Predictive Modeling, Interpretability, and Sustainability in Public University Admissions: A Case Study from Bangladesh

Abstract

This study explores the complex dynamics of undergraduate admissions in public universities in Bangladesh, where intense competition arises from limited seats and a large pool of applicants. Using sophisticated data mining and machine learning techniques, we analyze diverse factors influencing admission outcomes. A comprehensive dataset sourced from online surveys integrates academic achievements, socio-economic backgrounds, and other critical criteria to derive actionable insights. Machine learning classifiers such as Bagging, Random Forest, Gradient Boosting, and Extra Trees are employed, with hyperparameters fine-tuned via Grid Search and performance validated using 10-fold Cross-Validation to achieve accurate predictions of admission success. To enhance interpretability, SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) are utilized, offering clear insights into the contribution of individual features. These tools illuminate the significant impact of factors like previous academic achievements, socio-economic status, and access to preparatory resources in shaping admission outcomes. The findings underscore the need for targeted interventions to support students, particularly those from underrepresented or disadvantaged backgrounds. By addressing these inequities, this research emphasizes the importance of sustainability in higher education, aiming to foster long-term, equitable access to academic opportunities. Moreover, the integration of machine learning and explainable AI methodologies demonstrates their transformative potential in enhancing decision-making processes in education. This study contributes to a sustainable framework for improving transparency, equity, and inclusivity in university admissions systems.

Keywords: Undergraduate Admission, Public Universities, Bangladesh Education. Machine Learning, Grid Search, Cross-Validation, SHAP, LIME

1. Introduction

Undergraduate admission tests are pivotal in many countries' educational frameworks, offering a standardized approach to evaluate candidates' academic aptitude and readiness for higher education. These tests serve as a crucial tool for universities and colleges to assess applicants' potential, ensuring that they possess the necessary skills and knowledge to thrive in their chosen fields of study. In countries like the United States, the SAT (Scholastic Assessment Test) and ACT (American College Testing) are widely recognized and utilized for undergraduate admissions, providing a common benchmark for evaluating students from diverse educational backgrounds. Similarly, in the United Kingdom, the University and College Admission Service (UCAS) oversees the administration of standardized tests such as the SAT Subject tests or the Advanced Placement (AP) exams, which help universities assess the academic abilities of international applicants.

Undergraduate admission tests are essential not only for assessing academic proficiency but also for promoting fairness and equity in the admissions process. By utilizing standardized assessments, institutions can reduce the impact of subjective biases and socioeconomic factors, ensuring that all applicants are evaluated based on merit. Furthermore, these tests provide a comprehensive evaluation of candidates' cognitive abilities, such as critical thinking and problem-solving skills, crucial for success in higher education and beyond. Additionally, admission tests help universities maintain high academic standards by selecting candidates with the potential to excel in demanding academic environments through minimum score requirements or cutoffs, thereby upholding their reputation for excellence.

In Bangladesh, undergraduate admission tests also play a significant role in the country's educational landscape. A prime illustration of this is the admission tests for public universities, which are known for their intense competition and rigorous nature. These tests are conducted to select candidates for undergraduate programs in a wide range of disciplines, such as engineering, medicine, social sciences, and humanities. The admission tests in Bangladesh are specifically designed to assess students' academic abilities, including their proficiency in subjects like mathematics, physics, chemistry, and language. Moreover, they also evaluate students' critical thinking, problem-solving skills, and analytical reasoning, which are crucial for achieving success in higher education. However, the competition for admission to public universities in Bangladesh is intense, with a vast pool of applicants contending for a limited number of seats. For instance, in 2023, a staggering 13,74,488 students took part in the HSC and equivalent exams [1], while the number of available seats in the 49 public universities, as reported by Bangladesh Education Statistics 2021 [2], was only 51,152. However, it's worth noting that this figure has since increased to around 60,000. This glaring incongruity underscores the fierce competition and the daunting challenge of securing a coveted spot in public institutions.

The intense competition and limited seats in public universities often result in a large number of students being unable to secure admission or failing to get into the programs of their choice. Consequently, many students opt for private universities as an alternative route. Private universities in Bangladesh collectively offer a substantial number of seats, totaling 2,38,323 across 107 institutions [3]. However, the higher tuition fees associated with private institutions pose a significant financial challenge for many families in Bangladesh, especially considering the country's status as a developing nation. This financial burden often complicates access to higher education for students from less affluent backgrounds. Additionally, the varying quality of education and resources among private universities further complicates the decision-making process for prospective students. So, identifying the reasons why students are unable to secure admission to public universities is crucial.

Traditional educational approaches have historically relied on a comprehensive evaluation of student preferences, academic performance, and obstacles to achieving desired outcomes in order to customize educational experiences effectively. However, the process of collecting data through surveys, interviews, and observations faces significant challenges that limit its efficiency. One such challenge is the lack of scalability in traditional data collection methods; manual procedures are often time-consuming and labor-intensive, which can delay decision-making and overlook important insights as educational institutions expand. Moreover, the accuracy of data obtained through conventional means is at risk of human error, compromising the integrity of the analysis. These limitations, coupled with the inability of manual methods to capture the nuances of student preferences and objectives in evolving educational landscapes, underscore the need for modern solutions.

To tackle these challenges, educational researchers are increasingly utilizing data mining through online surveys and implementing Machine Learning (ML) classifiers to analyze the vast amount of data generated in educational settings. Online surveys provide a scalable and efficient approach to gathering data from a large number of participants, surpassing the constraints of conventional data collection methods [4]. Through the utilization of ML techniques, researchers can extract valuable insights from the collected data, uncovering patterns, trends, and correlations that may not be immediately evident through manual analysis [5, 6]. Furthermore, ML techniques enable the automation of repetitive tasks, reducing the potential for human error and enhancing the accuracy of the analysis [7]. By combining data mining with ML, educators can acquire a deeper comprehension of student preferences and behaviors, enabling more personalized and effective educational interventions [8, 9]. This integration of technology into the educational research process signifies a significant advancement in overcoming the limitations of traditional methods and unlocking new possibilities for enhancing student outcomes. The integration of technology reaches beyond academia, promoting sustainable practices and influencing various sectors [10, 11].

