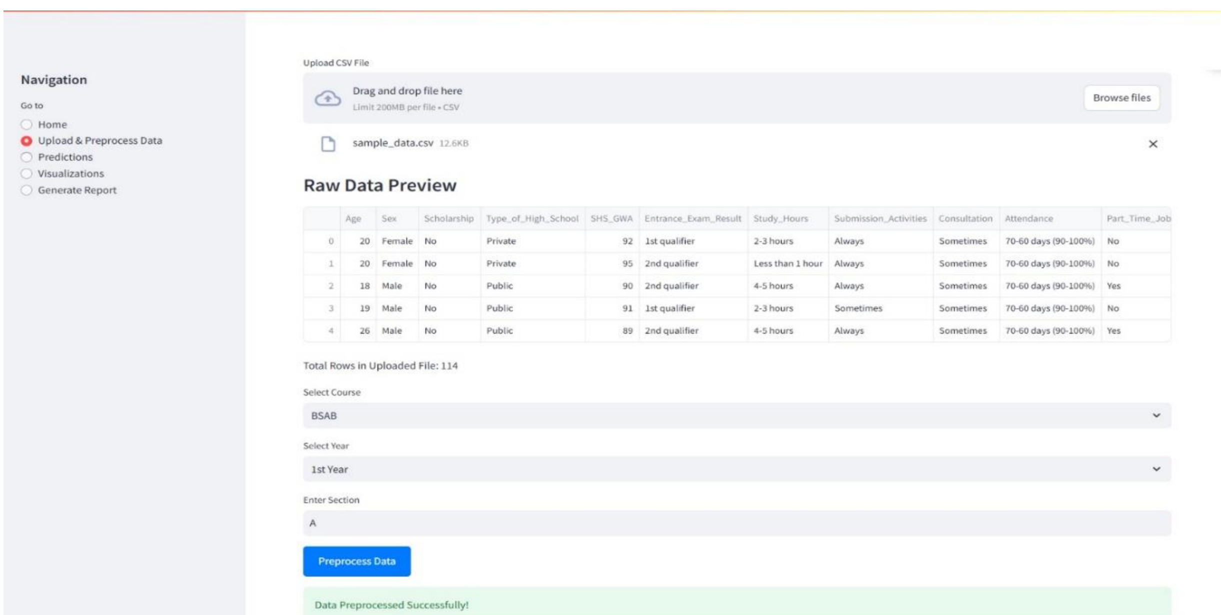


Fig. 6 Uploading and preprocessing features of the developed HEAPS

The Upload & Preprocess Data page enables users to upload student datasets in CSV format. This function ensures that the given data is automatically preprocessed using techniques including categorical encoding and addressing missing values as needed. Before generating predictions, a preprocessing button allows one to manipulate data. The system alerts users when the uploaded dataset deviates from the specified format, therefore maintaining data integrity. Figure 6 displays the Upload & Preprocess Data Tab. The Predictions tab is where the trained ERF model is applied to classify students as either "Low Risk" or "At Risk." Users can view prediction results in an intuitive tabular format and input individual student records for real-time predictions. The system ensures that all preprocessed data is correctly formatted before running predictions to enhance accuracy. After the prediction is performed, users can download the CSV file containing the predicted students classified as either at risk or low risk.

