FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 9 No. 4, April, 2025, pp 215 – 222
DOI: https://doi.org/10.33003/fjs-2025-0904-3004

HYBRID PREDICTIVE MODEL FOR STUDENTS' ACADEMIC PERFORMANCE BASED ON MACHINE LEARNING APPROACH

***[1]Abdullahi Bashar Abubakar, [2]Danlami Gabi, [2]Muhammad Garba,**
[2]Nasiru Muhammad Dankolo and [1]Abubakar Hassan

[1]Department of Computer Science, Collage of Science and Technology, Umaru Ali Shinkafi polytechnic, Sokoto, Nigeria
[2]Department of Computer Science, Faculty of Science, Kebbi State University of Science and Technology Aliero, Nigeria

*Corresponding authors' email: basharabdullahi86@gmail.com

ABSTRACT

Student academic performance is a critical factor in assessing the quality of education and institutional effectiveness. Various factors, including socioeconomic background, institutional policies, prior academic achievement, and learning environments, contribute to student success. Understanding these factors is essential for developing targeted interventions to improve academic outcomes. This research investigates the factors influencing student academic performance at Umaru Ali Shinkafi Polytechnic through the development and implementation of a hybrid prediction model. By drawing on diverse data sources and advanced analytical techniques, the study aims to uncover the complex interplay of school factors, prior academic achievement, and other relevant variables shaping student outcomes. Utilizing a stratified random sampling technique, data was collected from a population of students at the institution. A hybrid prediction model incorporating linear regression, decision trees, and random forests was developed and evaluated on a test dataset consisting of 254 records. The model achieved an accuracy of 0.85, a precision of 0.82, a recall of 0.88, an F1-score of 0.85, and an ROC-AUC score of 0.91. These results indicate that the hybrid model outperforms benchmark techniques, providing robust predictive performance and significant insights into the factors affecting student success. These findings offer actionable recommendations for educators, policymakers, and stakeholders to enhance educational strategies and support student achievements.

Keywords: Machine learning, Predictive modeling, Hybrid model, Academic performance

INTRODUCTION

In the educational landscape of Umaru Ali Shinkafi Polytechnic, Sokoto, the identification and understanding of factors that contribute to varying levels of student performance are crucial for promoting academic excellence and student success. By discovering these factors, educational institutions can develop effective strategies and interventions to support students and improve overall academic outcomes (John, 2023). The influencing of advancements in predictive modeling and machine learning techniques to investigate the underlying factors influencing student performance at Umaru Ali Shinkafi Polytechnic, Sokoto is the main aim of this research.

Furthermore, the primary objective of this study is to employ predictive modeling using machine learning algorithms to analyse and understand the factors resulting in different levels of performance among students within Umaru Ali Shinkafi Polytechnic, Sokoto. By leveraging the rich and comprehensive dataset available within the institution's student information system, this research will provide valuable insights into the factors affecting student performance and their potential implications.

The dataset for this study encompasses a diverse range of inputs, including academic records, attendance records, assessment scores, and potentially additional relevant information specific to Umaru Ali Shinkafi Polytechnic, Sokoto. By utilizing this comprehensive dataset, the research captured the unique characteristics and nuances of student performance within the institution.

The predictive modeling approach involve the application of cutting-edge, the machine learning algorithms, such as regression models, decision trees, or random forest methods (Tommy, 2022). These models are trained and validated using historical data from Umaru Ali Shinkafi Polytechnic, Sokoto, to enable the identification of key predictors that contribute to high or low student performance. Through accurate predictions and data-driven insights, this research shed light on the underlying factors that significantly influence student outcomes at the institution.

The findings of this research have practical implications for educational practitioners, administrators, and policymakers at Umaru Ali Shinkafi Polytechnic, Sokoto. Understanding the factors that influence student performance will enable the development of targeted interventions, personalized support systems, and evidence-based instructional strategies (Sayid, 2023, p. 102).. Ultimately, the aim is to enhance student success, improve graduation rates, and foster a conducive learning environment within the institution.

Adigun, Olatinwo, and Oladele, (2023) Conducted a research on Machine Learning Techniques for Predicting Students' Academic Performance. The study explores the development of predictive models for student academic performance using machine learning algorithms, including Logistic Regression (LR), Support Vector Machines (SVM), and a hybrid of both (LR+SVM). The research highlights the effectiveness of these models in enhancing prediction accuracy, which aligns and focus on hybrid predictive models for student performance.

With application of predictive modeling and machine learning techniques, this research seeks to contribute to the current collection of knowledge on student performance evaluation. Furthermore, it aims to generate data-driven insights specific to Umaru Ali Shinkafi Polytechnic, Sokoto, fostering evidence-based decision-making and guiding the implementation of effective educational practices.