In a similar vein, this research endeavors to employ data mining techniques and ML classifiers to delve into the underlying reasons why students in Bangladesh may face challenges securing admission to public universities. Through the analysis of data collected via online surveys and advanced tools, the study aims to transparently uncover underlying factors, paving the way for targeted interventions to ensure sustainable and equitable access to higher education. The primary contributions of this investigation are as follows:

- Creation of a dataset sourced from a Google document, incorporating responses from both public and private universities. This dataset encompasses 15 different attributes crucial for gaining insights pertinent to the research objectives.
- Application of various machine learning classifiers on preprocessed data to ascertain their efficacy. A thorough examination is conducted to identify the classifier that offers the most reliable performance.
- Performing thorough evaluations to tackle data imbalance. Undersampling, oversampling, and no sampling techniques are applied, followed by comprehensive assessments of machine learning techniques under each scenario.
- Utilization of Shapley explanation techniques to determine the most influential features contributing to the outcomes. This analysis sheds light on why students encounter difficulties in gaining admission to public universities by highlighting significant factors.
- Integration of Lime explanation techniques to validate the findings obtained through Shapley. This approach serves as a complementary validation method, enhancing the robustness of the insights gleaned from the Shapley analysis.

Section 2 comprises related works, while Section 3 provides a description of the dataset. In Section 4, the research methodology is detailed, followed by Section 5, which presents the results and initiates the discussion. Section 6 addresses potential threats to validity, and finally, Section 7 concludes the paper.

2. Related Work

The amalgamation of machine learning and educational data mining is revolutionizing the educational landscape by examining student preferences, objectives, and performance indicators through comprehensive datasets. These methods reveal factors influencing university choices and academic paths, addressing challenges in achieving objectives and fostering economic growth by aligning education with industry demands [12]. Advanced algorithms identify key influences on student decisions, enabling targeted recruitment and resource optimization. This analytical approach, grounded in data, reveals patterns frequently overlooked by conventional methods, thereby fostering personalized interventions and predictive analytics aimed at improving student outcomes and individualizing learning experiences.

The study by Protikuzzaman et al. [13] presented a method for predicting undergraduate admission in the engineering faculty at BSMRSTU, Bangladesh. It utilized machine learning algorithms like XGBoost, LightGBM, and GBM on collected data to estimate admission probabilities pre and post-admission test. Factors such as academic performance, living area, family background, study habits, and social activities were considered crucial before the exam, while obtained marks, admission test year, and similar factors were important afterward. Evaluation metrics showed GBM achieving the highest accuracy of 95% post-exam. These models offered insights for students to assess and improve their admission chances, potentially reducing stress and enhancing preparation strategies.

Walid et al. [14] investigated factors leading to undergraduate admission failures using machine learning. They used a dataset with ten attributes and 343 observations to simplify model complexity. The combination of edited nearest neighbor (ENN) and borderline SVM-based SMOTE performed best, followed by borderline SVM-based SMOTE with Adaboost. The study aimed to help underprivileged and middle-income families improve admission prospects for public universities in Bangladesh. By applying data mining and ML, it provided insights into predicting exam outcomes and suggested the potential for mobile apps to assess admission chances. The models combining ENN, SVM-based SMOTE, and Adaboost showed promise for predicting test outcomes and supporting educational practices and student mental health.

The study [15] investigated the prediction of students' final grades in courses offered by private universities in Bangladesh using machine learning techniques. Seven classifiers, including Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Logistic Regression, Decision Tree, AdaBoost, Multilayer Perceptron (MLP), and Extra Tree Classifier, were trained to classify students' grades into four quality classes: Excellent, Good, Poor, and Fail. The weighted voting approach was employed to aggregate the outputs of these classifiers, resulting in an accuracy of 81.73%. The study aimed to assist students in taking proactive measures to improve their grades before final examinations and provided insights into predicting students' performance using diverse machine learning algorithms.

The study [16] focused on assisting undergraduate students in Bangladesh in selecting the most suitable department based on their prior academic performance. By developing a system that utilized students' previous data, the research aimed to predict which department would be best aligned with their future career aspirations. Utilizing SPSS for data analysis and WEKA for algorithm selection, the study concluded that the K-Nearest Neighbors (KNN) model was the most effective in predicting department suitability, achieving an accuracy of over 90% with proper training and labeling. The research addressed the common confusion among students regarding department selection after completing higher secondary education, aiming to provide a solution that could guide students in making informed decisions about their academic future.

Ahmed et al. [17] investigated the issue of dropout rates in engineering universities in Bangladesh by employing machine learning methodologies, including neural networks and educational data mining. Through the analysis of data collected from 480 students, they pinpointed significant predictors of dropout, which included factors such as living conditions, involvement in student politics, distance from the university, academic performance, and the level of institutional support. Their predictive model achieved an accuracy rate of 91.5%, surpassing those that relied exclusively on personal or academic information. This research provides valuable insights into the dynamics of student dropout and potential prevention strategies, highlighting the importance of addressing both academic and personal obstacles to enhance student retention rates.

Hemal et al. [18] investigate the influence of internet usage on students' academic performance in Noakhali, Bangladesh, employing machine learning algorithms for prediction. Findings reveal that students predominantly utilize the internet for educational and entertainment purposes, with a notable preference for academic browsing among school and university students. Those with higher academic achievement tend to focus on educational platforms, while others engage more with social media. The random forest algorithm emerges as the most effective, achieving an 85% accuracy rate. These results underscore the importance of policy interventions to encourage focused internet usage for enhanced educational outcomes.

Islam et al. [19] commenced their study by addressing the importance of Educational Data Mining (EDM) in uncovering meaningful insights from educational data, particularly in predicting students' performance in programming. They proposed an EDM paradigm aimed at classifying students' programming performance more accurately using real collected data. Through rigorous experimentation and analysis, they explored the effectiveness of feature engineering and ensemble machine learning techniques, focusing on the Random Forest (RF) classifier. The study's findings highlighted the RF classifier's impressive performance, achieving a prediction accuracy of 94%. This research contributed significantly to the field of educational data processing, offering valuable insights for educators and practitioners to support students in improving their programming skills.

Mia et al. [20] investigated the application of machine learning techniques to predict student registration status in the context of private universities in Bangladesh, considering factors like grades, due amounts, and enrollment status. Seven classifiers were tested, with Support Vector Machine (SVM) emerging as the most effective, achieving an accuracy of 85.76%. SVM demonstrated high sensitivity (98.83%) and precision (84.67%), outperforming other classifiers. The study underscored the significance of early prediction for enhancing university planning and sustainability, offering valuable insights for future research in optimizing enrollment strategies.