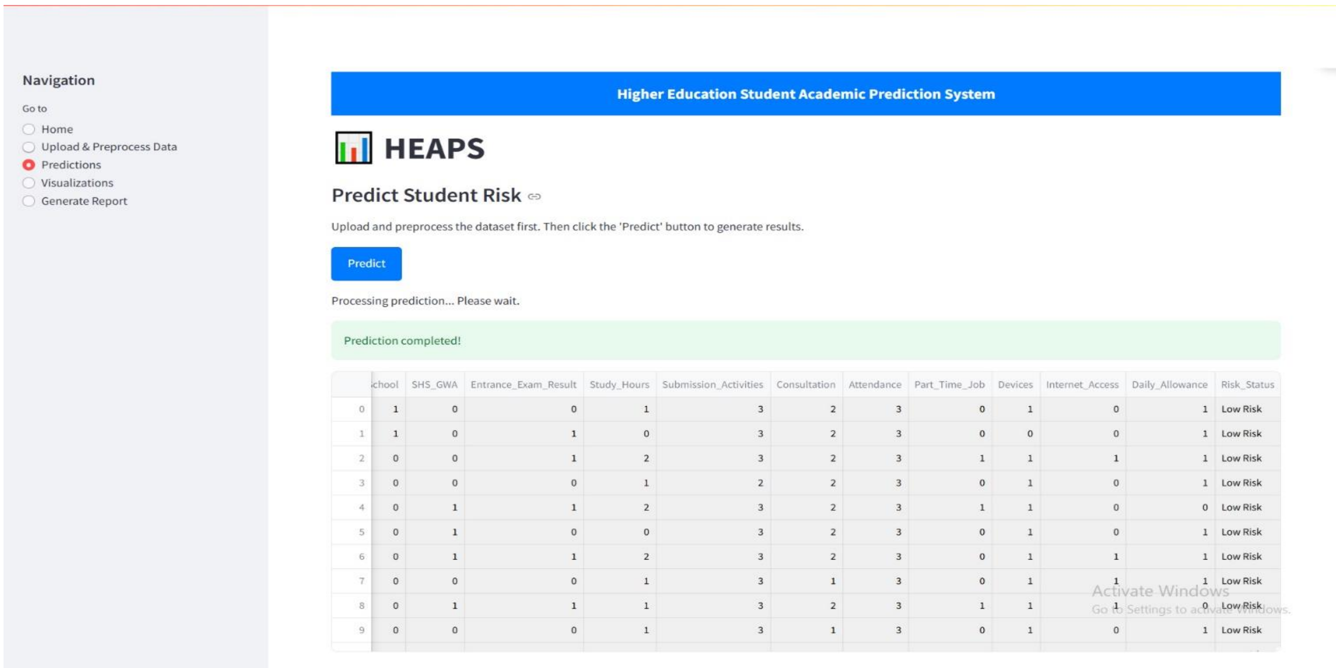


Fig. 7 Predicted at risk status of students using HEAPS

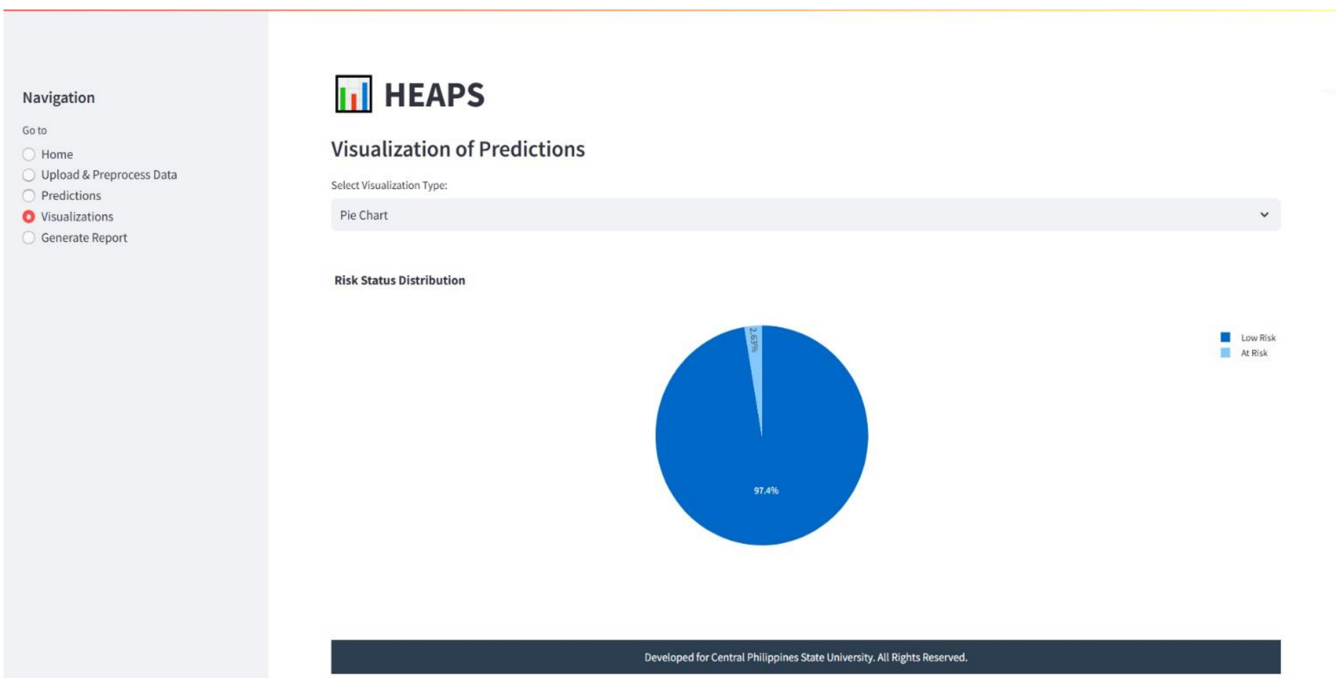


Fig. 8 Visualization of students' at risk

Figure 7 illustrates the Predictions Tab. The Visualizations tab depicted in Figure 8 provides graphical representations of student academic risk levels. This tab contains various visualizations: a pie chart illustrating the percentage of students classified as "Low Risk" versus "At Risk," a bar chart depicting the total number of students in each category, a feature importance chart to assist faculty and administrators in identifying the most significant factors impacting student performance, and a histogram to represent the distribution of each feature based on the gathered data.

These visualizations enable users to examine trends and facilitate data-driven decision-making. The Generate Report option allows users to export prediction results for documentation and additional analysis. Users may filter reports by course, year level, or part prior to being generated. The application enables the downloading of reports in PDF or CSV formats for academic record keeping. These features significantly improve the usability of the HEAPS system, guaranteeing a seamless experience for users while offering useful information on student performance and academic risk levels.

5.2. HEAPS Performance and Usability Evaluation Results

The practicality of HEAPS was evaluated by a survey questionnaire administered to 30 educators from diverse HEIs. Using a Likert scale, participants evaluated every question to show their level of agreement with system usability assertions. The percentages were utilized to assess the overall usability of HEAPS in important areas like navigation ease, learnability, efficiency, error management, output clarity, design, system responsiveness, and accessibility.

Table 4. Collected survey findings for the usability of HEAPS

Characteristics	Description	5-point Likert Scale				
		5	4	3	2	1
Ease of Navigation	The web application is easy to navigate	27	3	0	0	0
	Menus, buttons and functions are organized and user-friendly	20	9	1	0	0
Learnability	I was able to understand how to use the application without requiring extensive instructions or training	19	10	1	0	0
Efficiency	The web application allows me to complete tasks (e.g., uploading data, preprocessing, generating reports) quickly and efficiently.	20	9	1	0	0
