CAMBRIDGE
UNIVERSITY PRESS

RESEARCH ARTICLE

Enhancing student performance in African smart cities: a web-based approach through advanced ensemble modeling and genetic feature optimization

Hayat Sahlaoui[1] , El Arbi Abdellaoui Alaoui[2], Abdelaaziz Hessane[1], Said Agoujil[3], Stéphane Cédric Koumetio Tekouabou[4] and Abdellah Barebzi[5]

[1]Department of Computer Science, Faculty of Sciences and Techniques at Errachidia, University of Moulay Ismail, Errachidia 52000, Morocco
[2]IEVIA Team, IMAGE Laboratory, Department of Sciences, Ecole Normale Supérieure, University of Moulay Ismail, Meknes, Morocco
[3]Ecole Nationale de Commerce et de Gestion, Moulay Ismail, University of Meknes, El Hajeb 51000, Morocco
[4]Center of Urban Systems (CUS), Mohammed VI Polytechnic University (UM6P), Ben Guerir, Morocco
[5]Research Team "Education and Cognitive Neuroscience", Ecole Normale Supérieure, University of Moulay Ismail, Meknes, Morocco
Corresponding author: Hayat Sahlaoui; Email: ha.sahlaoui@edu.umi.ac.ma

Abstract

In the burgeoning landscape of African smart cities, education stands as a cornerstone for sustainable development and unlocking future potential. Accurate student performance prediction holds immense social importance, enabling early intervention, improved learning outcomes, and equitable access to quality education, aligning with sustainable development goals. Traditional models often falter in Africa due to imbalanced datasets and irrelevant features. This research leverages machine learning in Nigerian classrooms to predict underperforming students. Techniques like synthetic minority oversampling, edited nearest neighbors, and the Boruta algorithm for feature selection, alongside genetic algorithms for efficiency, enhance model performance. The ensemble models achieve AUCs of 90–99.7%, effectively separating low-performing and high-performing students. Implemented via Streamlit and Heroku, these models support real-time, data-driven decisions, enhancing early intervention, personalized learning, and informing policy and public service design. This ensures equitable education and a brighter future across Africa. By leveraging ML, this research empowers universities to support struggling students, optimize educational costs, and promote inclusive development, fostering data-driven decision-making and resource allocation optimization. Ultimately, this research paves the way for a future where data empowers education within African smart cities, unlocking the full potential of data-driven solutions and ensuring equitable educational opportunities across the continent.

Policy Significance Statement

The research explores machine learning techniques in Nigerian classrooms, creating predictive models to identify underperforming students. This aids universities globally in supporting struggling students, optimizing educational costs, and fostering inclusive development. Leveraging the synthetic minority oversampling and edited nearest neighbors (SMOTE–ENN) for imbalanced data, Boruta for feature selection, and genetic algorithms to enhance efficiency. Beyond technical advancements, the research translates findings into real-

world impact via a Streamlit and Heroku web app. This empowers universities worldwide to customize models, view real-time results, and make informed, data-driven decisions. Ultimately, this work empowers stakeholders to utilize data science for equitable and impactful education in African smart cities, fostering sustainable development.

1. Introduction

In the vibrant landscape of African smart cities, educatôn stands as a cornerstone for fostering sustainable development and unlocking the future potential of its youngest citizens. Accurately predicting student performance becomes crucial in this context, allowing for early intervention, improved learning outcomes, and ultimately, equitable access to quality education, aligning with the core principles of the sustainable development goal. However, this task faces unique challenges in the African context. Educational data often exhibit unpredictability, noise, and imbalanced class distributions, presenting significant hurdles for traditional prediction models (Ghorbani and Ghousi, 2020). These models can easily falter, favoring the majority class and leading to misleading outcomes (Fernández et al., 2011; Phua and Batcha, 2020). This is where machine learning (ML) emerges as a powerful tool for informed decision-making and improved learning experiences, empowering both educators and policymakers. One key application of ML lies in predicting student academic performance (Shaukat et al., 2016; 2017). However, imbalanced datasets, a common issue in education, hinder the effectiveness of such models. Furthermore, the complexity of this issue deepens when we consider the vast dimensions of educational datasets (Sammut and Webb, 2011; Yin et al., 2013) and their inherent class imbalances (Friedman, 1997; Zhang and Hu, 2014). Predicting student performance under such conditions becomes an intricate task. Moreover, if the data are cluttered with irrelevant features, it hampers ML models, resulting in suboptimal performance and heightened computational costs (Arif, 2020). In such scenarios, it becomes imperative to weed out insignificant attributes from the data, paving the way for more efficient and cost-effective ML processes. Feature selection methods, a suite of techniques designed to identify and retain only the most valuable features within a dataset, emerge as a solution to this challenge (Chandrashekar and Sahin, 2014). These methods fall into three distinct categories: filtering, wrapping, and embedding. Filtering algorithms select a subset of features based on criteria such as correlation or information entropy, offering a rank-ordered set of the most relevant attributes (Spolaôr et al., 2013). Wrapper procedures employ ML techniques to evaluate the quality of feature subsets. Embedded approaches, on the other hand, directly enhance ML algorithms' accuracy for predicting student performance. Feature selection has proven its worth in a multitude of applications, from gene array analysis to intrusion detection and text mining (Bharti and Singh, 2014). In the context of predicting student achievement, researchers harness these techniques to create more effective subsets of characteristics (Dhar and Jodder, 2020; Sobnath et al., 2020). Among these strategies, genetic algorithms shine as a popular tool to optimize feature selection for enhanced predictive accuracy (Turabieh, 2019; Yousafzai et al., 2020). At the heart of this endeavor is the deployment of dependable classification models. Education, as a sector, can harness the full potential of ML by translating these models into real-world decisions as swiftly as possible. This empowers universities and stakeholders to harness data science for equitable and impactful education in African smart cities.

1.1. Research questions

Question 1: How can hybrid sampling techniques be used to address imbalances in the data, ensuring all student groups are fairly represented and providing more reliable predictions?

Question 2: How effective are wrapper-based feature selection methods in identifying key factors impacting student performance, allowing efficient resource allocation, and reducing computational costs?

Question 3: Which of the two feature selection algorithms experimented with, the Boruta and the genetic algorithms, is better for feature selection in student performance models using the Friedman test as a statistical validity check?

Question 4: To what extent can these suggested models be deployed to improve their real-world accessibility? How to deploy models in the form of web applications to empower universities within and outside Africa To personalize models and view real-time results, optimizing student support through data-driven insights, and to offer accessible tools for policymakers to design effective and legitimate education policies aligned with the SDG, informed by accurate predictions and data analysis?

1.2. Objective

This study aims to leverage ML to enhance the prediction of student performance and optimize educational decision-making by:

- Mitigating class imbalance biases within student data to ensure fair and accurate representation of all student groups.
- Employing wrapper-based feature selection methods to identify the most critical factors influencing student academic outcomes.
- Validating the efficacy of different feature selection techniques using the Friedman test to determine the most statistically significant approach.
- Deploying the resulting ensemble model as an interactive web application accessible to universities both within and outside Africa. This application will empower institutions to personalize student support strategies and enable policymakers to design data-driven and SDG-aligned education policies informed by accurate predictions.

1.3. Organization of the paper

The remainder of this work is organized as follows: The related works are introduced in Section 2. In Section 3, we discuss the research methodology, and in Section 4, we present the experimental results and a discussion of the performance of different crafted strategies and the deployment methods. Section 5 is dedicated to our conclusions, while Section 6 offers insights into limitations and suggestions for future work.

2. Related work

This part reviews the pertinent literature and current state of the art techniques in the area of ML, specifically addressing student performance classification, the class imbalance problem, and the features dimensionality reduction problem. ML model deployment.

2.1. Student performance classification with ensemble model

Enhancing prediction accuracy poses a challenging endeavor. Because of ensemble modeling, which includes well-known algorithms like arcing (Breiman, 1998), boosting (Freund and Schapire, 1997), bagging (Breiman, 1996), and random forests (Breiman, 2001), this target is within reach. They may be divided into two broad categories: homogeneous and heterogeneous. In contrast to homogeneous approaches, which mix algorithms of a similar sort, heterogeneous approaches integrate algorithms of varying types, such as classification trees, support vector machines, and neural networks. Stacking is by far the most common (Wolpert, 1992).

Student performance predictions were enhanced by using the ensemble technique, as described in his study (Almasri et al., 2019). According to the findings, the application of ensemble approaches. It helps improve the model's accuracy, which achieves 98.5%. The suggested method, based on bagging and boosting ensemble-based techniques, achieved an accuracy of over 80%, validating its efficacy (Amrieh

et al., 2016). By combining many categorization methods into a single model, Ajibade et al. (2020) were able to double the accuracy of their predictions based on student data.

In their study (Onan et al., 2016), they evaluated five statistical keyword extraction methods in scientific text document classification, finding that the most-frequent-based method, when combined with the bagging ensemble of Random Forest, achieves the highest predictive performance, thereby enhancing text classification scalability. Onan (2019) also introduces a two-stage framework, combining an improved word embedding scheme and an enhanced clustering ensemble approach to extract significant topics from scientific literature. This innovation leads to improvements in predictive performance and clustering results. In their research (Onan et al., 2017), they further present a hybrid ensemble pruning scheme that significantly enhances sentiment classification in text documents, outperforming existing ensemble methods and pruning algorithms.