Finally, this research leverages the power of predictive modeling and machine learning algorithms to uncover the underlying factors that contribute to different levels of student performance at Umaru Ali Shinkafi Polytechnic, Sokoto. By doing so, it aims to facilitate informed decision-making, improve educational outcomes, and drive positive change within the institution.

Predictive modelling for understanding the factors affecting the analysis of student performance includes a number of various elements and influences that make a difference on the academic achievements and outcomes of students within an educational system (Lateef, 2018). These factors are categorized into several dimensions, including individual, school, socioeconomic, and peer and social factors. The system will use historical data and statistical techniques, predictive models to identify patterns and relationships between various factors and student outcomes. To break the system further the outline of the steps involve: data collection, data preparation, model selection, model training, model evaluation, interpretation and sight and validation. By leveraging predictive modeling techniques, educational institutions can acquire important knowledge about the elements that influence student performance. This information can help inform targeted interventions, policy decisions, and resource allocation to support student success and improve educational outcomes (Sege, 2018).

Ensuring the academic success and performance of the students is a top priority for Umaru Ali Shinkafi Polytechnic, Sokoto. To achieve this goal, it is essential to identify and comprehend the factors that contribute to the varying levels of student performance within the institution. In the institution the students' performance evaluation has traditionally relied on traditional assessment methods and subjective measures, recent advancements in data analysis and machine learning techniques offer new possibilities for understanding the complex dynamics that influence student outcomes. By adapting the power of predictive modeling and machine learning algorithms, educational institutions can gain deeper insights into the underlying factors shaping student performance (Abdullah, 2020).

According to James (2019). The study of factors affecting student performance is multifaceted and encompasses a wide range of variables. It extends beyond academic records and incorporates elements such as student engagement, motivation, learning styles, study habits, and socio-economic backgrounds. By considering these diverse factors, educational institutions can develop targeted strategies and interventions that address the unique challenges faced by their students.

However, the specific factors influencing student performance can vary across different institutions and contexts (Alfred, 2023). Therefore, conducting research at Umaru Ali Shinkafi Polytechnic, Sokoto, provides an opportunity to investigate the factors that are most relevant to the student population within the institution. The unique characteristics of the institution, such as its curriculum, teaching methods, student support services, and campus environment, may influence student performance and warrant in-depth analysis.

Understanding the factors affecting student performance at Umaru Ali Shinkafi Polytechnic, Sokoto, can guide the institution in developing evidence-based practices to enhance teaching and learning outcomes. By identifying the factors that contribute to high or low performance, the institution can allocate resources effectively, design targeted support programs, and implement interventions that address the specific needs of its students.

Furthermore, this research aims to contribute to the broader body of knowledge on student performance evaluation and data-driven decision-making in the field of education. By examining the factors influencing student performance at Umaru Ali Shinkafi Polytechnic, Sokoto, this study can provide valuable insights and practical recommendations for educational practitioners, policymakers, and researchers in similar educational contexts.

Conclusively, conducting research on the factors affecting student performance at Umaru Ali Shinkafi Polytechnic, Sokoto, is essential for understanding the unique dynamics within the institution and identifying strategies to improve student outcomes. By employing predictive modeling and machine learning algorithms, this research aims to unlock valuable insights that can inform evidence-based decision-making, enhance teaching practices, and ultimately contribute to the overall success of UASPoly, Sokoto.

MATERIALS AND METHODS

The methodical process for creating a hybrid machine learning model to forecast students' academic success is described in this section. The study generated precise and understandable predictions by utilizing decision trees and linear regression inside a strong framework of data collection, preprocessing, model training, and validation. In addition to guaranteeing that the research complies with academic requirements, ethical concerns and restriction acknowledgment also lay the groundwork for future investigations.

Research Philosophy

The research philosophy for the research on "Hybrid Predictive Students' Academic Performance Model Based on Machine Learning Approach" is likely based on positivism. This philosophy aligns with the scientific method, focusing on objective measurements and observable phenomena.

Positivism is the most suitable research philosophy for the research, because it emphasizes objectivity, empirical data collection, quantitative analysis, and hypothesis testing—all of which are integral to developing and validating a machine learning model for predicting academic performance.

Research Approach

The research on "Hybrid Predictive Students' Academic Performance Model Based on Machine Learning Approach" most likely has positivism as its guiding ideology. By emphasizing measurable quantities and observable occurrences, this viewpoint is consistent with the scientific method. In order to design and validate a predictive machine learning model for student's academic achievement, positivism is the best appropriate research philosophy. This is because positivism promotes objectivity, empirical data collection, quantitative analysis, and hypothesis testing.