	The system minimizes unnecessary steps, allowing for a smooth and fast workflow	22	8	0	0	0
Error Prevention & Handling	The application effectively prevents common errors, such as incorrect file uploads or missing values	19	11	0	0	0
	Error messages are clear, informative, and help me understand how to resolve issues	16	12	2	0	0
Clarity of Outputs	The prediction results are presented in a clear and easy-to-understand format	21	9	0	0	0
	The generated reports contain well-structured and useful information for decision-making	20	10	0	0	0
Aesthetic & Design	The application has a visually appealing and professional design	21	9	0	0	0
	The use of colors, fonts, and layouts enhances readability and user experience	20	8	2	0	0
System Responsiveness	The application responds quickly to my actions without significant delays	17	13	0	0	0
	The system processes data (e.g., uploading, prediction, report generation) in a reasonable amount of time	20	9	1	0	0
Accessibility	The application is user-friendly for both technical and non-technical users	23	7	0	0	0
	The text, buttons, and overall layout are easy to read and interact with	22	8	0	0	0

Footnote: Each number in the table denotes the frequency of participant responses for each question.

The findings presented in Table 4 demonstrate a significant degree of usability satisfaction across various criteria. Regarding navigation, 86.67% of participants strongly agreed that the application is user-friendly, while 96.67% noted that the menus and buttons are well-organized. In terms of learnability, 96.67% of participants concurred that the system necessitates minimal training. HEAPS exhibited notable efficiency, as 73.33% of respondents strongly agreed that the workflow is both smooth and fast. Positive comments were given for the system's error prevention and management features; 63.33% of respondents strongly agreed on their efficacy in preventing mistakes, and 93.33% of respondents stated the error messages given are informative. Regarding the clarity of outputs, all of the respondents agreed that predictions and reports are well-structured and easy to understand. The application received a high score in aesthetics and design, with 70.00% of respondents strongly agreeing on its visual appeal. Positive evaluations of system responsiveness came from all of the respondents since they agreed on its effective data processing capacity.