The study [21] predicted students' academic performance at an Indian technical institution using educational data mining techniques. Data was collected via a survey and academic records, followed by preprocessing and factor analysis. Machine learning algorithms, including multiple linear regression (MLR) and support vector regression (SVR), were compared. SVR with a linear kernel achieved the highest accuracy (83.44%), with MLR close behind. SVR with polynomial and radial basis function kernels had lower accuracy. The study highlights the effectiveness of data mining in predicting academic outcomes and emphasizes the role of past performance in forecasting future results.

Nieto et al. [22] investigated machine learning algorithms to support decision-making at Higher Educational Institutions (HEIs), focusing on predicting graduation rates of South American engineering students. Analyzing data from 6100 students over ten years, they compared three supervised classification algorithms. Random Forest achieved the highest accuracy (84.11%) and a recall rate of 91.93%, outperforming others. Logistic Regression had a slightly higher area under the curve (AUC). The study highlights machine learning's potential in improving decision-making, resource planning, and curriculum design at HEIs.

The study [23] aimed to predict students' difficulties in a digital design course using machine learning algorithms on data from the DEEDS TEL system. Various algorithms, including ANNs, SVMs, logistic regression, Naïve Bayes, and decision trees, were evaluated. Nine key variables correlated with session grades were identified. ANNs and SVMs achieved up to 80% accuracy in predicting student difficulties. Integrating these models into TEL systems could help identify struggling students early and improve teaching effectiveness.

Rois et al. [24] surveyed 355 Bangladeshi university students from twenty-eight institutions to predict stress prevalence using advanced machine learning (ML) techniques compared to logistic regression (LR). Significant prognostic factors, including pulse rate, blood pressure, sleep, smoking habits, and academic background, were identified using the Boruta algorithm. Results revealed a one-third stress prevalence rate within the last year. Among ML models, random forest (RF) performed best, with high accuracy (0.8972), precision (0.9241), sensitivity (0.9250), specificity (0.8148), and area under the ROC curve (0.8715). The study suggests ML frameworks improve stress prevalence prediction, aiding mental health promotion strategies and university counseling services.

Abuzinadah et al. [25] presented a machine learning framework that leverages deep convolutional features to forecast student academic performance, tackling issues prevalent in educational data mining, including low accuracy, data imbalance, and challenges in feature engineering. The research utilizes a tailored Convolutional Neural Network (CNN) for feature extraction and implements the Synthetic Minority Oversampling Technique (SMOTE) to balance the datasets. The findings indicate that the use of deep convolutional features significantly enhances prediction accuracy compared to the original features. The Extra Trees Classifier (ETC) achieved an impressive accuracy rate of 99.9%, surpassing existing advanced methodologies. This research is significant as it provides a reliable method for precisely forecasting student performance, enabling educational institutions to proactively assist students and decrease dropout rates.

Trivedi et al. [26] conducted a comprehensive study on student retention in higher education, focusing on the application of machine learning techniques, specifically Support Vector Machines (SVM) and Neural Networks. SVM demonstrated an accuracy of over 70% with modest misclassification, particularly excelling in identifying non-completers. Neural Networks showed even greater accuracy than SVM, especially in categorizing students into at-risk, intermediate, and advanced groups based on their GPA. Despite the study's limited sample size, the findings suggest promising potential for accurately forecasting student retention, with implications for understanding dropout risk and attrition factors.

AlGhamdi et al. [27] tackled the challenge of helping graduate students find suitable postgraduate universities using machine learning. They evaluated three regression models—Linear Regression, Decision Tree, and Logistic Regression—using Kaggle data. Logistic Regression proved the most accurate with the lowest Root Mean Square Error (RMSE) of 0.072. The study demonstrated machine learning's potential to enhance the university selection process and optimize admission chances. Future work could involve developing this algorithm into a software tool to better assist students in aligning their profiles with suitable universities.

Drawing inspiration from prior research in educational data mining and machine learning, this study explores the experiences of students in Bangladesh who, after failing to gain admission to public universities, opt for private institutions. Given that public universities are typically the preferred option in the country, this research aims to identify the reasons behind students' exclusion from public institutions and their transition to private universities. By investigating the obstacles these students encounter in fulfilling the requirements of public universities, the study seeks to shed light on the factors that shape their educational choices and trajectories, with an emphasis on promoting sustainable solutions. These findings aim to provide valuable insights to enhance access and support for students navigating the higher education system, ensuring long-term equity and sustainability in educational opportunities.

3. Dataset Description

In the highly competitive educational landscape of Bangladesh, where admission to public universities is fiercely sought after, many students inevitably consider private institutions as a secondary option. However, for most students, securing a spot in a public university remains the preferred choice. To gain a deeper understanding of the dynamics of university admissions in Bangladesh and the factors that influence students' success in securing a seat in public institutions, a comprehensive dataset known as the Undergraduate Admission Test Survey has been compiled. This dataset aims to provide analytical insights into the factors that contribute to students' inability to secure admission to public universities in Bangladesh. By examining various parameters such as academic performance, socioeconomic background, and other relevant factors within the context of Bangladesh's educational system, valuable insights are sought to unravel the complexities of university admissions in the country. The survey focuses on understanding the barriers that hinder access to public universities in Bangladesh and aims to provide actionable insights to policymakers, educators, and aspiring students.

3.1. Data Collection Process

A Google document has been created to facilitate a collaborative effort in exploring the various dimensions of student achievement and setbacks during the admission process. This document consists of 15 thought-provoking questions and has been shared with students from both public and private universities in Bangladesh. The main goal is to encourage widespread participation and gather a diverse range of perspectives on the admissions experience. By actively involving students from different educational backgrounds, this study aims to gain a comprehensive understanding of the numerous challenges and opportunities encountered throughout the complex journey of seeking admission. Anonymous survey data was collected from university students who participated voluntarily, with informed consent implied through their participation. At the beginning of the survey, participants were informed about the study's purpose, its voluntary nature, and the confidentiality of their responses. Table 1 presents a summary of the 15 attributes and their potential values. This collaborative endeavor aimed to deepen the understanding of admission dynamics by incorporating varied viewpoints and experiences.