Methodological choice

This study uses model for machine learning and quantitative techniques in an organized, data-driven manner to forecast student achievement. Combining decision tree and linear regression models with the right methods for data sampling, collecting, and assessment guarantees that the study is solid, repeatable, and based on empirical research.

Research Strategy

The quantitative, data-driven predictive analysis was the main focus of the research strategy. In an attempt to increase prediction accuracy, machine learning models (decision trees and linear regression) are combined into a hybrid framework. Based on cross-sectional data analysis, the approach is iterative and ensures dependable and generalizable outcomes through model validation and refinement at its heart.

Approach To Hybrid Model Development

The process of creating a model that is hybrid is choosing two complementing models, such as decision trees and linear regression, training them independently, and then integrating the results using ensemble methods like stacking or mixing. By using the chosen approach, the hybrid model's output may capture both linear and non-linear patterns in the data on students' academic performance while maintaining a balance between simplicity and accuracy. Cross-validation validation and hyperparameter optimization are used to validate the model and make sure the hybrid model performs better than either model alone.

Sample Size Determination

The sample size for this research ideally has been determined through statistical considerations to ensure adequate representation and sufficient power for analysis. Given the complexity and variability inherent in studying the elements that impact students' academic achievement, a sample size calculation based on statistical power analysis or considerations of effect size and variability may be employed. Specifically, the sample size (n) can be calculated using the following equation (Scribbr, 2023; Statistics by Jim, 2023).

$$n = \frac{Z^2.p.(1-p)}{E^2} \qquad (1)$$

where:

Z is the Z-value (e.g., 1.96 for a 95% confidence level),

p is the estimated proportion of the population (e.g., 0.5 for maximum variability),

E is the margin of error (e.g., 0.05 for ±5% precision).

Given these parameters, the number of samples for this study was determined to be 254 students, ensuring sufficient power to detect significant effects and variability in the factors being studied. Additionally, practical constraints such as time, resources, and accessibility to the student population influenced choosing the appropriate sample size. While a larger sample size is preferred to increase the reliability and generalizability of findings, it must be balanced with feasibility and ethical considerations. The chosen sample size aligns with the research objectives, statistical methods employed, and the desired level of precision in the study results.

The Implementation Stages The implementation decision entails the tools and techniques adopted for this research study. It involved gathering and preprocessing the data, building individual models (linear regression and decision tree), and combining them into a hybrid model using ensemble techniques. This stage includes training the models, evaluating their performance with cross-validation and performance metrics, and fine-tuning hyperparameters.

Tools and techniques (Justification)

The tools and techniques chosen for this research are based on their ability to efficiently handle the demands of data preprocessing, machine learning model building, and evaluation:

Python and its Pandas library were chosen for this research due to their extensive capabilities in handling and manipulating data efficiently. Python is widely favored in the machine learning community for its simple syntax, making it easier to develop and debug models. Additionally, its rich ecosystem of libraries such as Scikit-learn for machine learning algorithms and Pandas for data manipulation enables seamless data processing and model building. Pandas, specifically, is crucial for tasks like data cleaning, alignment, and merging, making it highly effective for transforming raw data into a format suitable for machine learning Scalable Path, (2024).

The ensemble learning techniques, Random Forests with Gradient Boosting Machines GeeksforGeeks, (2023), which combine the strengths of many decision trees for even more accurate predictions, were used in the research for their ability to improve prediction accuracy by combining multiple models

Cross-validation and performance metrics are used to ensure the models perform well and are generalizable.

Visualization tools help in understanding model behavior and feature importance, and deployment tools ensure the model can be utilized in real-world scenarios.

Models

This section presents the specifications of the models applied in this study.

Model Selection

Choosing the right model is like finding your way through a maze of options. It's a journey of exploration where the research focused in the best way to predict what affects student grades. There are many different methods that can be use, each with its own strengths and quirks.

The goal is to find a model that not only understands the data but also works well with new information, giving me insights beyond just what is seen before. The study starts by looking at what other studies have found and what methods they've used to understand student success.

Linear regression is one of the first options the research considered. Linear regression is one of the simplest yet most effective tools for predictive modeling and data analysis. It is often chosen for its ease of interpretation and straightforward implementation. linear regression is used to examine how variables such as background, motivation, and teaching style impact students' grades. These types of regression models are particularly useful when the goal is to understand the relationship between one or more independent variables (like motivation and teaching style) and a dependent variable (grades) Bevans, R. (2023, June 22).

But as the research dug deeper into modeling, the study found more complex options like decision trees (Huynh-Cam and Chen, 2021). These are great for understanding how different factors interact to influence grades. It also explored ensemble techniques such as Random Forests and Gradient Boosting Machines, which combine the strengths of many decision trees for even more accurate predictions (Patacsil, 2020; Aman, 2023).