2.2. Class imbalance problem

Real-world datasets often exhibit imbalanced class distributions, impacting ML models that perform best with roughly equal class examples (Longadge and Dongre, 2013). Imbalances lead to biased classifiers favoring dominant groups, resulting in inconsistent results (Kotsiantis et al., 2006). Various strategies address imbalanced data using over-sampling (Yap et al., 2013), under-sampling (Liu et al., 2008), and hybrid-sampling (Galar et al., 2011) methods. Oversampling techniques like random over-sampler (Li et al., 2013), synthetic minority oversampling (SMOTE; Chawla et al., 2002), borderline SMOTE (Han et al., 2005), and Support Vector Machine (SVM)-SMOTE (Tang et al., 2008) counteract imbalances. Common undersampling methods, such as edited nearest neighbors (ENN) and TOMEK-links (Guan et al., 2009), reduce overrepresented classes. Hybrid methods, like SMOTE–ENN and SMOTE–TOMEK (Batista et al., 2004), combine oversampling and undersampling strategies.

In their study (Chaudhury et al., 2016), they examine the influence of data preprocessing on classification algorithms using a dataset of student performance with unequal classroom distribution. Undersampling and oversampling methods are used with support vector machines, decision trees, and Naive Bayes classification algorithms, respectively, to rectify the issue of imbalanced class distribution. For example, SMOTE, an oversampling-type method (Chaudhury et al., 2016), performs better in experiments. Using the open-source Weka program, Salal et al. (2019) evaluate the performance of several categorization methods on a dataset of student performance. Experiments showed that certain factors have a greater impact on student performance during the data preprocessing stage, whereas classification algorithm accuracy improves following this step (Mangina and Psyrra, 2021). In their research (Sahlaoui et al., 2021), they applied a kernel shape value to the problem of categorizing students' academic achievements. They contributed by suggesting a three-layer architecture built on top of the XGBoost concept and SMOTE technologies. As a result, the model's efficiency and interpretability will increase while the load of an imbalanced dataset is reduced. This cutting-edge approach improves classifiers' predicting accuracy and interpretability by using their strengths through boosting, kernel shape value, and SMOTE procedures. To improve their student performance prediction approach, Costa et al. (2017) used a data preprocessing and fine-tuning strategy. Superior results may be achieved using the proposed approach with an SVM classifier. Dimensional reduction in the form of the Information Gain method and data balancing using the SMOTE algorithm constitute part of the preprocessing technique. In addition, the grid search technique is utilized to find the hyper-parameter's optimal values for the SVM algorithm.

In his study, Onan (2023) presents Genetic Algorithm-based Global Trajectory Reconstruction (GTR-GA), an approach that combines graph-based neural networks and genetic algorithms to create diverse, high-quality augmented text data, addressing data scarcity in natural language processing and enhancing downstream performance. Onan et al. (2019) also tackle class imbalance in ML with a consensus clustering-based undersampling approach, enhancing predictive performance across various imbalanced classification benchmarks by incorporating diverse clustering and classification methods.

2.3. *Features dimensionality reduction problem*

Optimizing ML involves crucial feature selection to enhance accuracy and understand variables influencing students' academic success. This process strategically selects a subset of input features (Bhat, 2019), categorized into filtering, wrapping, and embedded methods. Filtering methods, like information gain and the chi-square test, offer computational efficiency. Wrapping approaches delegate feature selection to the classifier, incurring higher costs for heightened accuracy (Paddalwar et al., 2022). Embedded selection integrates the process, extracting significant features during each training iteration. Well-recognized embedded approaches, including random forest, decision tree, and LASSO feature selection (Butvinik, 2021), collectively optimize model performance.

Success in mortality prediction was shown by subsequent studies (Moreno García et al., 2014; Liu et al., 2018), which addressed dimensionality reduction and unbalanced class issues. Leppänen et al. (2017) use linear regression, decision trees, and Naïve Bayes techniques to estimate the student performance on two separate student datasets. The feature selection method improves the accuracy values of the classification algorithms for both datasets, as shown by the experimental findings (Leppänen et al., 2017). In his study (Ünal, 2020), he used a decision tree-based feature selection technique to assess the efficacy of several classification algorithms on the student dataset. As stated by Ünal (2020), the best accuracy values are obtained using the random forest technique. In his research (Zaffar et al., 2020), he used three different student performance datasets to determine the impact of the quick correlation-based filtering strategy on the performance of the support vector machine classifier. The feature selection technique enhances the student academic performance prediction model, as concluded by experimental data (Zaffar et al., 2020). In order to construct a performance prediction model for students, researchers employ filter-based or wrapper-based feature selection techniques. The vast majority of studies use correlation-based filtering techniques (Dhar and Jodder, 2020; Sobnath et al., 2020), which evaluate the degree to which a trait is associated with a subject's performance. Filtering techniques based on information gained evaluate each characteristic and the student's performance to determine which one is more informative (Amrieh et al., 2016; Garcia and Skrita, 2019). Wrapper approaches based on gene algorithms have been utilized to tackle global optimization issues, and gene-based algorithms are widely employed in this context. To simulate biological evolution's natural selection process, Wutzl et al. (2019) used a genetic-based algorithm (GA), and they iteratively fine-tuned the feature selection approach by analyzing past prediction outcomes. Similar to natural evolution, Yousafzai et al. (2020) and Turabieh (2019) employed a binary GA to constantly improve feature selection via the application of the three operations of selection, crossover, and mutation on the Feature Selection Scheme. Begum and Padmannavar (2022), in their research, have checked whether the database features are enough to develop successful models for forecasting student performance and analyze genetic algorithm-based feature selection to guarantee redundancy management and appropriate training in supervised classification. In his research, Anand et al. (2021) use academic data to demonstrate an FS strategy using the Boruta algorithm. This research makes an effort to enhance the classifier performance of MAIT's undergraduate students.

In his study (Onan, 2022), he introduces a state-of-the-art bidirectional convolutional recurrent neural network for sentiment analysis, effectively processing sequences, reducing feature dimensionality, and emphasizing essential features. Additionally, Onan and Korukoğlu (2017) propose an ensemble feature selection method using a genetic algorithm that aggregates feature lists from various methods, significantly improving efficiency and classification accuracy in sentiment analysis. In the context of text genre analysis, Onan (2018) developed an ensemble classification scheme with a high average predictive performance of 94.43%. This approach combines the random subspace ensemble of a random forest with multiple feature types to achieve outstanding results. Furthermore, Kou et al. (2021) introduce a novel bankruptcy prediction model for small and medium-sized enterprises (SMEs) using transactional data. This model incorporates a two-stage feature-selection method, reducing the number of features while preserving classification performance. Finally, in their study (Kou et al., 2020), they employed multiple criteria decision-making (MCDM) methods to comprehensively evaluate feature selection techniques in

the domain of text classification with small sample datasets, offering valuable recommendations to enhance this area of research.

2.4. Machine learning model deployment

Model deployment is the crucial final phase of a ML project, enabling live use of business insights. Integration into web or mobile applications using cloud services facilitates swift identification of at-risk students by teachers for timely intervention (Prakash, 2021). Understanding both frontend (HTML, CSS, and JavaScript) and backend frameworks (Linux administration, networking, and security) is crucial for web program creation (Kim, 2021). Flask, an open-source Python framework, suits modern web applications, while Django, more advanced, requires additional mastery time. TensorFlow.js excels at storing web-compatible models but has limitations in handling complex ML implementations (Vadapalli, 2020; Buvaneshwaran, 2022). Streamlit, an open-source platform, empowers data scientists and MLOps developers to craft engaging interfaces for datasets. As a Python-based frontend framework, Streamlit facilitates swift development of ML web applications without extensive frontend development knowledge (Kimnaruk, 2022). Serving as a free model deployment solution, Streamlit streamlines the process, allowing rapid implementation without expertise in front-end development or other programming languages (Flynn, 2021; Thetechwriters, 2021). Unlike Flask and Express, Streamlit seamlessly integrates these functions (Kimnaruk, 2022). While deploying Streamlit applications on cloud services like Amazon, GCD, or Azure may seem daunting (Sharm, 2020), Heroku offers a user-friendly alternative. This free online app hosting service simplifies and accelerates hosting cloud-based apps in any language. Operating as a Platform as a Service (PaaS), Heroku facilitates deploying and starting an app in under 5 min, with system management handled by Heroku, freeing developers from the complexities of setting up their server infrastructure (Melo, 2023).

In his study (Dempere Guillermo, 2018), he has built a full-featured mentoring platform that is integrated with the results of ML algorithms, and this effort is focused on the deployment of the student performance model. The authors create a Mentor Academic Guidance Support Platform based on predictive data-driven dashboards. The "APPA" mobile app developed by the researcher makes use of the benefits offered by the React-Native platform in order to provide answers to problems associated with predicting students' performance (Moharekar and Pol, 2021).

2.5. Research contribution and significance

While previous studies have explored feature selection to enhance student performance models, this research explores the use of ML techniques within the unique environment of Nigerian classrooms, focusing on developing predictive models to identify underperforming students. This empowers universities both within and outside Africa to proactively support struggling students, optimize educational costs, and promote inclusive development. We move beyond previous studies by tackling the intricacies of predicting student performance in demanding scenarios, particularly within small and imbalanced datasets. We do this by leveraging a unique combination of techniques: SMOTE–ENN resampling methods address imbalances in the data, ensuring all student groups are fairly represented, and feature selection through Boruta and the genetic Algorithm: To identify and utilize only the most relevant factors impacting academic achievement, reducing model complexity and enhancing generalizability. This approach significantly boosts prediction accuracy, offering valuable insights for universities within and outside Africa. Our research transcends technical advancements by translating these findings into real-world impact through an accessible web application built with Streamlit and Heroku. This empowers universities within and outside Africa to customize model settings and view real-time results, fostering informed decision-making based on data-driven insights and predictions. Moreover, this research demonstrates how data science can inform policy by improving situation analysis through a deeper understanding of factors impacting student performance, enhancing predictions for timely interventions and resource allocation, and supporting public service design for effective

and legitimate education policies. Ultimately, this research empowers stakeholders to harness data science for equitable and impactful education in African smart cities, paving the way for a future where data empower education, fosters sustainable development, and unlocks the potential of every student in African smart cities.