Table 1: Attributes and their Possible Values

Id	Feature Explanation	Possible Value
1	SSC GPA	Any value between 0 and 5.00
2	HSC GPA	Any value between 0 and 5.00
3	Family's economic condition	Good, Medium, Average, Below Average
4	Location during exam preparation	Village, Town
5	Educational status of family	Educated, Uneducated, Highly Educated
6	Involvement in politics during exam preparation	Yes, No
7	Time spent on social media/other activities	0-1 Hour, 1-3 Hours, 3-5 Hours, More than 5 Hours
8	Stayed with family during exam preparation	Yes, No
9	Average duration of study per day during preparation	2-3 Hours, 3-5 Hours, 5-7 Hours, More than 7 Hours
10	Area of college	Village, Town
11	Area of school	Village, Town
12	Presence of bad habits like smoking/drug addiction	Yes, No
13	Involvement in any type of relationship	Yes, No
14	External factors affecting exam performance	Yes, No
15	Current institution	Institution name

Subsequently, the Google document yielded 634 responses, indicative of significant engagement from the student community. These responses were meticulously organized into a spreadsheet format to facilitate systematic analysis. Recognizing the need for further data processing, the spreadsheet data were converted into a CSV (Comma-Separated Values) file, which is a widely accepted format for handling large datasets. This conversion was essential as it enabled streamlined processing and efficient utilization of machine learning techniques, paving the way for deeper insights into the data. The CSV format not only simplified data handling but also allowed for seamless integration with various analytical tools and algorithms. This methodological approach ensured a rigorous examination of the collected data, leading to a more comprehensive understanding of the factors influencing student outcomes during the admission process.

3.2. Data Cleaning

After the Google document was widely distributed, a significant number of 634 submissions were received. However, in order to maintain the accuracy of the data, a careful data cleaning protocol was implemented to carefully review the influx of responses. This thorough procedure was designed to identify and correct any instances of misinformation or fraudulent entries, ensuring the reliability of the dataset. As a result, 38 responses were flagged as unreliable and subsequently removed, effectively preserving the integrity of subsequent analyses. The remaining 595 submissions were subjected to thorough examination and confirmed as legitimate, showcasing a wide array of authentic viewpoints and experiences concerning the admission process from various educational backgrounds. This carefully assembled data holds significant promise for application in machine learning algorithms, providing enhanced understanding of the intricacies involved in student admissions.

3.3. Description of Attributes

The study employed a dataset gathered through a Google Form survey, consisting of 15 columns and 595 rows, where each row corresponds to an individual student's response from both public and private universities. The columns encompassed a range of elements related to the admission process and student demographics, such as personal details, preferences, and perceptions. The objective of this analysis was to explore the factors that affect student choices and experiences in higher education.

SSC GPA & HSC GPA: The characteristics outlined reflect the academic achievements of participants in their SSC and HSC examinations, respectively. Among the 595 students, the SSC GPA varied from 2.99 to 5.00, while the HSC GPA ranged from 3.34 to 5.00. Elevated GPAs suggest a sustained commitment to academic excellence and proficient study techniques, which play a crucial role in influencing admission outcomes and future opportunities. In contrast, lower GPAs may indicate academic difficulties, which could restrict future possibilities.

Family's Economic Condition: This feature illustrates the socioeconomic backgrounds of respondents and their financial resources for education, emphasizing the inequalities in access and achievement. Students hailing from more affluent families frequently gain advantages through tutoring and extracurricular activities, whereas those from less privileged backgrounds may encounter obstacles such as the necessity of part-time employment and limited study materials. The retrieved data indicates that 39.0% of students originate from medium economic backgrounds, with 32.9% coming from average backgrounds, representing the largest segment.

Where Stayed During Exam Preparation: This attribute emphasizes the living conditions of respondents while they prepare for exams, differentiating between urban and rural environments. Students in urban areas typically enjoy greater access to resources such as libraries, tutoring services, and technology, which can improve their academic outcomes. Conversely, students in rural areas may encounter obstacles, including a scarcity of educational materials and infrastructural deficiencies. The gathered data indicates that 84.4% of students resided in town during their exam preparation, with the remaining students living in the village.

Educational Status of Family: This trait reflects the educational attainment within respondents' families, offering insights into the influence of family background on students' academic aspirations and support networks. Families with higher educational attainment often provide stronger support, fostering higher academic achievement. In contrast, students from less educated families may face different challenges and rely more on external resources. The data shows that 76.1% of students come from educated families, while 11.9% belong to either uneducated or highly educated families.

Involvement in Politics: Analyzing the political engagement of respondents during their exam preparation offers valuable insights into the influence of socio-political factors on academic performance. Participation in political activities may influence time management, stress levels, and overall well-being, which can disrupt study habits. The findings from the collected data indicate that 97.1% of students refrained from political involvement during their exam preparation, suggesting that the majority prioritize their academic obligations, whereas 2.9% were active in political pursuits.

Time Spent on Social Media/Other Activities: This characteristic assesses the involvement of respondents in non-academic pursuits, including social media usage, while preparing for examinations. It offers valuable insights into possible distractions or coping strategies that may influence academic outcomes. A high level of engagement could suggest distractions that detract from study time, although it may also function as a means of stress relief. The findings indicate that 43.2% of students dedicated 1-3 hours to social media during their exam preparation, making up the largest part.

Stayed with Family During Exam Preparation: This attribute shows whether respondents lived with their families during exam preparation. Family support can enhance students' study environment and emotional well-being, positively influencing academic performance. The compiled data reveals that 64.2% of students stayed with their families, benefiting from their support, while 35.8% lived elsewhere, facing potential challenges like increased responsibilities and reduced familial support.

Average Duration of Study in a Single Day: This aspect assesses the amount of time respondents dedicate to studying each day in preparation for exams, providing insights into their study practices and time management skills. An increased study duration typically signifies greater dedication and self-discipline, whereas a reduced amount may suggest difficulties in balancing responsibilities or a lower sense of urgency. The findings from the sourced data reveal that 34.1% of students engaged in more than 7 hours of study daily, while 33.1% devoted between 5 to 7 hours to their studies, occupying the largest section.

College Area & School Area: These attributes indicate the geographic distribution of the respondents' educational institutions, underscoring the disparities in educational resources between urban and rural environments. Typically, urban regions provide superior facilities and greater opportunities, thereby enriching the academic experience. The acquired data reveals that 91.3% of students attended colleges located in towns, whereas 67.6% attended schools situated in towns, suggesting a trend of students migrating from rural areas to urban centers for their higher education.

Bad Habits & Involvement in Relationships: These traits explore respondents' engagement in behaviors like smoking, drug use, and personal relationships during exam preparation. Such behaviour may divert attention from academic pursuits, hinder cognitive abilities, and contribute to stress, ultimately impacting academic success. The findings from the gathered data indicate that 6.4% of students acknowledged engaging in detrimental habits like smoking or drug use, while 22.5% were involved in personal relationships during their exam preparation period.