Finally, HEAPS has shown great accessibility, with 76.67% of respondents strongly agreeing that it is easy for both technical and non-technical users. The notable percentage of Strongly Agree and Agree replies points to HEAPS as a useful tool for predicting student academic risk since it is efficient, quick, and easily available. This tool greatly benefits educators in HEIs.

6. Conclusion and Recommendation

The study focused on using HEAPS, a web-based tool that employs machine learning algorithms, to identify students at risk in higher education. By means of extensive experimentation, the Hybrid RF + XGB model, utilizing Random Forest as the meta-classifier, showcased exceptional performance and obtained the best results on multiple key metrics. In contrast, ERF achieved competitive outcomes, demonstrating the highest precision alongside the fastest training and inference times, thereby illustrating a balance between computational efficiency and predictive capability. This comparative analysis demonstrates that the stacking-based hybrid approach provides superior performance, whereas a single model, such as ERF, is still a feasible choice for faster deployments in resource-limited settings. This study contributes by integrating ensemble learning techniques specifically designed for educational contexts and implementing these optimized models in a fully operational, user-friendly web application. In contrast to prior research that concludes with model development, HEAPS converts predictive analytics into a practical academic tool.

The usability evaluation of the HEAPS demonstrates its effectiveness, efficiency, and accessibility in identifying at-risk students in higher education. The application exhibits high satisfaction ratings from educators, characterized by ease of use, fast processing, clear output presentation, and robust error-handling mechanisms, thereby rendering it appropriate for academic monitoring and decision-making. Integrating HEAPS into institutional practices allows higher education institutions to leverage data-driven insights, facilitating early intervention strategies that enhance student success. Upon full implementation, HEAPS may enhance student retention, improve academic support programs, and contribute to institutional success.

In conclusion, performance metrics serve as important indicators of model capability; however, they should be interpreted in conjunction with additional important deployment factors, including computational efficiency and system latency. In contexts such as HEAPS, where timely risk prediction and common institutional adoption are targets, a faster and simpler model may offer an ideal combination of accuracy and practicality, rendering it the more suitable option despite the minor performance advantage of more complex hybrid models.

6.1. Future Works and Ethical Considerations

Future research may investigate hybrid techniques that optimize computational efficiency while maintaining model performance. Investigating the optimization of stacking models to minimize training and inference times while preserving high accuracy is warranted. Moreover, the inclusion of feature selection techniques, lowering dimensionality and enhancing interpretability could increase efficiency. Testing the models on larger and more varied datasets would help to confirm their generalizability among several educational institutions.

Future researchers are urged to collect student data from several higher education institutions all throughout the Philippines to more equally portray the national student population and increase the applicability of the model to institutions with different academic environments and student demographics. Including additional relevant variables not covered in this research might improve the predictive power of the model. Combining HEAPS with current Learning Management Systems (LMS) could also enable customized support and timely notifications, thereby enhancing academic advice and student services.

Strategic planning and institutional decision-making in higher education are acknowledged to depend on the inclusion of machine learning. Still, its application has to be carefully aligned with legal and ethical standards. Within the framework of the Philippines, HEIs have to establish and implement thorough data governance policies that are compliant with the Data Privacy Act of 2012.

These guidelines have to ensure that data gathered from institutional systems, faculty, and students is handled to the best degree of confidentiality and consent. Data usage transparency, secure storage, and individuals' rights to access and amend their data are essential elements of responsible AI integration in education. A strategic implementation of tools like HEAPS requires careful planning. Adoption must be systematic and integrated, guided by institutional policies, readiness of infrastructure, and a culture that supports data-informed decision-making.

Acknowledgments

The researchers would like to express their sincere gratitude to Central Philippines State University for granting them access to essential data, which made this research study possible. Their support and commitment to academic excellence have been invaluable in the completion of this work.