3. Research methodology

In this section, we describe our proposed approach, explaining the workflow adopted in this research. We detail the various approaches to testing the dataset, including the sampling methods, wrapper-based feature selection algorithms, and learning classifiers utilized. Additionally, we discuss the criteria used to evaluate these techniques during the learning and training process.

3.1. Proposed system

To assess the efficacy of ML in predicting Nigerian classroom outcomes, this study adopts a multi-step methodology. First, data preprocessing will address potential class imbalances using the SMOTE–ENN hybrid sampling method. Next, wrapper-based feature selection techniques, Boruta and the genetic algorithm, will be employed to identify the most relevant features and potentially enhance model performance through dimensionality reduction. Finally, model development and evaluation will involve employing a diverse range of ML classifiers with varying theoretical foundations. The selection of the optimal model will be based on established metrics like accuracy, precision, recall, F1-score, and area under the curve (AUC). Tenfold stratified cross-validation will ensure robust evaluation, and hyper-parameter tuning will further refine the chosen model's performance. The workflow for this investigation is elucidated in Figure 1.

3.2. Data preprocessing

3.2.1. Dataset information

Accessing educational data sources for educational data mining (EDM) research can pose challenges. To overcome this, we utilized a publicly available dataset from Nigerian educational institutions (https://www.kaggle.com/c/datasciencenigeria/data). This dataset comprises 650 students, each characterized by 23 different elements. The study focuses on the final grade point average (GPA) as the outcome variable, categorizing students into poor, medium, and good based on their grade-point average. Table 1 summarizes the dataset, including various psychological, individual, and contextual variables such as student attendance, travel time to school, reading time, educational assistance, health, and parental education level. These elements collectively contribute to understanding and analyzing factors influencing student performance, addressing the multi-categorization challenge posed by the study.
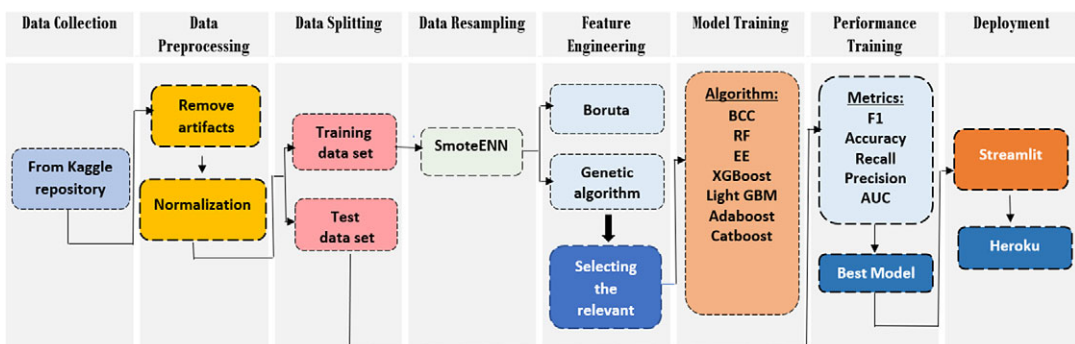

Figure 1. *Global ensemble-based system for the classification of students' academic performance.*

Table 1. *Explanation of the Nigeria dataset variables used to forecast student outcomes*

Feature	Description	Feature type
Age	Student's age from 10 to 17.	Numerical
Gender	The student gender ("Male" or "Female").	Nominal
Location	Student's home address type urban or rural.	Nominal
Famsize	Family size: "LE3"—less or equal to 3 or "GT3"—greater than 3.	Nominal
Pstatus	Parent's cohabitation status Living Together "T" or Apart "A."	Nominal
Medu	Mother's education (from G–01 to G–12).	Numerical
Fedu	Father's education.	Nominal
Semester	School year semester as (first or second).	Nominal
Traveltime	Home to school travel time.	Numerical
Studytime	Weekly study time.	Numerical
Failures	Number of past class failures.	Numerical
Schoolsup	Extra educational support.	Nominal
Famsup	Educational support from family.	Nominal
Paid	Extra paid classes.	Nominal
Activities	Extra–curricular activities.	Nominal
Nursery	Attended nursery school.	Nominal
Higher	Desire for higher education.	Nominal
Internet	Internet access at home.	Nominal
Famrel	The quality of family relationships.	Nominal
Freetime	Free time after school	Numerical
Health	health status of student	Numerical
Absences	Number of school absences	Numerical
Score	Score of the student	Numerical

3.2.2. Data sampling

Real-world datasets often exhibit class imbalances, impacting model performance data (Longadge and Dongre, 2013) and yielding unpredictable results (Kotsiantis et al., 2006). This study addresses this by applying SMOTE–ENN, a hybrid sampling method that combines SMOTE (generating synthetic minority class data) and ENN for data cleaning. This approach aims to create a more balanced class distribution within the Nigerian student performance dataset, which is heavily skewed toward the low-performance class (73.8%). The remaining classes are underrepresented, with medium performance at 15.1% and high performance at 11.1%, as depicted in Figure 2.

Utilizing SMOTE–ENN aims to mitigate this imbalance and ensure the model is trained on a more representative distribution of student performance outcomes, as shown in Figure 3.

3.2.3. Feature selection

Feature selection refers to the procedure of picking out and keeping just the most relevant features of a model (Bhat, 2019). In this study, we employ two wrapper-based feature selection methods, namely Boruta and the genetic algorithm, with the aim of enhancing the precision of predicted student performance. This section delves into the detailed explanation of these feature selection techniques, shedding light on their way of optimizing model performance by strategically choosing the most influential features. However, before applying these techniques, it's crucial to assess potential redundancy within the features themselves. We assessed feature redundancy using Pearson correlation to identify highly correlated features. The analysis revealed no strong correlations, as depicted in Figure 4's heatmap, suggesting a
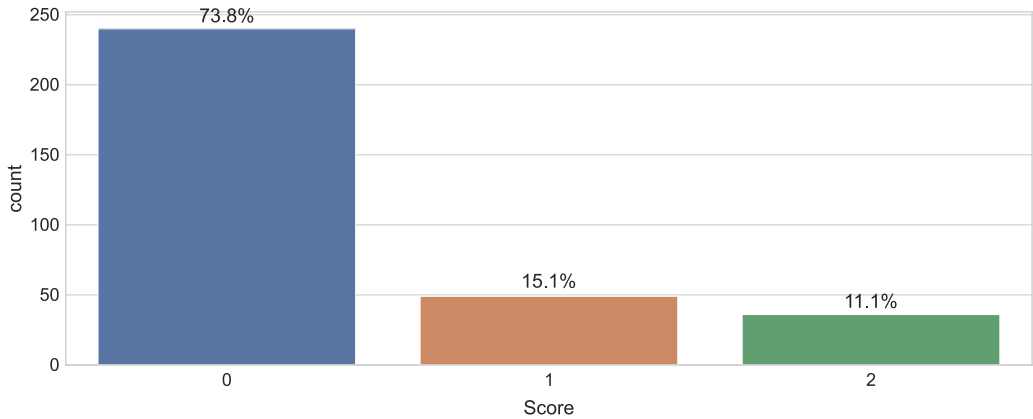
Figure 2. *Distribution of student performance in relation to the different classes.*
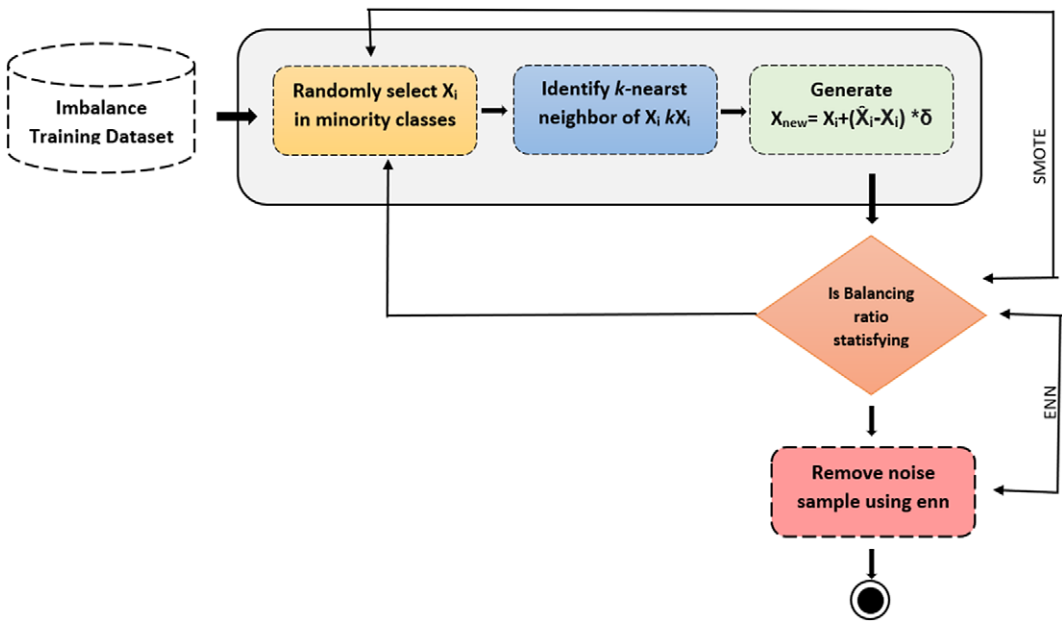

Figure 3. *Schematic flow of SMOTE–ENN.*

minimal risk of multicollinearity. This allows us to proceed with wrapper-based feature selection methods (Boruta and genetic algorithm) without additional preprocessing for redundancy.