External Factors Affecting Performance: This characteristic assesses the impact of external influences such as personal difficulties, health-related issues, or financial obstacles on the examination performance of respondents. These stressors can distract attention and impede academic achievement by diminishing the time and energy allocated for studying. Furthermore, the emotional burden associated with these stressors may negatively affect cognitive abilities, thereby intensifying challenges in academic performance. The assembled data shows that 50.9% of students reported being impacted by these external factors during their exam preparation, while the remaining students did not.

Name of Current Institution: This attribute functions as the dependent variable, shaped by several independent factors such as academic performance and socioeconomic status. By classifying students' institutions as either public or private, we can examine the impact of these categories on academic results. The recorded data indicates that there are 369 students enrolled in public universities and 226 in private universities, resulting in a distribution of 62% for public institutions and 38% for private ones.

The dataset derived from the Undergraduate Admission Test Survey provides significant insights into the elements that impact student admissions to public universities in Bangladesh. Through the examination of academic performance, socioeconomic status, and various other factors, along with the application of machine learning and visualization methods, the dataset uncovers trends that affect admission results. This information contributes to academic discourse and offers practical recommendations for enhancing fairness and inclusivity in the admission processes.

4. Research Methodology

The methodology initiates by creating the dataset, which then proceeds through two distinct stages: one pathway involves addressing the class imbalance, while the other utilizes the unaltered dataset without handling the imbalance. Following this bifurcation, both pathways enter the data preprocessing stage, ensuring the data is clean and ready for model training. After preprocessing, a comprehensive classifier pipeline is defined, comprising 14 different classifiers, each accompanied by meticulously specified hyperparameter grids. The next step involves leveraging grid search in conjunction with 10-fold cross-validation to rigorously assess the performance of each classifier across various hyperparameter settings. This process ensures a robust evaluation of the models, enabling the identification of the most effective classifier configurations. Upon obtaining the performance metrics, the final stage involves a thorough evaluation of the classifiers, complemented by a detailed visualization of feature importance, which highlights the most influential features contributing to the model predictions. Figure 1 illustrates the workflow of the proposed methodology.

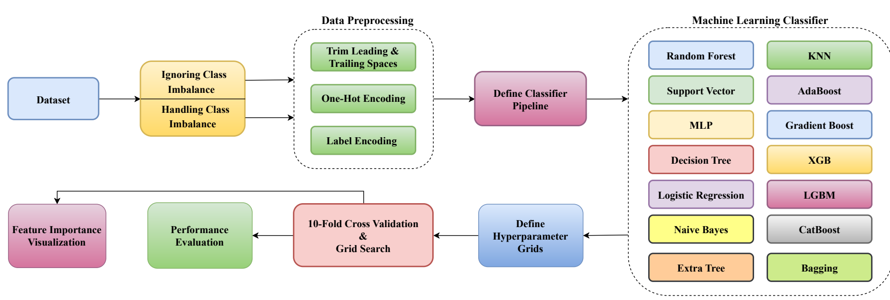


Figure 1: Workflow of the Proposed Methodology

4.1. Class Imbalance Handling

The dataset for the undergraduate admission test survey comprises 595 samples, with 226 pertaining to private universities and 369 to public universities, highlighting a notable class imbalance. This disparity can distort model performance, leading to a preference for the majority class (public universities) at the expense of the minority class (private universities). Consequently, this may diminish predictive accuracy and overlook essential patterns within the minority class. Such imbalances impede the model's capacity to generalize effectively. To mitigate this issue, two main techniques are utilized: oversampling and undersampling. Oversampling enhances the representation of the minority class by either duplicating existing samples or creating new ones through methods such as SMOTE, thereby enabling the model to learn from the minority class's patterns. In contrast, undersampling decreases the size of the majority class by randomly eliminating samples, thereby aligning it more closely with the minority class to avert bias towards the majority class.

This study utilizes both oversampling and undersampling methods to tackle the issue of class imbalance within the undergraduate admission test survey dataset. Initially, the 'RandomOverSampler' technique is employed to augment the number of samples in each class to 500, thereby ensuring a balanced representation between private and public universities. This approach alleviates the bias towards the majority class by replicating instances from the minority class. Furthermore, a strategy is implemented where the majority class (public universities) is reduced to 226 instances through random selection to match the minority class (private universities), which also consists of 226 instances. This balanced dataset, with 226 instances per class, mitigates the original class imbalance, ultimately leading to the creation of a more robust and generalizable model. Both methods are assessed not only to evaluate their effectiveness but also to compare their performance against the unbalanced dataset, thereby identifying which strategy yields superior model performance and generalization.

4.2. Data Preprocessing

Preprocessing is crucial in data analysis and machine learning because it lays the foundation for accurate and reliable results. By cleaning and preparing the data before feeding it into algorithms, we ensure that the information is consistent, free from errors, and suitable for analysis. Without proper preprocessing, the data may contain inconsistencies, missing values, or irrelevant information, which can significantly impact the performance and reliability of machine learning models. In the preprocessing phase of this dataset, several critical steps were undertaken to enhance data quality and suitability for analysis and machine learning tasks.

The initial task involved cleaning the target column by removing any leading and trailing spaces. This step is essential because extraneous whitespace can lead to inconsistencies in data interpretation and processing. For example, the entries "Private University " (with a trailing space) and "Private University" (without a trailing space) would be treated as different categories by the algorithm, potentially skewing the results. Therefore, by standardizing the format of the data and removing unnecessary whitespace, uniformity, and accuracy are ensured in subsequent analyses.

Next, one-hot encoding was utilized to transform categorical variables such as "Where did you stay while preparing for the exam?" and "What was the educational background of your family?" into a format that is compatible with machine learning algorithms. This process is vital, as these algorithms generally require numerical input, and using categorical variables in their original form can lead to complications. One-hot encoding generates binary columns for each category within a variable, thereby preventing any erroneous ordinal relationships from being established. For instance, the categories "Town" and "Village" within the "Where did you stay?" variable are converted into separate binary columns. In the absence of this encoding, the model could mistakenly interpret an ordinal relationship among the categories, resulting in skewed outcomes.