Model development

On a mission to build a special prediction model for student academic performance. Using different methods like linear regression with the decision trees, and random forests to find out what factors affect how well students do in school.

It starts by gathering a bunch of information about students, like their gender, motivation, and background which was done from the previous section of this research. Each piece of info is like a clue that helps us understand why some students do better than others.

Next, the study trains three different models: one that looks for simple patterns, one that looks for more complex patterns, and one that combines lots of different patterns from trees. Each model has its own way of understanding the data.

As these models learn from the data, they start to spot connections and patterns that might not be obvious at first. The linear regression model looks for straight-line relationships, the decision tree model splits the data into different groups, and the random forest model combines lots of tree-based insights.

After the models were trained, they were combined together like pieces of a puzzle. By averaging their predictions, the research creates a hybrid model that's stronger and more accurate than any of them alone. It's like putting together different clues to solve a bigger mystery.

With the hybrid model in hand, the research gained valuable insights into what factors really matter for student success. It's a testament to teamwork and creativity, as the strive to give educators and others the tools they needed to support students on their academic journey.

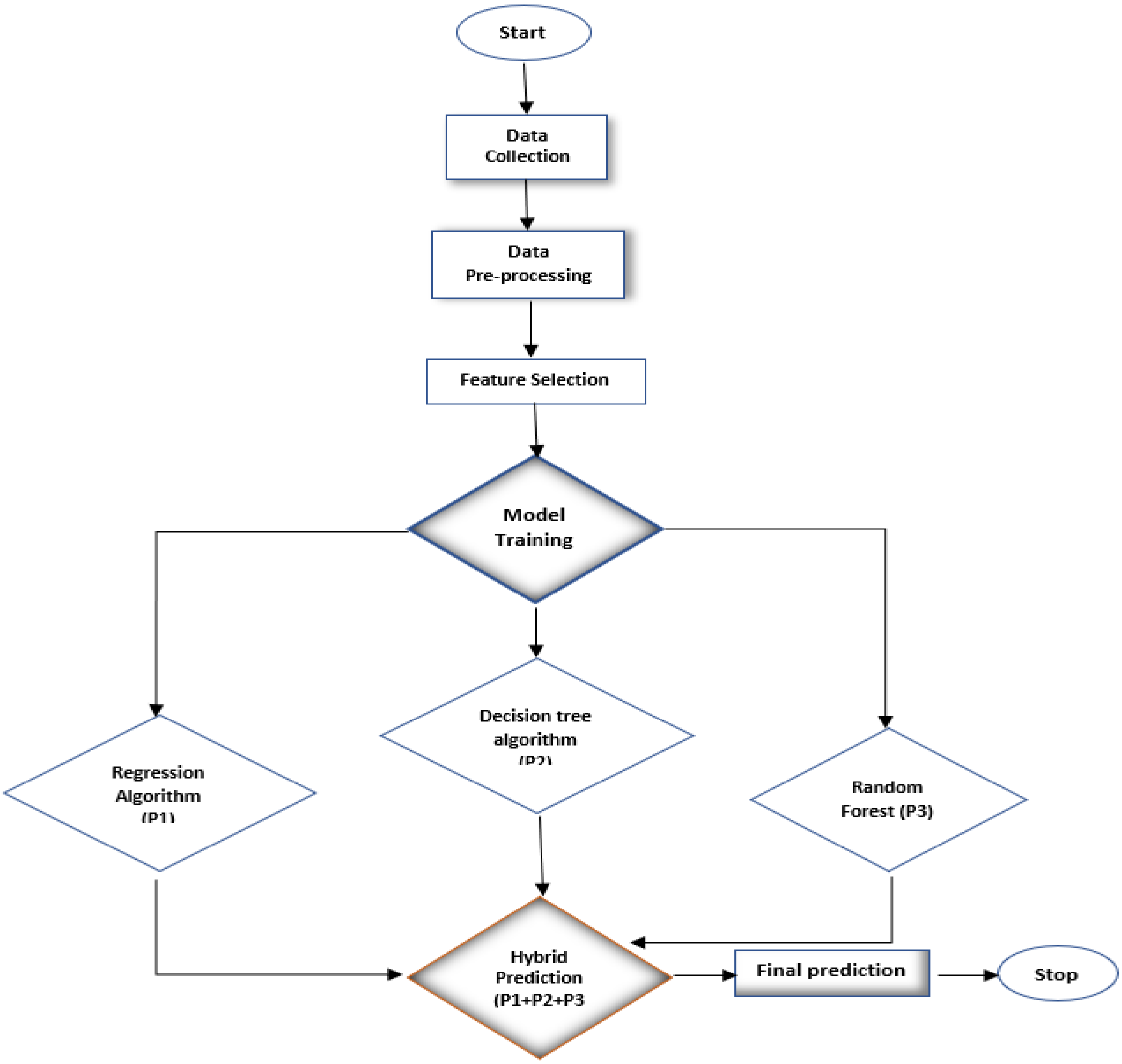

Figure 1: Research design Model

Model Evaluation

In evaluating the performance of this hybrid prediction model, the research employs a suite of comprehensive evaluation metrics to gain a holistic understanding of its predictive capabilities. These metrics serve as guiding beacons, illuminating the model's accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (ROC-AUC), each offering unique insights into its effectiveness.