• *Genetic algorithm*

A genetic algorithm is a method for resolving optimization issues by utilizing processes like mutation, crossover, and selection (kindsonthegenius, 2018), as shown in Figure 5.

Starting with a large population, it uses heuristics (Algorithm 1) to narrow it down until only the best candidates remain. The process begins with a group of people, or the population, whose characteristics are determined by factors called genes. An individual's fitness level is measured by a score based on the fitness function, where higher scores indicate more fitness. As a way to decide which member of the current generation will be used to produce the next, "selection" is a concept worth considering. In order to create a new generation, two people are chosen to act as "parents." Crossover is a technique used to create
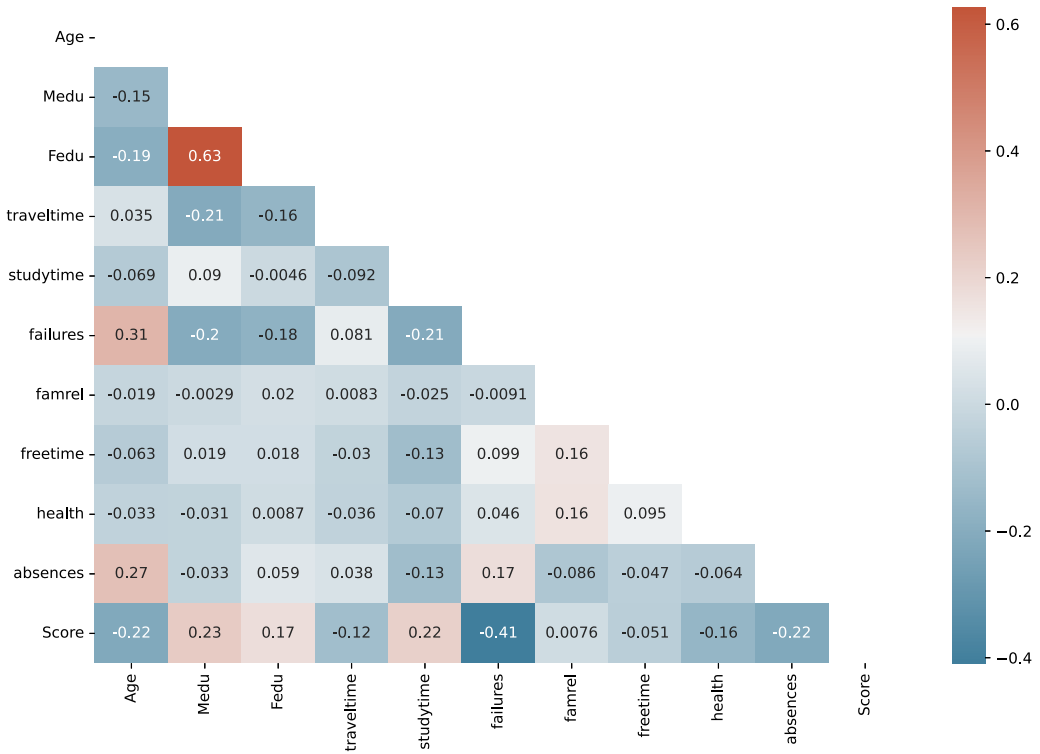
Figure 4. *Correlation heatmap for student performance data.*

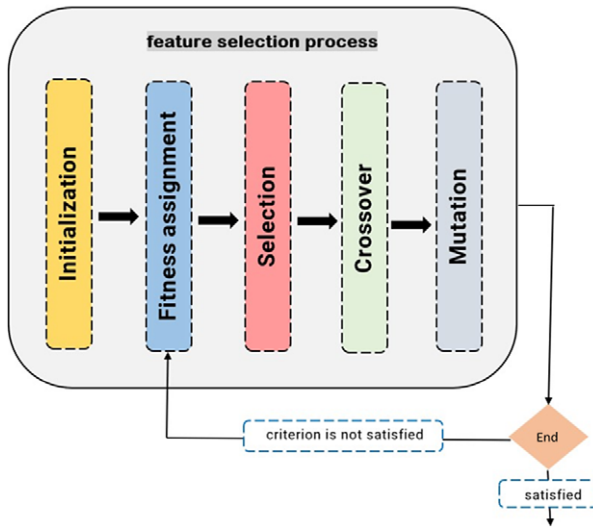**Figure 5.** *The features selection process using genetic algorithm.*

hybrid offspring by combining the genetic information of both parents. In a process known as mutation, the order of genes on a chromosome is changed, resulting in a whole new chromosome. When termination requirements are met, the algorithm stops and a new generation of offspring emerges that is not significantly different from the preceding one.

Algorithm 1 Genetic algorithm

Begin
Generate the initial population
Compute Fitness
REPEAT
Perform Selection
Perform Crossover
Perform Mutation
Compute Fitness
UNTIL Population converges
END

• *Boruta algorithm*

The Boruta algorithm is a ML technique for choosing which features to employ. Specifically, it works by duplicating the features in the original dataset into a new one, which are then referred to as shadow features (Ruto, 2022) as shown in Figure 6.

After including these characteristics, the resulting feature space has double the dimension of the original dataset. The method next creates a classifier (Random Forest Classifier) on the expanded feature space and uses a statistical test called the [Z-Score] to rank the features in order of relevance. A feature is regarded as important and retained if its value is greater than the maximum importance of its shadow characteristics; otherwise, it is discarded. This continues until either a certain number of iterations has been reached or all features have been validated or abandoned.

3.3. Classifiers and learning algorithms

The classifiers tested in the experiments are the following:

Random Forest: Conceived by Breiman in the 2000s, operates as an ensemble method utilizing a myriad of diverse decision trees to enhance prediction and classification (Wang et al., 2021). The
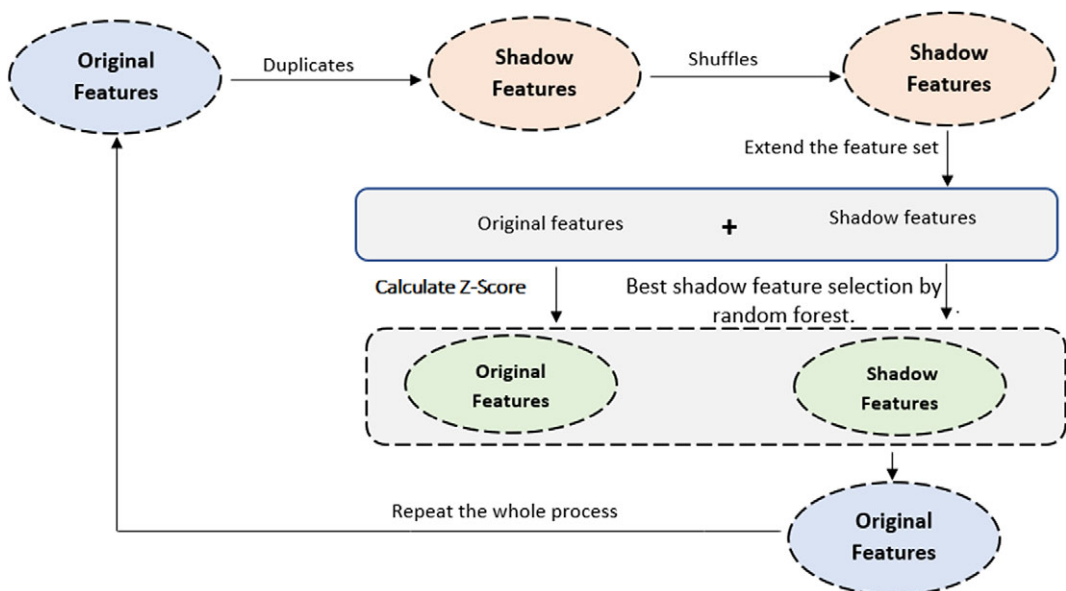

Figure 6. *The features selection process using Boruta algorithm.*

strength of Random Forest lies in its incorporation of randomness during both tree construction and feature selection (Altman and Krzywinski, 2017). This intentional diversity equips the model to adeptly handle intricate challenges, resulting in consistent success across various studies (Bock et al., 2005).

Balanced Bagging Classifier: A counterpart to Scikit-Learn's bagging method, adeptly handles imbalanced data. By skillfully sampling smaller portions from the majority class, this classifier ensures fair and representative results (imbalanced-learn.org).

Easy Ensemble Classifier: A formidable solution for imbalanced data, employs a unique approach. It systematically creates a more balanced dataset by carefully shrinking the dominant class. This strategy enables training a classifier on a more even distribution, resulting in fairer predictions for all (Liu et al., 2008).

XGBOOST: A leader in gradient boosting, continually evolves and adapts. Introduced by Chen and Guestrin (2016) and built upon J. H. Friedman (2001) is foundational work, it demonstrates exceptional performance in various classification and regression challenges (Hongshan et al., 2019).

ADABOOST: An ensemble learning method, envisions individual learners uniting to form a formidable entity. In this iterative process, less effective learners gain more weight and are retrained, culminating in a robust and accurate final model (Kavish, 2022).

LIGHTGBM: Elevates tree-based methods, pushing the boundaries of gradient boosting. Outperforming XGBoost (Sahlaoui et al., 2021), it adopts a leaf-wise growth strategy, constructing a tree one leaf at a time. With its versatility in handling continuous and categorical features, LightGBM proves invaluable for data scientists. Its popularity is reinforced by support for parallel processing and GPU training, enhancing efficiency with large datasets.