Label encoding plays a vital role in the preprocessing of ordinal variables, such as the question regarding a family's economic condition, which includes categories like "Good," "Medium," and "Average." This technique assigns a distinct integer to each category according to its rank, thereby preserving the natural order within the dataset. It enables the machine learning model to recognize the hierarchy among the categories, acknowledging that "Good" is superior to "Average," which is in turn ranked above "Medium." In the absence of label encoding, the model may mistakenly treat ordinal variables as nominal, overlooking their sequential nature and potentially leading to erroneous interpretations. This misinterpretation could cause the model to incorrectly equate "Medium" with or even elevate it above "Good," thereby skewing the analysis. Consequently, label encoding is indispensable for upholding the ordinal characteristics of variables.

The preprocessing steps encompassed addressing potential inconsistencies and ensuring proper formatting for each variable. For instance, the variable "How much time did you spend on social media/other activities while preparing for the exam?" contained categories like "0-1 Hour" and "1-3 Hours," which were standardized to ensure uniformity and prevent misinterpretation by algorithms. Similarly, the variable "Did you stay with your family while preparing for the exam?" was encoded to ensure binary responses ("Yes" or "No") were accurately interpreted. This preprocessing ensured that the dataset was ready for effective analysis by mitigating errors, enhancing model accuracy, and ensuring the overall reliability of the data. By carefully executing these preprocessing tasks, the groundwork was laid for building robust and insightful machine-learning models. These models can now effectively interpret and predict outcomes based on the given dataset.

4.3. Classifier Pipeline Definition and Hyperparameter Grid Specification

Creating a strong machine learning model necessitates a meticulously organized classifier pipeline and precisely outlined hyperparameter grids, both of which are critical for achieving reliable research results. The classifier pipeline streamlines the data processing stages and model training, incorporating feature scaling, dimensionality reduction, and training into a unified framework that promotes reproducibility and scalability. Concurrently, establishing hyperparameter grids facilitates systematic optimization across different parameter configurations, enabling algorithms to identify intricate patterns and generalize effectively. This iterative tuning process enhances predictive accuracy, mitigates the risks of overfitting or underfitting, and fortifies the model's robustness and generalization abilities.

This study utilizes a diverse array of classifiers, comprising 14 unique models, aimed at effectively capturing the distinctive features of the dataset. The models range from traditional methods such as Logistic Regression and Decision Trees to more sophisticated ensemble techniques like Random Forests and Gradient Boosting. These models are organized within pipelines that integrate preprocessing steps, which include one-hot encoding for categorical variables and the standardization of numerical features. This structured pipeline framework guarantees uniform data processing and training across the classifiers, thereby optimizing the workflow and improving reproducibility.

For each classifier, carefully constructed hyperparameter grids delineate the values to be investigated during the grid search optimization process. Table 2 presents the hyperparameter grids for the various classifiers used in this study. This approach facilitates an in-depth analysis of the hyperparameter space, aimed at identifying the optimal settings for achieving peak model performance. Serving as essential tools in the optimization journey, these grids offer meticulous control over model functionality and efficacy, thereby improving generalization and mitigating the risks of overfitting. Their flexibility ensures that the optimization process is tailored to the unique characteristics of the dataset and the specific requirements of the research, promoting efficient and resilient model development. Collectively, classifier pipelines and hyperparameter grids support a methodical exploration and enhancement of machine learning models.

Table 2: **Hyperparameter Grids for Various Classifiers**

Classifier	**Parameters & Values**
RandomForestClassifier	`n_estimators`: [50, 100, 200]; `max_depth`: [None, 10, 20]; `min_samples_split`: [2, 5, 10]; `min_samples_leaf`: [1, 2, 4]
SVC	`C`: [0.1, 1, 10]; `kernel`: ['linear', 'rbf', 'poly']; `gamma`: ['scale', 'auto']
MLPClassifier	`hidden_layer_sizes`: [(50,), (100, 50), (100, 50, 20)]; `max_iter`: [200, 500, 1000]; `alpha`: [0.0001, 0.001, 0.01]
DecisionTreeClassifier	`max_depth`: [None, 5, 10, 20]; `min_samples_split`: [2, 5, 10]; `min_samples_leaf`: [1, 2, 4]
LogisticRegression	`C`: [0.1, 1, 10]; `max_iter`: [50, 100, 200]
KNeighborsClassifier	`n_neighbors`: [3, 5, 7]; `weights`: ['uniform', 'distance']
AdaBoostClassifier	`n_estimators`: [50, 100, 200]; `learning_rate`: [0.1, 0.5, 1]
GradientBoostingClassifier	`n_estimators`: [50, 100, 200]; `learning_rate`: [0.01, 0.1, 0.5]; `max_depth`: [3, 5, 10]
XGBClassifier	`n_estimators`: [50, 100, 200]; `learning_rate`: [0.01, 0.1, 0.5]; `max_depth`: [3, 5, 10]
LGBMClassifier	`n_estimators`: [50, 100, 200]; `learning_rate`: [0.01, 0.1, 0.5]; `max_depth`: [3, 5, 10]
CatBoostClassifier	`n_estimators`: [50, 100, 200]; `learning_rate`: [0.01, 0.1, 0.5]; `max_depth`: [3, 5, 10]
ExtraTreesClassifier	`n_estimators`: [50, 100, 200]; `max_depth`: [None, 5, 10, 20]; `min_samples_split`: [2, 5, 10]; `min_samples_leaf`: [1, 2, 4]
BaggingClassifier	`n_estimators`: [50, 100, 200]; `max_samples`: [0.5, 0.7, 1.0]; `max_features`: [0.5, 0.7, 1.0]

4.4. Grid Search Optimization

In this research, Grid Search is a key tool for fine-tuning machine learning models through the systematic exploration of predefined hyperparameter sets for each classifier in the established pipeline. By evaluating various combinations of hyperparameters within specified ranges, Grid Search seeks to identify configurations that maximize performance metrics such as accuracy, precision, recall, and F1-score. This approach enables a thorough assessment of classifiers across diverse hyperparameter settings, ensuring optimal performance tailored to the dataset's characteristics. The optimization process utilizes the hyperparameter grids outlined in Table 2, detailing the parameter values for each classifier. This structured investigation supports a rigorous search, enhancing adaptability and performance. By leveraging Grid Search, iterative refinements can be made, uncovering the hyperparameter settings that yield the most effective predictive outcomes.

Grid Search serves a dual purpose: it not only determines the most effective hyperparameters but also sheds light on the relationships between various hyperparameters and their impact on model performance, clarifying the operational characteristics of each algorithm. This process of iterative optimization improves predictive accuracy while bolstering the robustness and generalization abilities of the model. By incorporating cross-validation methods, Grid Search reduces the likelihood of overfitting by validating against unseen data, thereby ensuring that the chosen hyperparameter configurations are capable of generalizing well. This thorough investigation of the hyperparameter landscape is fundamental to our approach, ultimately enhancing the effectiveness of the modeling process and fostering progress in machine learning methodologies within this research.