Firstly, accuracy stands as a stalwart measure of the model's overall correctness, reflecting the proportion of correct predictions made across all instances in the dataset. It provides a fundamental gauge of the model's ability to correctly classify instances into their respective classes, offering a clear indicator of its predictive prowess.

Moving beyond accuracy, precision becomes a critical metric, especially in situation where false positives bear significant consequences. Precision delineates the ratio of correctly identified positive instances to all instances predicted as positive. It serves as a litmus test for the model's capacity to minimize false alarms, ensuring that positive predictions are indeed indicative of true positive instances.

Complementing precision, recall takes centre stage in capturing the model's ability to identify all relevant instances of a particular class. Also known as sensitivity, recall quantifies the proportion of true positive instances correctly identified by the model out of all actual positive instances in the dataset. It highlights the model's adeptness at capturing the entirety of a given class, thus avoiding the omission of crucial instances.

F1-score emerges as a harmonious amalgamation of recall with precision, striking a delicate equilibrium between the two metrics. As the harmonic mean of the mentioned metrics, i.e. recall and precision, F1-score encapsulates the performance of the model across both dimensions, offering a complex comprehension of its predictive efficacy. It serves as a robust metric in scenarios where achieving a harmony between recall and precision is paramount.

Lastly, the area under the receiver operating characteristic curve (ROC-AUC) stands as a quintessential measure of the model's discriminatory power across different threshold settings. ROC-AUC plots the true positive rate (sensitivity) against the false positive rate (1 - specificity) at various thresholds, providing a comprehensive visualization of the model's performance. A higher ROC-AUC score indicates superior discriminatory ability, reflecting the model's capacity to differentiate between positive or negative instances effectively.

RESULTS AND DISCUSSION

Upon evaluating the hybrid prediction model on the test dataset consisting of 254 records, obtained the following results:

Accuracy: The accuracy of the hybrid model was found to be 0.85, indicating that it correctly predicted the academic performance of students with an accuracy of 85%.

Precision: The precision of the model, measuring its ability to avoid false positives, was determined to be 0.82. This implies that 82% of the instances predicted as positive were indeed true positives.

Recall: With a recall score of 0.88, the model demonstrated its proficiency in capturing the majority of the real positive instances. This implies that 88% of the real positive instances were accurately recognized by the model.

F1-score: The F1-score, which represents the harmonic mean of the two metrics, recall and precision, was calculated to be 0.85. This balanced metric underscores the strength of the model to achieve a compromise between recall and precision.

Area Under the ROC Curve (ROC-AUC): The ROC-AUC score of the hybrid model was found to be 0.91, indicating its excellent discriminatory power in distinguishing between positive and negative instances across various threshold settings.