CATBOOST: Excels in handling categorical data, utilizing ordered boosting for superior prediction accuracy in this domain (Kavish, 2022). Its meticulous approach involves carefully ordering and analyzing categories, uncovering hidden patterns for more precise predictions.

3.4. Metrics of performance or evaluation methods

3.4.1. Metrics used in the experiment

To compare and determine the best model, it is necessary to evaluate the performance of the classifiers. There are many ways to measure and verify ML algorithms. This research uses five different criteria to compare the different tactics. These metrics are F1, precision, recall, accuracy, and AUC. They are widely regarded as the best metrics to gauge performance in unbalanced situations:

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}, \tag{3.1}$$

$$Precision = \frac{TP}{TP + FP}, \tag{3.2}$$

$$Recall = \frac{TP}{TP + FN}, \tag{3.3}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \tag{3.4}$$

$$AUC = \frac{1 + TP_{rate} - FP_{rate}}{2}. \tag{3.5}$$

Moreover, the statistical assessment approach allows for stronger analysis and comparison.

3.4.2. Statistical analysis

To evaluate the effectiveness of the proposed methods, this study will utilize established statistical techniques. The Shapiro–Wilk test will assess data normality (Brownlee, 2018; Jäntschi and Bolboacă, 2018), followed by either ANOVA (for normally distributed data; Fisher, 1959) or the Friedman test (for non-normal data; Kaur and Kaur, 2018) to identify significant differences in performance metrics (accuracy, precision, recall) across the four experimental approaches: SMOTE–ENN with Boruta, SMOTE–ENN with genetic algorithm, and the baseline strategy with various classifiers. Post-hoc analysis (if applicable) will pinpoint specific method differences. Additionally, rankings based on performance metrics will be generated to identify the most effective approach (Friedman, 1940).

4. Experiments and result analysis

In this section, we delve into a comprehensive examination of the gathered findings. Initially, we provide a detailed elucidation of the specific hyper-parameter settings for each of the ML models and the key parameter settings for each feature selection method employed in this work. Subsequently, we discuss the actual implementation and deployment of the model. We outline the experimental process used to assess the accuracy of the proposed models for predicting student performance, comparing various ML techniques through training and cross-validation score analysis. We track accuracy, precision, recall, and AUC values for model comparison. Our aim is to evaluate the performance gains achieved by employing tailored strategies on the Nigerian dataset, utilizing ensemble techniques, hybrid data sampling methods, such as SMOTE–ENN, combined with wrapper-based feature selection techniques, particularly incorporating Boruta and genetic algorithms, and comparing them against baseline strategies. Additionally, we investigate the feasibility of deploying ensemble techniques using Streamlit on the front end and Heroku on the back end.

4.1. The performance of the different classifiers

4.1.1. Hyper-parameters setting

In this study, all evaluations were performed using LightGBM, XGBoost, AdaBoost, CatBoost, Random Forest, Easy Ensemble, and Balanced Bagging. Table 2 shows the specific hyper-parameter settings for each of the ML models used in this work.

In addition, the key parameter settings for each feature selection method employed in this work are listed in Table 3.

List of hyper-parameters used in Boruta and genetic algorithm

With these settings, optimal outcomes are achieved, and the replicability and reproducibility of our study are enhanced. To achieve this, this study employs GeneticSelectionCV, a feature selection technique

Table 2. *Machine learning models with their specific hyper-parameters settings*

Methods	Hyper-parameters
Random Forest	N_estimators = 113, min_samples_leaf = 2.
Balanced Bagging Classifier	n_estimators = 500.
LightGBM	Learning_rate = 0.1, max_depth = −5, num_leaves = 31.
XGBoost	olsample_bytree = 0.3, gamma = 0.1, learning_rate = 1.0, max_depth = 8, min_child_weight = 3, objective = multi:softprob, n_estimators, loss = deviance.
Easy Ensemble	random_state = 42.
AdaBoost	algorithm = 'SAMME', base_estimator = None, learning_rate = 0.3. n_estimators = 300, random_state = 1.
CatBoost	random_state = 1.

Table 3. *Hyper-parameters used in Boruta and genetic algorithm*

Methods	Hyper-parameters
Boruta	N_estimators = 'auto', verbose = 2, random_state = 1.
Genetic algorithm	cv = 5, verbose = 0, scoring = 'accuracy', max_features = 11, n_population = 120, crossover_proba = 0.5, mutation_proba = 0.2, n_generations = 10, crossover_independent_proba = 0.04, tournant_size = 3, n_gen_no_change = 10, caching = True, n_jobs = −1.

within scikit-learn that utilizes a genetic algorithm to optimize the feature subset for a ML model. Notably, GeneticSelectionCV incorporates cross-validation during the selection process, mitigating the risk of overfitting. BorutaPy, another feature selection library, is implemented to identify relevant features from a potentially large pool. This approach aims to improve model performance and reduce the risk of overfitting by discarding irrelevant features. Scikit-learn provides a comprehensive suite of ML algorithms, including ensemble methods like Random Forest, AdaBoost, and Gradient Boosting. These ensemble methods leverage the power of multiple models to achieve superior performance compared to individual models.

4.1.2. The results of classifiers performance with and without SMOTE–ENN

In Table 4, we observe that the classifiers' performance on the imbalanced dataset demonstrates the challenges in predicting student performance. CatBoost exhibits the highest accuracy at 85%, but even this leading classifier falls short of achieving impressive accuracy. Notably, all classifiers show precision results below 70%, indicating the overall difficulty in correctly classifying student performance. This is further substantiated by F1-scores, which reveal that predictive classifiers struggle with a few classes, with CatBoost achieving an F1-score of 59%, reflecting suboptimal classifier performance.

Table 4 also highlights the impact of addressing class imbalances using SMOTE–ENN. With this resampling technique, CatBoost significantly improves its accuracy to 97%, demonstrating its efficacy on balanced data. However, despite the improvement in accuracy, certain classes continue to pose challenges, emphasizing the need to address the issue of imbalanced data. Furthermore, the table shows varying

Table 4. *The results of classifiers performance with and without SMOTE–ENN*

	Algorithm	Accuracy	F1-score	Precision	Recall	AUC	Time (s)
NONE	Random Forest	0.84	0.56	0.66	0.54	0.831	0.215482
	Balanced Bagging Classifier	0.70	0.54	0.54	0.62	0.769	1.620828
	LightGBM	0.79	0.55	0.65	0.54	0.775	0.320182
	XGBoost	0.80	0.58	0.62	0.58	0.791	0.140504
	Easy Ensemble	0.44	0.34	0.42	0.41	0.663	0.658205
	AdaBoost	0.74	0.28	0.25	0.33	0.701	0.397638
	CatBoost	0.85	0.59	0.63	0.60	0.831	3.013832
SMOTE ENN	Random Forest	0.96	0.96	0.97	0.95	0.997	0.254074
	Balanced Bagging Classifier	0.94	0.93	0.94	0.93	0.991	2.559989
	LightGBM	0.94	0.94	0.94	0.93	0.994	0.252148
	XGBoost	0.96	0.95	0.96	0.95	0.997	0.164361
	Easy Ensemble	0.83	0.83	0.84	0.84	0.925	0.754161
	Adaboost	0.77	0.77	0.80	0.75	0.914	0.437412
	CatBoost	0.97	0.97	0.98	0.96	0.999	3.069313

execution times for the classifiers, highlighting the trade-off between classifier accuracy and the time required to execute them.

4.1.3. The results of classifiers after applying the Boruta feature selection method

In our examination of feature selection in Tables 5 and 6, Boruta adeptly identifies 75% of features as pivotal, showcasing its proficiency in pinpointing pertinent student attributes. Noteworthy is its sustained impact on classifier AUC, consistently maintaining a high value of 99%, validating its reliability in forecasting student outcomes. This unwavering stability underscores Boruta's effectiveness in systematically choosing essential predictors for tailored interventions in education, offering valuable perspectives for both practitioners and policymakers in the field.

4.1.4. The result of classifiers after applying genetic algorithms and the feature selection method

When applying the genetic algorithm, we discover that attributes such as study time, parents' education, and school absence significantly impact student performance. The results in Table 6 highlight that both Boruta and genetic algorithms lead to improvements in classifier performance, with AUC values indicating high prediction model performance with a value of 99%. Importantly, the genetic algorithm retains less than 50% of the student attributes, contributing to efficient model building. These numerical findings underscore the utility of feature selection methods in enhancing the precision of the prediction models while reducing the computational and time costs associated with feature-rich datasets. The choice of genetic algorithm for feature selection stands out due to its ability to swiftly construct accurate models. In essence, these results provide empirical validation of the strategies employed in our study, substantiating their effectiveness in addressing the challenges of predicting student performance. In conclusion,

Table 5. *Selected features by Boruta and genetic algorithm*

Feature	Boruta feature	Genetic feature
Age	✓	✓
Gender	✓	✓
Location	✓	×
Famsize	✓	✓
Pstatus	✓	×
Medu	✓	✓
Fedu	✓	✓
Semester	×	×
Traveltime	✓	×
Studytime	✓	✓
Failures	×	×
Schoolsup	×	✓
Famsup	✓	×
Paid	×	✓
Activities	✓	×
Nursery	✓	×
Higher	×	×
Internet	✓	×
Famrel	✓	×
Freetime	✓	×
Health	✓	✓
Absences	✓	✓

Table 6. *The results of classifiers performance using SMOTE–ENN + Boruta and SMOTE–ENN + genetic algorithm*