References

[1] Erik W. Aslaksen, *The Role of Education*, The Evolution of Society: An Information-Processing Perspective, Springer, pp. 207-227, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[2] Dilnavoz Shavkidinova, Feruza Suyunova, and Jasmina Kholdarova, "Education is an Important Factor in Human and Country Development," *Current Research Journal of Pedagogics*, vol. 4, no. 1, pp. 27-34, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[3] AnneDuffy et al., "Predictors of Mental Health and Academic Outcomes in First-Year University Students: Identifying Prevention and Early-Intervention Targets," *BJPsych open*, vol. 6, no. 3, pp. 1-8, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[4] Richard J. Daker et al., "First-Year Students' Math Anxiety Predicts STEM Avoidance and Underperformance throughout University, Independently of Math Ability," *npj Science of Learning*, vol. 6, no. 1, pp. 1-13, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[5] Kevin L. Rand et al., "Hope and Optimism as Predictors of Academic Performance and Subjective Well-Being in College Students," *Learning and Individual Differences*, vol. 81, pp. 1-35, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[6] D. Van Staden, and P. Naidoo, "Future-Proofing Imperatives for Remote Online Teaching, Learning and Student Support in the Context of Pandemic Change and Beyond: A Case for South African Higher Education Transformation," *South African Journal of Higher Education*, vol. 36, no. 3, pp. 269-281, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[7] Sky Mkuti, and Elzabé Aucamp, "Contextualising the South African Higher Education Academic Interventions for Student Success: Challenges and Opportunities," *Proceedings of the 1st International Conference on Education Research*, vol. 1, no. 1, pp. 194-201, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[8] Muhammad Imran et al., "Early Identification and Intervention: Amplifying the Voice of Slow Learners," *AITU Scientific Research Journal*, vol. 2, no. 2, pp. 17-25, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[9] Amal Abdulla Al Hassani, and Stephen Wilkins, "Student Retention in Higher Education: The Influences of Organizational Identification and Institution Reputation on Student Satisfaction and Behaviors," *International Journal of Educational Management*, vol. 36, no. 6, pp. 1046-1064, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[10] Chaka Chaka, "Educational Data Mining, Student Academic Performance Prediction, Prediction Methods, Algorithms and Tools: An Overview of Reviews," *Journal of E-Learning and Knowledge Society*, vol. 18, no. 2, pp. 58-69, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[11] Sunita M. Dol, and Pradip M. Jawandhiya, "Systematic Review and Analysis of EDM for Predicting the Academic Performance of Students," *Journal of The Institution of Engineers (India): Series B*, vol. 105, no. 4, pp. 1021-1071, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[12] Opeyemi Ojajuni et al., "Predicting Student Academic Performance using Machine Learning," *International Conference on Computational Science and Its Applications*, Cagliari, Italy, pp. 481-491, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[13] Juan L. Rastrollo-Guerrero et al., "Analyzing and Predicting Students' Performance by Means of Machine Learning: A Review," *Applied Sciences*, vol. 10, no. 3, pp. 1-16, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[14] Nasir Abdul Jalil et al., "Machines Learning Trends, Perspectives and Prospects in Education Sector," *ICEMT '19: Proceedings of the 3rd International Conference on Education and Multimedia Technology*, Nagoya Japan, pp. 201-205, 2019,. [CrossRef] [Google Scholar] [Publisher Link]

[15] Yuri Nieto et al., "Usage of Machine Learning for Strategic Decision Making at Higher Educational Institutions," *IEEE Access*, vol. 7, pp. 75007-75017, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[16] Lydia T. Liu et al., "Reimagining the Machine Learning Life Cycle to Improve Educational Outcomes of Students," *Proceedings of the National Academy of Sciences*, vol. 120, no. 9, pp. 1-12, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[17] Hui Luan, and Chin-Chung Tsai, "A Review of Using Machine Learning Approaches for Precision Education," *Educational Technology & Society*, vol. 24, no. 1, pp. 250-266, 2021. [Google Scholar] [Publisher Link]

[18] Mustafa Yağcı, "Educational Data Mining: Prediction of Students' Academic Performance using Machine Learning Algorithms," *Smart Learning Environments*, vol. 9, no. 1, pp. 1-19, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[19] Iddrisu Issah et al., "A Systematic Review of the Literature on Machine Learning Application of Determining the Attributes Influencing Academic Performance," *Decision Analytics Journal*, vol. 7, pp. 1-11, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[20] P. Ajay et al., "Prediction of Student Performance using Random Forest Classification Technique," *International Research Journal of Engineering and Technology (IRJET)*, vol. 7, no. 4, pp. 405-408, 2020. [Google Scholar] [Publisher Link]