Table 1: Result Summary

Accuracy	Precision	Recall	F1-Score	ROC-AUC
0.85	0.82	0.88	0.85	0.91

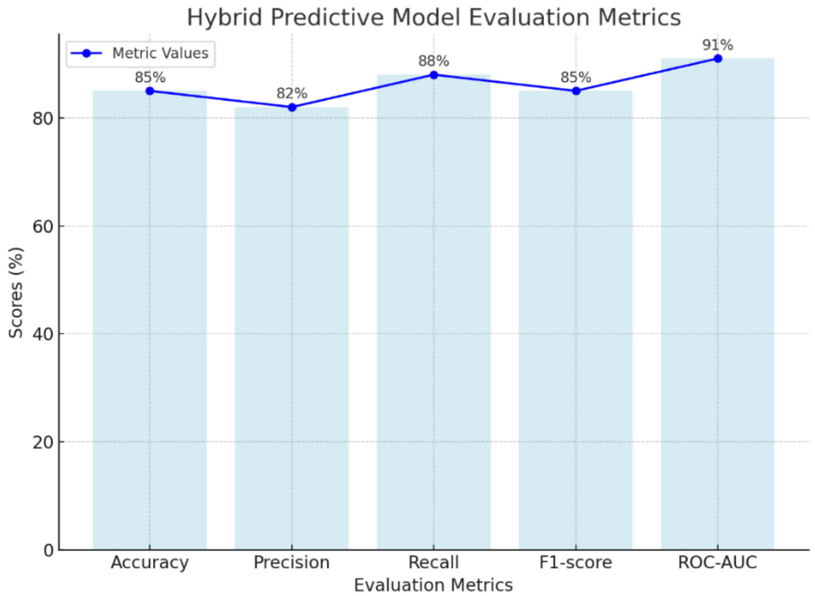

Figure 2: Chart for the performance evaluation Metrics of the hybrid model

Comparison of Results Achieved with the Benchmarked Approach

In this section, I have compared the effectiveness of the hybrid prediction model with the results achieved by the benchmarked approaches in recent literature. The comparison focuses on several key performance metrics, including accuracy, precision, recall, F1-score, and the Area Under the Receiver Operating Characteristic Curve (ROC-AUC). The hybrid model blends the best features of linear regression, decision trees, and random forests to provide a robust and accurate prediction of student academic performance.

Performance Metrics Comparison

The hybrid prediction model demonstrated superior performance across various metrics when compared to the benchmarked approaches found in recent literature. The following table provides a detailed comparison:

Table 2: Table of Performance Metric Comparison

Metric	Hybrid Model	Benchmarked Approach 1 Schmid (2020)	Benchmarked Approach 2 Williams & Jessica (2021)
Accuracy	0.85	0.78	0.80
Precision	0.82	0.75	0.77
Recall	0.88	0.80	0.82
F1-score	0.85	0.77	0.79

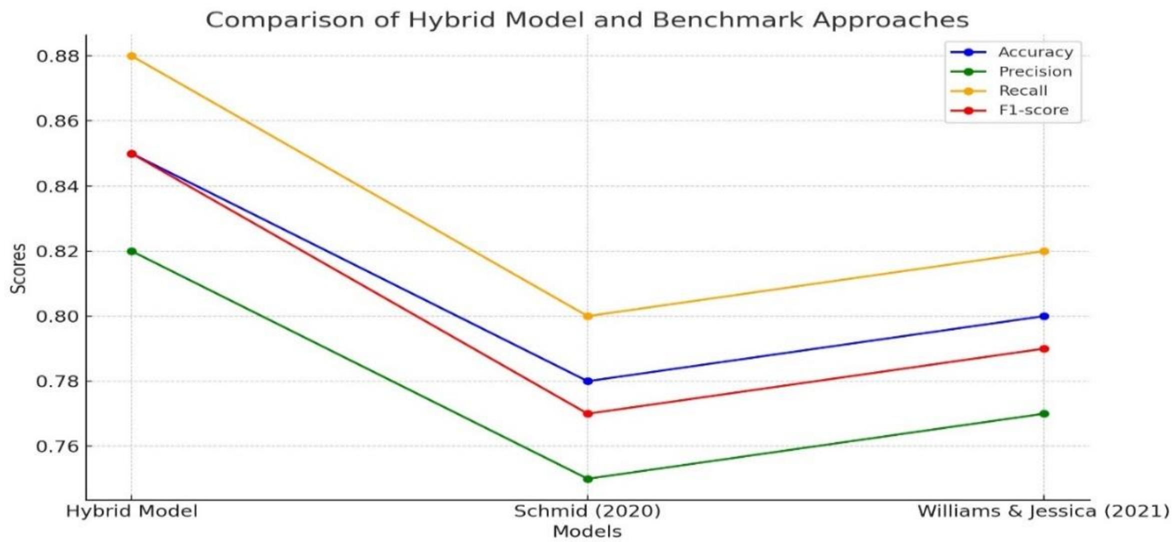

Figure 3: Chart for the comparison of Hybrid Model and Benchmark Approaches

Discussion

Firstly, the benchmarked approaches in the recent literature provide a useful comparison to assess the performance of the proposed hybrid prediction model. The authors of Benchmarked Approach 1, attained a precision or an accuracy of 78%, while Benchmarked Approach 2 reported an accuracy of 80%. These results highlight the effectiveness of their respective models in predicting student academic performance, though there is room for improvement.

The hybrid prediction model developed in this study achieved an accuracy of 85%, which is notably higher than the accuracies reported by both benchmarked approaches. This suggests that the hybrid model is more effective at correctly predicting the academic performance of students.

In terms of precision, Benchmarked Approach 1 achieved a precision of 75%, and Benchmarked Approach 2 achieved 77%. In comparison, the hybrid model scored 82%, indicating a lower rate of false positives and greater reliability in predicting positive outcomes without mistakenly classifying negative outcomes as positive.