	Algorithm	Accuracy	F1-score	Precision	Recall	AUC	Time (s)
BORUTA	Random Forest	0.96	0.96	0.97	0.95	0.998	0.757156
	Balanced Bagging Classifier	0.93	0.93	0.94	0.93	0.991	2.196038
	LightGBM	0.94	0.94	0.95	0.93	0.993	0.084831
	XGBoost	0.95	0.94	0.95	0.95	0.990	0.149496
	Easy Ensemble	0.82	0.81	0.81	0.83	0.915	1.214677
	AdaBoost	0.77	0.76	0.79	0.76	0.911	0.432469
	CatBoost	0.97	0.97	0.98	0.96	0.999	2.995022
GENITIC	Random Forest	0.97	0.97	0.97	0.97	0.997	0.506400
	Balanced Bagging Classifier	0.96	0.95	0.96	0.95	0.996	1.132528
	LightGBM	0.97	0.97	0.97	0.96	0.996	0.069106
	XGBoost	0.96	0.95	0.96	0.95	0.993	0.146872
	Easy Ensemble	0.82	0.82	0.83	0.83	0.910	0.604066
	AdaBoost	0.80	0.79	0.83	0.78	0.90	0.275252
	CatBoost	0.98	0.97	0.98	0.97	0.997	1.212168

Table 7. *Shapiro test*

Statistics	Mean	Variance	Standard deviation	*P*-value
1.520	0.971	0.0014	0.037	0.0000086

this discussion underscores the importance of addressing class imbalance and applying feature selection methods in the context of predicting student performance. These techniques not only enhance model accuracy but also contribute to computational efficiency. The choice of classifier also plays a significant role in the overall performance. By combining these strategies, we aim to provide educational institutions and policymakers with valuable insights into enhancing academic achievements.

4.2. The result of statistical analysis

Based on the Shapiro–Wilk normality test results, the data for the Nigeria dataset do not follow a normal distribution ($p < 0.05$), as displayed in Table 7. Consequently, parametric statistical tests like ANOVA are not suitable for analyzing these results. Therefore, a non-parametric test, specifically the Friedman test, is the more appropriate choice for this dataset to assess statistically significant differences in precision-recall, F1, and AUC metrics across the various classifiers and feature selection and sampling techniques.

Friedman test for Nigerian dataset

Table 8 demonstrates that the *p* values of the data for the Friedman test are greater than 0.05 ($a = 0.05$) for the Nigeria dataset. This indicates that the null hypothesis has been accepted. This suggests that the experimental procedures are not statistically significantly different. This implies that all three approaches could potentially yield comparable impacts on the academic performance of Nigerian students; however, additional research is imperative to confirm these findings.

4.3. Deployment results

Streamlit, a user-centric web application development framework, plays a crucial role in disseminating the developed model's capabilities to a wider audience. This platform streamlines the deployment process

Table 8. *Friedman test*

	Source	*W*	ddof1	*Q*	*P*-value
Friedman	Within	0.108571	2	1.52	0.467

by consolidating both front-end and back-end functionalities within a single environment. Unlike frameworks like Flask or Express, Streamlit eliminates the need for separate development of user interfaces and server-side logic, facilitating a more rapid and efficient deployment (Sharma, 2020). The deployed application leverages Streamlit's intuitive interface to provide users with a user-friendly experience for interacting with the model. This interface incorporates several input fields, allowing users to enter relevant data points that influence student performance, as identified by the model. Once the user submits this input, the application seamlessly integrates with the trained model, performing real-time predictions on the user's academic standing as depicted in Figure 7.

The results are then presented in a clear and concise manner, enabling users to understand their predicted performance outcomes (favorable or unfavorable). Furthermore, the interactive nature of the application empowers users to explore various "what-if" scenarios. By modifying the input values, users can observe how different factors might influence their academic performance. This dynamic feedback empowers users to gain valuable insights and potentially make informed decisions to optimize their academic trajectory. In essence, Streamlit facilitates the deployment of a user-friendly and interactive web application that leverages the predictive capabilities of the developed model. This application provides a valuable tool for students to gain insights into their academic standing and explore strategies for potential improvement.

4.4. Discussion

This study explores the use of ML techniques within the unique environment of Nigerian classrooms, focusing on developing predictive models to identify underperforming students. This empowers
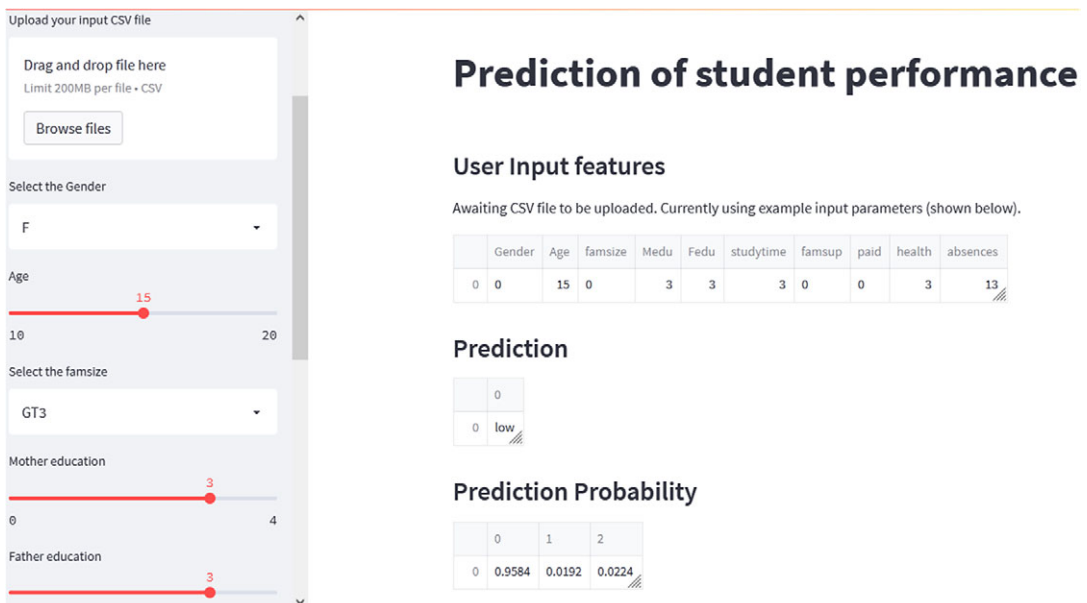

Figure 7. *Web-based student performance model application.*

universities both within and outside Africa to proactively support struggling students, optimize educational costs, and promote inclusive development. We move beyond previous studies by tackling the intricacies of predicting student performance in demanding scenarios, particularly within small and imbalanced datasets. The dataset's limited size and imbalanced class distribution posed significant obstacles. With only three categories in the dataset, capturing meaningful trends was particularly difficult. Typically, larger datasets are required to effectively handle multi-classification problems. To tackle class imbalance, we applied SMOTE–ENN techniques to oversample data on student performance. This approach not only mitigates bias but also reduces computational costs. Balancing class distribution is crucial for ensuring accurate predictions across all categories. To address the challenges of irrelevant features, feature selection became a critical aspect of our research. We needed to ensure the efficiency of our ML models by eliminating redundant or minimally informative predictions. Our choice of feature selection methods included the genetic algorithm and Boruta, aiming to leverage the strengths of both approaches. Genetic algorithms have proven effective in various domains, while Boruta is a robust technique for identifying relevant features. Two classifiers, XGBoost and LightGBM, emerged as strong performers in our experiments. These algorithms demonstrated their ability to handle the challenges posed by the small dataset and class imbalance. It's worth noting that the choice of classifier has a substantial impact on predictive accuracy. Our analysis highlighted attributes such as study time, parents' education, school support, the number of school absences, health, and participation in extra-paid classes as significant factors influencing students' academic achievements. Understanding the importance of these attributes is invaluable for educational institutions. Incorporating modern deployment tools, such as Streamlit and Heroku, further facilitates the practical application of our research. Streamlit simplified the development of ML applications, allowing more time for model improvement, while Heroku streamlined system management for quicker and simpler app development. Our research transcends technical advancements by translating these findings into real-world impact through an accessible web application built with Streamlit and Heroku. This empowers universities within and outside Africa to customize model settings and view real-time results, fostering informed decision-making based on data-driven insights and predictions. Moreover, this research demonstrates how data science can inform policy by improving situation analysis through a deeper understanding of factors impacting student performance, enhancing predictions for timely interventions and resource allocation, and supporting public service design for effective and legitimate education policies. Ultimately, this research empowers stakeholders to harness data science for equitable and impactful education in African smart cities, paving the way for a future where data empower education, fosters sustainable development, and unlocks the potential of every student in African smart cities.

5. Conclusion

This research explores the use of ML techniques within the unique environment of Nigerian classrooms, focusing on developing predictive models to identify underperforming students. This empowers universities both within and outside Africa to proactively support struggling students, optimize educational costs, and promote inclusive development. This study harnessed the power of ML to predict student performance and employed feature selection techniques to expedite the resolution of this critical issue. The results from our experimentation underscore the viability of the genetic algorithm and Boruta feature selection methods when coupled with SMOTE–ENN and CatBoost classifiers. Notably, this approach proved effective in reducing the number of features by over 50% while significantly expediting the model-building process. The genetic algorithm stood out by offering the swiftest training times, demonstrating its applicability to extensive and complex training datasets. Feature selection emerges as a pivotal aspect of ML and prediction, as it serves to enhance prediction accuracy across various student classifiers while simultaneously reducing processing costs. This study identified a range of factors, including study time, mother's and father's education levels, school support, school absences, health status, and participation in extra-paid classes, as significant contributors to students' academic achievements. These attributes, when integrated into the model, significantly enhance prediction accuracy and interpretability. Classifier

analysis demonstrated that robust classifiers like CatBoost and Random Forest deliver promising results across all performance indicators. However, it's essential to acknowledge the higher maintenance costs associated with these classifiers compared to XGBoost and LightGBM. The findings equip software developers and researchers with insights into the selection of sampling and feature selection strategies, enabling them to optimize performance under various scenarios and evaluation criteria. Moreover, the utilization of modern technologies like Streamlit and Heroku for web deployment underscores the importance of user-friendly ML applications and efficient model building. This aspect not only enhances the accessibility of ML models but also supports the continuous improvement of predictive outcomes. Our research transcends technical advancements by translating these findings into real-world impact through an accessible web application built with Streamlit and Heroku. This empowers universities within and outside Africa to customize model settings and view real-time results, fostering informed decision-making based on data-driven insights and predictions. Moreover, this research demonstrates how data science can inform policy by improving situation analysis through a deeper understanding of factors impacting student performance, enhancing predictions for timely interventions and resource allocation, and supporting public service design for effective and legitimate education policies. Ultimately, this research empowers stakeholders to harness data science for equitable and impactful education in African smart cities, paving the way for a future where data empower education, fosters sustainable development, and unlocks the potential of every student in African smart cities.