4.5. Cross-Validation for Rigorous Evaluation

Cross-Validation assumes a critical role in rigorously assessing the performance of the classifiers and validating their generalization capabilities. By partitioning the dataset into multiple subsets and iteratively training and testing the models on different combinations of these subsets, Cross-Validation offers a more accurate estimate of the model's performance compared to traditional train-test splits. This technique is instrumental in mitigating the risk of overfitting, as it evaluates the model's performance on multiple independent subsets of the data.

In this research, 10-fold Cross-Validation is employed to assess the performance of each classifier under different hyperparameter configurations. The dataset is partitioned into 10 equal-sized segments, facilitating a systematic approach to model training and evaluation, which leads to more accurate performance assessments. This technique promotes robustness and enhances the models' ability to generalize to previously unseen data by testing them on varied subsets. Furthermore, Cross-Validation evaluates the consistency of performance metrics across different data divisions, offering valuable insights into the stability and reliability of the models. It also plays a significant role in hyperparameter optimization by pinpointing configurations that consistently deliver high performance across multiple folds. This iterative approach improves generalization capabilities, ensuring dependable predictions on new data. In summary, Cross-Validation is an essential aspect of the research methodology, allowing for a thorough evaluation of model performance and instilling confidence in the reliability and generalization capabilities of the developed machine learning models.

4.6. Performance Evaluation Metrics

In assessing the effectiveness of the classifiers, various performance evaluation metrics are employed, including accuracy, precision, recall, F1 score, macro average, and weighted average. Each metric provides valuable insights into different aspects of the models' predictive capabilities.

- **Accuracy**: Accuracy measures the proportion of correctly classified instances out of the total instances in the dataset. It is calculated as:

$$\text{Accuracy} = \frac{\text{Number of correctly classified instances}}{\text{Total number of instances}} \tag{1}$$

- **Precision**: Precision quantifies the ratio of correctly predicted positive observations to the total predicted positive observations. It focuses on the accuracy of positive predictions. Precision is calculated as:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{2}$$

- **Recall**: Recall, also known as sensitivity or true positive rate, measures the ratio of correctly predicted positive observations to all actual positive observations in the dataset. It assesses the classifier's ability to identify all relevant instances. Recall is calculated as:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{3}$$

- **F1 Score**: F1 score is the harmonic mean of precision and recall. It provides a balance between precision and recall and is particularly useful when there is an imbalance between the number of positive and negative instances. F1 score is calculated as:

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

- **Macro Average**: Macro average computes the average of the metrics (accuracy, precision, recall, F1 score) for each class without considering class imbalance. It is calculated as the arithmetic mean of the individual class scores.

$$\text{Macro Average} = \frac{1}{N} \sum_{i=1}^{N} \text{Metric}_i \tag{5}$$

  Where:

  - N is the total number of classes.
  - Metric_i represents the performance metric (e.g., accuracy, precision, recall, F1 score) for the i^{th} class.

- **Weighted Average**: Weighted average computes the average of the metrics (accuracy, precision, recall, F1 score) for each class while taking into account class imbalance. It is calculated as the weighted sum of the individual class scores, where the weights are proportional to the number of instances in each class.

$$\text{Weighted Average} = \frac{\sum_{i=1}^{N}(\text{Metric}_i \times \text{Instances}_i)}{\sum_{i=1}^{N}\text{Instances}_i} \tag{6}$$

Where:

- N is the total number of classes.
- Metric_i represents the performance metric (e.g., accuracy, precision, recall, F1 score) for the i^{th} class.
- Instances_i is the number of instances in the i^{th} class.

4.7. Experiment Setup and Tools

The experiment utilizing Python-3 was carried out solely on Kaggle Notebooks, leveraging its features for machine learning support. Essential libraries incorporated included pandas for data manipulation, scikit-learn for algorithm implementation, and seaborn and matplotlib for data visualization. The preprocessing phase involved the use of StandardScaler, LabelEncoder, ColumnTransformer, and OneHotEncoder to effectively manage data transformations. The experimental setup was supported by robust hardware, featuring 13GB of RAM and a 16GB P100 GPU, which facilitated the execution of computationally demanding tasks. A diverse array of classifiers was utilized, including RandomForestClassifier, SVC, and sophisticated ensemble models such as XGBClassifier, LGBMClassifier, and CatBoostClassifier, with GridSearchCV employed for hyperparameter optimization. To address class imbalance, StratifiedKFold cross-validation and RandomOverSampler from the imblearn library were implemented, with cross_val_score and cross_val_predict aiding in evaluation, while the classification_report provided essential performance metrics. Visualizations generated with seaborn and matplotlib enhanced result interpretability, enabling a thorough analysis of undergraduate admission test survey data within a streamlined and effective experimental framework.

5. Results & Discussion

5.1. Performance Analysis Without Sampling

This section presents a performance analysis of various classifiers on the given dataset without applying any sampling techniques. The objective is to evaluate and compare the effectiveness of different machine learning classifiers using cross-validation and grid search for hyperparameter tuning. This analysis provides insights into the inherent capabilities of each classifier when applied to the original, unaltered data. It's noteworthy that the dataset comprises a total of 595 instances, with 226 belonging to the private university class and 369 to the public university class.

Table 3: Classification Report for Various Classifiers on the Unaltered Dataset

Classifier	Class	Precision (%)	Recall (%)	F1-Score (%)	Support
Random Forest	**Private University**	82	71	76	226
	Public University	84	91	87	369
	Accuracy (%)		83		595
	Macro Avg (%)	83	81	82	595
	Weighted Avg (%)	83	83	83	595
Support Vector	**Private University**	79	59	68	226
	Public University	78	90	84	369
	Accuracy (%)		79		595
	Macro Avg (%)	78	75	76	595
	Weighted Avg (%)	78	79	77	595
Multi-layer Perceptron	**Private University**	71	63	67	226
	Public University	79	84	82	369
	Accuracy (%)		76		595
	Macro Avg (%)	75	74	74	595
	Weighted Avg (%)	76	76	76	595
Decision Tree	**Private University**	77	68	72	226
	Public University	82	88	85	369
	Accuracy (%)		80		595
	Macro Avg (%)	80	78	79	595
	Weighted Avg (%)	80	80	80	595
Logistic Regression	**Private University**	79	60	68	226
	Public University	79	90	84	369
	Accuracy (%)		79		595
	Macro Avg (%)	79	75	76	595
	Weighted Avg (%)	79	79	78	595
Naive Bayes	**Private University**	72	54	62	226
	Public University	76	87	81	369
	Accuracy (%)		75		595
	Macro Avg (%)	74	71	71	595
	Weighted Avg (%)	74	75	74	595
K-Nearest Neighbors	**Private University**	72	51	60	226
	Public University	75	88	81	369
	Accuracy (%)		74		595
	Macro Avg (%)	74	70	70	595
	Weighted Avg (%)	74	74	73	595
AdaBoost	**Private University**	85	72	78	226
	Public University	84	92	88	369
	Accuracy (%)		**84**		595
	Macro Avg (%)	84	82	83	595
	Weighted Avg (%)	84	84	84	595