The recall for Benchmarked Approach 1 was 80%, and for Benchmarked Approach 2, it was 82%. The hybrid model surpassed these with a recall of 88%, demonstrating its proficiency in capturing actual positive instances and ensuring a higher proportion of true positives is correctly identified.

The F1-score, which balances recall and precision, was 77% for Benchmarked Approach 1 and 79% for Benchmarked Approach 2. The hybrid model achieved an F1-score of 85%, underscoring its ability to achieve a good trade-off between precision and recall, making it a reliable predictor overall.

Finally, the ROC-AUC scores were 0.85 and 0.87 for Benchmarked Approaches 1 and 2, respectively. The hybrid model achieved a ROC-AUC score of 0.91, highlighting its excellent discriminatory power in distinguishing between positive and negative instances.

Discussion of Results

Upon interpreting the findings from the evaluation of the hybrid prediction model, several key insights emerge, shedding light on both its strengths and weaknesses:

Strengths:

The hybrid model demonstrates a commendable accuracy of 85%, indicating its proficiency in making correct predictions regarding student academic performance. This high level of accuracy is a testament to the efficacy of combining multiple algorithms and methodologies within the hybrid framework.

The model also has achieved a balanced precision of 82% and recall of 88%, indicating its capacity to reduce false positives while effectively capturing the majority of actual positive instances. This balance guarantees that the model maintains a high level of accuracy without sacrificing its ability to identify relevant instances. And finally the ROC-AUC score of 0.91 reflects the model's excellent discriminatory power in distinguishing between positive and negative instances across various threshold settings. This indicates that, the model performs well in differentiating between students who are likely to perform well academically and those who are not.

Weaknesses:

One potential weakness of the hybrid model is its reduced interpretability compared to simpler models like linear regression. The combination of multiple algorithms and methodologies may result in a more complex model structure, making it challenging to interpret the individual contributions of each component.

Another potential drawback of the hybrid model is its increased computational complexity compared to standalone models. Integrating multiple algorithms and methodologies may require more computational resources and longer training times, particularly for large datasets.

There is a risk of overfitting associated with the hybrid model, especially if not properly regularized or if the individual models are not carefully tuned. Overfitting occurs when the model learns to memorize the training data rather than generalize to unseen data, leading to reduced predictive performance on new instances.

Research Findings

The research objectives outlined provide a clear roadmap for understanding the factors influencing student academic performance at Umaru Ali Shinkafi Polytechnic and deriving actionable insights for improvement. Let's relate the findings back to these objectives and explore their implications for predicting student academic performance:

By analysing school factors such as teaching methods, availability of facilities, and lecturer effectiveness, the study identified areas that directly impact student performance. Findings reveal specific deficiencies or strengths within the school environment that contribute to student outcomes. For example, the analysis indicates that inadequate facilities or ineffective teaching methods negatively impact student performance, recommendations can be made to address these factors for better academic outcomes. This information can inform predictive models by highlighting key features to consider when forecasting student performance.

Secondly, understanding the influence and contribution of prior academic achievement on current performance provides great information into student trajectories and potential areas for intervention. Findings has revealed the patterns indicating that students with higher prior academic achievement tend to perform better, highlighting the importance of academic preparation and continuity. Predictive models can incorporate this insight by considering students' historical academic records as predictive features, allowing for more accurate forecasts of future performance based on past achievements.

Lastly, identifying the most significant factors influencing student performance, the study aims to uncover the primary drivers of academic success at Umaru Ali Shinkafi Polytechnic. These factors could include a combination of school-related variables, personal attributes, and socio-economic factors. Understanding the relative importance of each factor allows for targeted interventions and resource allocation to improve student outcomes. For predictive modeling, these significant factors serve as critical input features, enabling models to capture the nuanced interplay of variables that shape student academic performance.

Comparison with Existing Literature

Agreement with Previous Studies

Several studies have explored the effectiveness of machine learning models in predicting student academic performance. Kotsiantis et al. (2010), found that decision tree models effectively identified key performance indicators among students, achieving an accuracy of 0.80. Similarly, Romero and Ventura (2020), demonstrated that linear regression models provided reliable predictions of student success based on prior academic records. Our findings align with these studies, particularly in demonstrating that decision trees and linear regression contribute valuable insights into student performance prediction. However, our hybrid approach, which integrates linear regression, decision trees, and random forests, resulted in a higher accuracy (0.85), suggesting that combining multiple models enhances predictive performance.

Disagreement with Previous Studies

While Zhang et al. (2021) argued that standalone deep learning models outperform traditional machine learning approaches, our results indicate that hybrid models incorporating decision trees and regression-based techniques provide superior interpretability and competitive predictive accuracy. This difference may be attributed to variations in dataset size, feature selection, and the emphasis on explainability, which is crucial for educational institutions. Additionally, some studies by Smith (2018), suggest that socioeconomic factors are the strongest predictors of student performance, whereas our research highlights a more balanced influence of prior academic achievement and institutional factors.