6. Limitations and future work

This study has certain limitations. The use of publicly available datasets rather than a custom student data collection and the relatively small dataset comprising less than a thousand records pose constraints on the scope of the study. Future research should aim to leverage larger, more feature-rich datasets, harness cutting-edge big data technologies, and explore deep learning techniques to further enhance model performance. The challenge of feature selection for academic performance prediction will be an area of focused investigation. Future efforts will involve amassing substantial datasets and conducting comprehensive experiments to evaluate the efficacy of different feature selection algorithms and assess their adaptability to various contexts. Furthermore, exploring the synergy between feature selection techniques and classification algorithms to enhance prediction performance will be a priority. The study's forward-looking approach extends to the exploration of individualized course recommendation systems and the development of a web application for deploying recommender system models. With an eye on mobile app development, we aim to address challenges associated with the individualized course recommender system. In conclusion, this study presents a robust methodology that effectively addresses the challenge of student performance prediction in the education sector. While acknowledging its limitations, the study offers valuable insights and paves the way for broader applications and extensive research expansion.

Author contribution. Conceptualization: E.A.A.A., S.A.; Formal analysis: H.S.; Methodology: H.S.; Software: H.S.; Supervision: E.A.A.A., S.A.; Validation: E.A.A.A., S.A.; Visualization: H.S.; Writing—original draft: H.S.; Writing—review and editing: H.A., S.C.K.T., B.A.

References

Ajibade S-SM, **Ahmad NB and Shamsuddin SM** (2020) A data mining approach to predict academic performance of students using ensemble techniques. In *Intelligent Systems Design and Applications: 18th International Conference on Intelligent Systems Design and Applications (ISDA 2018) Held in Vellore, India*, December 6–8, 2018, Volume *1*. India: Springer Cham, pp. 749–760.

Almasri A, **Celebi E and Alkhawaldeh RS** (2019) Emt: Ensemble meta-based tree model for predicting student performance. *Scientific Programming 2019*, 3610248.

Altman N and Krzywinski M (2017) Ensemble methods: Bagging and random forests. *Nature Methods 14*(10), 933–935.

Amrieh EA, **Hamtini T and Aljarah I** (2016) Mining educational data to predict student's academic performance using ensemble methods. *International Journal of Database Theory and Application 9*(8), 119–136.

Anand N, **Sehgal R**, **Anand S and Kaushik A** (2021) Feature selection on educational data using Boruta algorithm. *International Journal of Computational Intelligence Studies 10*(1), 27–35.

Arif A (2020) Build a machine learning web app with Streamlit and python [Heroku deployment], December. Available at https://dataaspirant.com/feature-selection-methods-machine-learning/ (accessed December 2022).

Batista GE, **Prati RC and Monard MC** (2004) A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD Explorations Newsletter 6*(1), 20–29.

Begum S and Padmannavar S (2022) Genetically optimized ensemble classifiers for multiclass student performance prediction. *International Journal of Intelligent Engineering and Systems 15*(2), 316–328.

Bharti KK and Singh PK (2014) A survey on filter techniques for feature selection in text mining. In *Proceedings of the Second International Conference on Soft Computing for Problem Solving (SocProS 2012)*, December 28–30, 2012. New Delhi, India: Springer, pp. 1545–1559.

Bhat S (2019) Feature selection: Filter method, wrapper method and embedded method. Available at https://www.datasciencesmachinelearning.com/2019/10/feature-selection-filter-method-wrapper.html (accessed December 2022).

Bock M, **Rossner G**, **Wissen M**, **Remm K**, **Langanke T**, **Lang S**, **Klug H**, **Blaschke T and Vrščaj B** (2005) Spatial indicators for nature conservation from European to local scale. *Ecological Indicators 5*(4), 322–338.

Breiman L (1996) Bagging predictors. *Machine Learning 24*, 123–140.

Breiman L (1998) Arcing classifier (with discussion and a rejoinder by the author). *Annals of Statistics 26*(3), 801–849.

Breiman L (2001) Random forests. *Machine Learning 45*, 5–32.

Brownlee J (2018) A gentle introduction to normality tests in python. Available at https://machinelearningmastery.com/a-gentle-introduction-to-normality-tests-in-python/ (accessed December 2022).

Butvinik D (2021) Feature selection: Benefits and methods. how to choose a feature selection method? Available at https://www.thetechplatform.com/post/feature-selection-benefits-and-methods-how-to-choose-a-feature-selection-method (accessed December 2022).

Buvaneshwaran K (2022) Deploy machine learning model using Streamlit. Available at https://copyassignment.com/deploy-machine-learning-model-using-streamlit/ (accessed December 2022).

Chandrashekar G and Sahin F (2014) A survey on feature selection methods. *Computers & Electrical Engineering 40*(1), 16–28.

Chaudhury P, **Mishra S**, **Tripathy HK and Kishore B** (2016) Enhancing the capabilities of student result prediction system. In *Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies*. Udaipur, India: Association for Computing Machinery, pp. 1–6.

Chawla NV, **Bowyer KW**, **Hall LO and Kegelmeyer WP** (2002) Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research 16*, 321–357.

Chen T and Guestrin C (2016) Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining*. San Francisco, CA: Association for Computing Machinery, pp. 785–794.

Costa EB, **Fonseca B**, **Santana MA**, **de Araújo FF and Rego J** (2017) Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses. *Computers in Human Behavior 73*, 247–256.

Dempere Guillermo G (2018) Predictive data driven dashboard as an academic guidance support platform for mentors. Available at https://diposit.ub.edu/dspace/handle/2445/130959 (accessed December 2022).

Dhar J and Jodder AK (2020) An effective recommendation system to forecast the best educational program using machine learning classification algorithms. *Ingénierie des Systèmes d'Information 25*(5), 559–568.

Fernández A, **García S and Herrera F** (2011) Addressing the classification with imbalanced data: Open problems and new challenges on class distribution. *HAIS 1*, 1–10.

Fisher RA (1959) *Statistical Methods and Scientific Inference*. Technical report. Edinburgh: Oliver and Boyd.

Flynn D (2021) Beautify your data with Primehub and Streamlit. Available at https://blog.infuseai.io/beautify-your-data-with-primehub-and-streamlit-c9caaa92f24e (accessed December 2022).

Freund Y and Schapire RE (1997) A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences 55*(1), 119–139.

Friedman JH (1997) On bias, variance, 0/1—Loss, and the curse-of-dimensionality. *Data Mining and Knowledge Discovery 1*, 55–77.

Friedman JH (2001) Greedy function approximation: A gradient boosting machine. *Annals of Statistics 29*, 1189–1232.

Friedman M (1940) A comparison of alternative tests of significance for the problem of m rankings. *Annals of Mathematical Statistics 11*(1), 86–92.

Galar M, **Fernandez A**, **Barrenechea E**, **Bustince H and Herrera F** (2011) A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 42*(4), 463–484.

Garcia JD and Skrita A (2019) Predicting academic performance based on students' family environment: Evidence for Colombia using classification trees. *Psychology, Society & Education 11*(3), 299–311.

Ghorbani R and Ghousi R (2020) Comparing different resampling methods in predicting students' performance using machine learning techniques. *IEEE Access 8*, 67899–67911.

Guan D, **Yuan W**, **Lee Y-K and Lee S** (2009) Nearest neighbor editing aided by unlabeled data. *Information Sciences 179*(13), 2273–2282.

Han H, **Wang W-Y and Mao B-H** (2005) Borderline-smote: A new over-sampling method in imbalanced data sets learning. In *Advances in Intelligent Computing: International Conference on Intelligent Computing, ICIC 2005*, Hefei, China, August 23–26, 2005, Proceedings, Part I *1*. Berlin: Springer Berlin Heidelberg, pp. 878–887.

Zhao H, **Yan X**, **Wang G and Yin X** (2019) Fault diagnosis of wind turbine generator based on deep autoencoder network and xgboost. *Automation of Electric Power Systems 43*(1), 81–90.

Jäntschi L and Bolboacă SD (2018) Computation of probability associated with Anderson–Darling statistic. *Mathematics 6*(6), 88.

Kaur A and Kaur I (2018) An empirical evaluation of classification algorithms for fault prediction in open source projects. *Journal of King Saud University-Computer and Information Sciences 30*(1), 2–17.

Kavish (2022) Ultimate guide to boosting algorithms. Available at https://www.analyticsvidhya.com/blog/2022/12/ultimate-guide-to-boosting-algorithms/ (accessed December 2022).

Kim A (2021) Deploying data science web apps to the cloud. Available at https://www.practicum.blog/posts/deploying-data-science-web-apps-to-the-cloud (accessed December 2022).