[21] Wei Zhang, Yu Wang, and Suyu Wang, "Predicting Academic Performance using Tree-Based Machine Learning Models: A Case Study of Bachelor Students in an Engineering Department in China," *Education and Information Technologies*, vol. 27, no. 9, pp. 13051-13066, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[22] Leena H. Alamri et al., "Predicting Student Academic Performance using Support Vector Machine and Random Forest," *ICETM '20: Proceedings of the 2020 3rd International Conference on Education Technology Management*, London United Kingdom, pp. 100-107, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[23] Arash Khosravi, and Ahmad Azarnik, “Leveraging Educational Data Mining: XGBoost and Random Forest for Predicting Student Achievement,” *International Journal of Data Science and Advanced Analytics*, vol. 6, no. 7, pp. 387-393, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[24] Muhammad 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. [CrossRef] [Google Scholar] [Publisher Link]

[25] Martijn Meeter, “Predicting Retention in Higher Education from High-Stakes Exams or School GPA,” *Educational Assessment*, vol. 28, no. 1, pp. 1-10, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[26] Diego Buenaño-Fernández, David Gil, and Sergio Luján-Mora, “Application of Machine Learning in Predicting Performance for Computer Engineering Students: A Case Study,” *Sustainability*, vol. 11, no. 10, pp. 1-18, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[27] Bashir Khan Yousafzai, Maqsood Hayat, and Sher Afzal, “Application of Machine Learning and Data Mining in Predicting the Performance of Intermediate and Secondary Education Level Student,” *Education and Information Technologies*, vol. 25, no. 6, pp. 4677-4697, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[28] Erwin T. Lau, Licheng Sun, and Qingping Yang, “Modelling, Prediction and Classification of Student Academic Performance using Artificial Neural Networks,” *SN Applied Sciences*, vol. 1, no. 9, pp. 1-10, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[29] Rüdiger Wirth, and Jochen Hipp, “CRISP-DM: Towards a Standard Process Model for Data Mining,” *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*, vol. 1, pp. 29-40, 2000. [Google Scholar] [Publisher Link]

[30] Aslina Saad, and Shahnita Shaharin, “The Methodology for Ontology Development in Lesson Plan Domain,” *Methodology*, vol. 7, no. 4, pp. 557-562, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[31] Edeh Michael Onyema et al., “Prospects and Challenges of using Machine Learning for Academic Forecasting,” *Computational Intelligence and Neuroscience*, vol. 2022, no. 1, pp. 1-7, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[32] Khurram Jawad, Muhammad Arif Shah, and Muhammad Tahir, “Students’ Academic Performance and Engagement Prediction in a Virtual Learning Environment using Random Forest with Data Balancing,” *Sustainability*, vol. 14, no. 22, pp. 1-15, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[33] Vaneza Flores, Stella Heras, and Vicente Julian, “Comparison of Predictive Models with Balanced Classes using the SMOTE Method for the Forecast of Student Dropout in Higher Education,” *Electronics*, vol. 11, no. 3, pp. 1-16, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[34] Tarid Wongvorachan, Surina He, and Okan Bulut, “A Comparison of Undersampling, Oversampling, and SMOTE Methods for Dealing with Imbalanced Classification in Educational Data Mining,” *Information*, vol. 14, no. 1, pp. 1-15, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[35] Thao-Trang Huynh-Cam, Long-Sheng Chen, and Huynh Le, “Using Decision Trees and Random Forest Algorithms to Predict and Determine Factors Contributing to First-Year University Students’ Learning Performance,” *Algorithms*, vol. 14, no. 11, pp. 1-17, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[36] Fuxing Yu, and Xinran Liu, “Research on Student Performance Prediction based on Stacking Fusion Model,” *Electronics*, vol. 11, no. 19, pp. 1-13, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[37] Lijuan Yan, and Yanshen Liu, “An Ensemble Prediction Model for Potential Student Recommendation using Machine Learning,” *Symmetry*, vol. 12, no. 5, pp. 1-17, 2020. [CrossRef] [Google Scholar] [Publisher Link]