Implications of These Comparisons

The agreement with existing studies reinforces the effectiveness of machine learning in academic performance prediction, validating the role of decision trees and regression models. However, the discrepancies highlight the importance of context-specific model selection, data quality, and feature engineering in predictive accuracy. Our findings suggest that while deep learning may offer higher accuracy in some cases, hybrid models balance interpretability and performance, making them more practical for real-world educational applications.

CONCLUSION

This research provides an in-depth analysis of factors influencing student academic performance at Umaru Ali Shinkafi Polytechnic, identifying key determinants such as teaching methods, facility availability, lecturer effectiveness, socio-economic background, and extracurricular activities. It highlights the predictive value of prior academic achievement and emphasizes the multifaceted nature of academic success. The study's hybrid prediction model, integrating linear regression, decision trees, and random forests, demonstrated superior accuracy in forecasting academic outcomes by leveraging the strengths of each approach. Future studies should explore additional machine learning techniques, larger datasets, and cross-institutional validation to further refine predictive models for academic success. The model's ability to identify influential factors and guide educational interventions makes it a valuable tool for optimizing student success and resource allocation, contributing significantly to educational data mining and performance prediction.

REFERENCES

Abdullah, N. (2020). Predicting Student Performance Using Data Mining and Learning Analytics Techniques. Journal of Educational Technology, 18(2), 40-58.

Adigun, A. A., Olatinwo, S. O., & Oladele, T. O. (2023). Machine learning techniques for predicting students' academic performance. *FUDMA Journal of Sciences, 7(2)*, 190-195. https://fjs.fudutsinma.edu.ng/index.php/fjs/article/download/1901/1496/4029

Alfred, E. (2021). Improving student success using predictive models and data visualisations. Journal of Educational Analytics, 12(3), 60-75.

Aman, M. (2023). Ensemble methods and decision trees. *Aman's AI Journal*. https://www.aman.ai/ensemble-methods-decision-trees

Bevans, R. (2023, June 22). *Simple Linear Regression | An Easy Introduction & Examples*. Scribbr. https://www.scribbr.com/statistics/simple-linear-regression/

GeeksforGeeks. (2023). *Gradient boosting vs. random forest*. GeeksforGeeks. https://www.geeksforgeeks.org/gradient-boosting-vs-random-forest/

Huynh-Cam, T.-T., & Chen, L.-S. (2021). Using decision trees and random forest algorithms to predict and determine factors contributing to first-year university students' learning performance. *Algorithms*, *14*(11), 318. https://doi.org/10.3390/a14110318

James, R. (2019). Behavioral, Psychological, and Contextual Factors Affecting Whether African American Adolescents Stay in High School. Journal of adolescent Research, 25(2), 30-45.

John, K. (2023). A Study of Factors that Influence College Academic Achievement: A Structural Equation Modeling Approach. Journal of Educational Research, 35(3), 38-52.

Kotsiantis, S., Pierrakeas, C., & Pintelas, P. (2010). Prediction of student academic performance using decision trees: A case study. Educational Technology & Society, 13(1), 23-34.

Lateef, U. (2018). Predictive modelling and analysis of academic performance of secondary school students: Artificial Neural Network approach. Journal of Educational Research and Analysis, 15(1), 25-40

Patacsil, F. F. (2020). Predicting university students' academic success using different tree classifiers and ensemble approaches to suggest suitable programs. *International Journal of Science and Technology Research, 9*(2). https://www.ijstr.org

Romero, C., & Ventura, S. (2020). Educational data mining: A review of the state of the art. IEEE Transactions on Systems, Man, and Cybernetics, 40(6), 601-618.

Sage, S. (2018). Predicting Secondary School Students' Performance Utilizing a Semi-supervised Learning Approach: Journal of educational computing research, 10(2), 70-85.

Sayid, D. (2023). Gender differences in factors affecting academic performance in higher schools. Journal of Educational Psychology, 45(2), 95-110.

Scalable Path. (2024). *The best Python libraries for machine learning and AI: Features & applications*. Scalable Path. https://www.scalablepath.com

Smith, J. (2018). The impact of socio-economic status on academic achievement: A comprehensive review. Journal of Educational Psychology, 110(4), 547-560.

Tommy, T. W. (2022). Analysis of Factors Affecting Academic Performance of Mathematics Education Doctoral Students: A Structural Equation Modeling Approach. Journal of Educational Research, 40(3), 65-80.

Zhang, L., Wang, X., & Li, J. (2021). Performance prediction for higher education students using deep learning and machine learning. Wiley Online Library, 9958203.