Kimnaruk Y (2022) Fast machine learning model deployment. Available at https://medium.com/mlearning-ai/fast-machine-learning-model-deployment-465973dff40a (accessed December 2022).

kindsonthegenius (2018) What is genetic algorithm? (a simple and detailed explanation). Available at https://kindsonthegenius.com/blog/what-is-genetic-algorithm-a-simple-and-detailed-explanation/ (accessed December 2022).

Kotsiantis S, **Kanellopoulos D and Pintelas P** (2006) Handling imbalanced datasets: A review. *GESTS International Transactions on Computer Science and Engineering 30*(1), 25–36.

Kou G, **Xu Y**, **Peng Y**, **Shen F**, **Chen Y**, **Chang K and Kou S** (2021) Bankruptcy prediction for SMEs using transactional data and two-stage multiobjective feature selection. *Decision Support Systems 140*, 113429.

Kou G, **Yang P**, **Peng Y**, **Xiao F**, **Chen Y and Alsaadi FE** (2020) Evaluation of feature selection methods for text classification with small datasets using multiple criteria decision-making methods. *Applied Soft Computing 86*, 105836.

Leppänen L, **Leinonen J**, **Ihantola P and Hellas A** (2017) Predicting academic success based on learning material usage. In *Proceedings of the 18th Annual Conference on Information Technology Education*, Rochester, NY, pp. 13–18.

Li H, **Li J**, **Chang P-C and Sun J** (2013) Parametric prediction on default risk of Chinese listed tourism companies by using random oversampling, isomap, and locally linear embeddings on imbalanced samples. *International Journal of Hospitality Management 35*, 141–151.

Liu J, **Chen XX**, **Fang L**, **Li JX**, **Yang T**, **Zhan Q**, **Tong K and Fang Z** (2018) Mortality prediction based on imbalanced high-dimensional ICU big data. *Computers in Industry 98*, 218–225.

Liu X-Y, **Wu J and Zhou Z-H** (2008) Exploratory undersampling for class-imbalance learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 39*(2), 539–550.

Longadge R and Dongre S (2013) Class imbalance problem in data mining review. *International Journal of Computer Science and Network 2*(1). Preprint, arXiv:1305.1707.

Mangina E and Psyrra G (2021) Review of learning analytics and educational data mining applications. *IATED*, pp. 949–954.

Melo A (2023) The best 10 alternatives to Heroku in 2023. Available at https://blog.back4app.com/heroku-alternatives/ (accessed February 2023).

Moharekar TT and Pol UR (2021) Academic performance prediction application (appa). *YMER 20*, 179–196.

Moreno García MN, **González Robledo J**, **Martín González F**, **Sánchez Hernández F and Sánchez Barba M** (2014) Machine learning methods for mortality prediction of polytraumatized patients in intensive care units–dealing with imbalanced and high-dimensional data. In *Intelligent Data Engineering and Automated Learning–IDEAL 2014: 15th International Conference*, Salamanca, Spain, September 10–12, 2014. Springer, Cham, pp. 309–317.

Onan A (2018) An ensemble scheme based on language function analysis and feature engineering for text genre classification. *Journal of Information Science 44*(1), 28–47.

Onan A (2019) Two-stage topic extraction model for bibliometric data analysis based on word embeddings and clustering. *IEEE Access 7*, 145614–145633.

Onan A (2022) Bidirectional convolutional recurrent neural network architecture with group-wise enhancement mechanism for text sentiment classification. *Journal of King Saud University- Computer and Information Sciences 34*(5), 2098–2117.

Onan A (2023) Hierarchical graph-based text classification framework with contextual node embedding and Bert-based dynamic fusion. *Journal of King Saud University-Computer and Information Sciences 35*, 101610.

Onan A and Korukoğlu S (2017) A feature selection model based on genetic rank aggregation for text sentiment classification. *Journal of Information Science 43*(1), 25–38.

Onan A, **Korukoğlu S and Bulut H** (2016) Ensemble of keyword extraction methods and classifiers in text classification. *Expert Systems with Applications 57*, 232–247.

Onan A, **Korukoğlu S and Bulut H** (2017) A hybrid ensemble pruning approach based on consensus clustering and multi-objective evolutionary algorithm for sentiment classification. *Information Processing & Management 53*(4), 814–833.

Onan, A. (2019) Consensus clustering-based undersampling approach to imbalanced learning. *Scientific Programming 2019*, 5901087.

Paddalwar S, **Mane V and Ragha L** (2022) Predicting students' academic grade using machine learning algorithms with hybrid feature selection approach. *ITM Web of Conferences 44*, 03036.

Phua EJ and Batcha NK (2020) Comparative analysis of ensemble algorithms' prediction accuracies in education data mining. *Journal of Critical Reviews 7*(3), 37–40.

Prakash J (2021) Deployment of machine learning models. Available at https://pianalytix.com/deployment-of-machine-learning-models/ (accessed December 2022).

Ruto N (2022) How to get started with the Boruta algorithm in machine learning. Available at https://www.section.io/engineering-education/getting-started-with-boruta-algorithm/ (accessed December 2022).

Sahlaoui H, **Alaoui EAA**, **Nayyar A**, **Agoujil S and Jaber MM** (2021) Predicting and interpreting student performance using ensemble models and Shapley additive explanations. *IEEE Access 9*, 152688–152703.

Salal Y, **Abdullaev S and Kumar M** (2019) Educational data mining: Student performance prediction in academic. *International Journal of Engineering and Advanced Technology 8*(4C), 54–59.

Sammut C and Webb GI (2011) *Encyclopedia of Machine Learning*. New York: Springer Science & Business Media.

Sharm P (2020) Deploying machine learning models using Streamlit – an introductory guide to model deployment. Available at https://www.analyticsvidhya.com/blog/2020/12/deploying-machine-learning-models-using-streamlit-an-introductory-guide-to-model-deployment/ (accessed December 2022).

Sharma P (2020) Build a machine learning web app with Streamlit and python [Heroku deployment]. Available at https://machinelearningknowledge.ai/build-a-machine-learning-web-app-with-streamlit-and-python-heroku-deployment/ (accessed December 2022).

Shaukat K, **Nawaz I**, **Aslam S**, **Zaheer S and Shaukat U** (2016) Student's performance in the context of data mining. In *2016 19th International Multi-Topic Conference (INMIC)*. Islamabad, Pakistan: IEEE, pp. 1–8.

Shaukat K, **Nawaz I**, **Aslam S**, **Zaheer S and Shaukat U** (2017) *Student's Performance: A Data Mining Perspective*. Koln: LAP Lambert Academic Publishing.

Sobnath D, **Kaduk T**, **Rehman IU and Isiaq O** (2020) Feature selection for UK disabled students' engagement post higher education: A machine learning approach for a predictive employment model. *IEEE Access 8*, 159530–159541.

Spolaôr N, **Cherman EA**, **Monard MC and Lee HD** (2013) Relieff for multi-label feature selection. In *2013 Brazilian Conference on Intelligent Systems*, pp. 6–11.

Tang Y, **Zhang Y-Q**, **Chawla NV and Krasser S** (2008) SVMs modeling for highly imbalanced classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 39*(1), 281–288.

Thetechwriters (2021) Machine learning model deployment using Streamlit. Available at https://www.analyticsvidhya.com/blog/2021/10/machine-learning-model-deployment-using-streamlit/ (accessed December 2022).

Turabieh H (2019) Hybrid machine learning classifiers to predict student performance. In *2019 2nd International Conference on New Trends in Computing Sciences (ICTCS)*. Amman, Jordan: IEEE, pp. 1–6.

Ünal F (2020) Data mining for student performance prediction in education. *Data Mining-Methods, Applications and Systems 28*, 423–432.

Vadapalli P (2020) A step-by-step guide to deploying ml models using Streamlit. Available at https://www.upgrad.com/blog/deploying-ml-models-using-streamlit/ (accessed December 2022).

Wang K, **Tian J**, **Zheng C**, **Yang H**, **Ren J**, **Li C**, **Han Q and Zhang Y** (2021) Improving risk identification of adverse outcomes in chronic heart failure using SMOTE+ ENN and machine learning. *Risk Management and Healthcare Policy 14*, 2453–2463.

Wolpert DH (1992) Stacked generalization. *Neural Networks 5*(2), 241–259.

Wutzl B, **Leibnitz K**, **Rattay F**, **Kronbichler M**, **Murata M and Golaszewski SM** (2019) Genetic algorithms for feature selection when classifying severe chronic disorders of consciousness. *PLoS One 14*(7), e0219683.

Yap BW, **Abd Rani K**, **Abd Rahman HA**, **Fong S**, **Khairudin Z and Abdullah NN** (2013) An application of oversampling, undersampling, bagging and boosting in handling imbalanced datasets. In *Proceedings of the First International Conference on Advanced Data and Information Engineering (DaEng-2013)*. Dordrecht: Springer, pp. 13–22.

Yin L, **Ge Y**, **Xiao K**, **Wang X and Quan X** (2013) Feature selection for high-dimensional imbalanced data. *Neurocomputing 105*, 3–11.

Yousafzai BK, **Hayat M and Afzal S** (2020) Application of machine learning and data mining in predicting the performance of intermediate and secondary education level student. *Education and Information Technologies 25*, 4677–4697.

Zaffar M, **Hashmani MA**, **Savita K**, **Rizvi SSH and Rehman M** (2020) Role of FCBF feature selection in educational data mining. *Mehran University Research Journal of Engineering & Technology 39*(4), 772–778.

Zhang X and Hu B-G (2014) A new strategy of cost-free learning in the class imbalance problem. *IEEE Transactions on Knowledge and Data Engineering 26*(12), 2872–2885.