Tables 3 and 4 present a comprehensive evaluation of various machine learning classifiers on the 'Undergraduate Admission Test Survey' dataset, which includes features relevant to university admission eligibility. This analysis emphasizes the metrics of precision, recall, F1-score, and accuracy, supplemented by both weighted and macro averages for enhanced understanding. The weighted average reflects overall accuracy while taking class distribution into account, whereas the macro average indicates performance balance between Public and Private University categories. The classifiers examined include ensemble methods such as Random Forest, Bagging, and AdaBoost, as well as advanced techniques like SVM, MLP, and Gradient Boosting, alongside traditional approaches including Decision Tree, Logistic Regression, Naive Bayes, and KNN. This diverse selection facilitates a comprehensive comparison between simpler, more interpretable methods and their more complex counterparts, with a total of 14 classifiers evaluated. The findings provide a nuanced perspective on how each classifier identifies patterns within the admission dataset, highlighting their unique strengths and limitations.

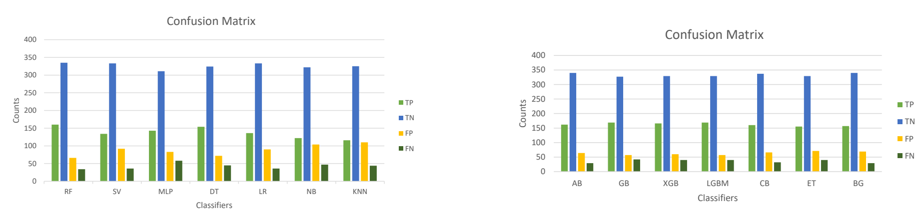
Table 4: Classification Report for Various Classifiers on the Unaltered Dataset Contd.

Classifier	Class	Precision (%)	Recall (%)	F1-Score (%)	Support
XGBoost	**Private University**	81	73	77	226
	Public University	85	89	87	369
	Accuracy (%)		83		595
	Macro Avg (%)	83	81	82	595
	Weighted Avg (%)	83	83	83	595
Gradient Boosting	**Private University**	80	75	77	226
	Public University	85	89	87	369
	Accuracy (%)		83		595
	Macro Avg (%)	83	82	82	595
	Weighted Avg (%)	83	83	83	595
Light Gradient Boosting Machine	**Private University**	81	75	78	226
	Public University	85	89	87	369
	Accuracy (%)		**84**		595
	Macro Avg (%)	83	82	82	595
	Weighted Avg (%)	84	84	84	595
CatBoost	**Private University**	83	71	77	226
	Public University	84	91	87	369
	Accuracy (%)		**84**		595
	Macro Avg (%)	83	81	82	595
	Weighted Avg (%)	84	84	83	595
Extra Trees	**Private University**	79	69	74	226
	Public University	82	89	86	369
	Accuracy (%)		81		595
	Macro Avg (%)	81	79	80	595
	Weighted Avg (%)	81	81	81	595
Bagging	**Private University**	84	69	76	226
	Public University	83	92	87	369
	Accuracy (%)		**84**		595
	Macro Avg (%)	84	81	82	595
	Weighted Avg (%)	84	84	83	595

The standout performers in the analysis are AdaBoost, LightGBM, CatBoost, and Bagging, each achieving an accuracy of 84%. These ensemble methods excel in capturing intricate patterns within the dataset by combining multiple weak learners into a strong predictor. AdaBoost demonstrates high precision and recall for Private (85% and 72%) and Public (84% and 92%) Universities, showcasing its robustness and reliability. Its adaptive reweighting of misclassified instances enhances model performance iteratively. Bagging shows similar strengths, with an 84% recall for "Public University" and 69% for "Private University," resulting in F1-scores of 87% and 76%. Bagging increases stability and accuracy by aggregating predictions from models trained on different data subsets, reducing variance. CatBoost and LightGBM also maintain high performance, with recall rates of 91% and 89% for "Public University," and 71% and 75% for "Private University," respectively. LightGBM, with similar metrics, excels in handling class imbalances and enhancing predictive accuracy, optimized for speed and efficiency in large datasets. CatBoost stands out for its superior handling of categorical features and reduced overfitting. These ensemble methods effectively manage diverse data characteristics and mitigate issues like overfitting and class imbalance, making them highly effective for this dataset.

Following closely, Random Forest, Gradient Boosting, and XGBoost each achieve an accuracy of 83%, demonstrating robust performance, especially in identifying "Public University" cases, with recall rates of 89-91%. However, their performance for "Private University" cases is slightly lower, with recall rates ranging from 71-75%. This decrease may be due to class imbalance or the complex patterns associated with this category. Ensemble methods excel by capturing interactions between features that simpler models may overlook. In contrast, classifiers like Decision Tree, SVM, Logistic Regression, MLP, Naive Bayes, and K-Nearest Neighbors (KNN) generally perform less effectively, with Decision Tree attaining 80% accuracy but showing inconsistency in results. SVM and Logistic Regression both reach 79% accuracy, yet they struggle with "Private University" recall (59-60%), reflecting challenges in addressing class imbalances. MLP, with 76% accuracy, also indicates a significant recall disparity between classes, suggesting room for improvement. Finally, Naive Bayes and KNN are the least effective, with accuracies of 75% and 74%, respectively, particularly faltering in their performance for the "Private University" class. These results underscore the limitations of simpler models in capturing the complexities of the data, highlighting the advantages of more sophisticated ensemble methods.

5.1.1. In-depth Confusion Matrix Analysis



(a) Confusion matrix showing classifier performance

(b) Confusion matrix showing classifier performance

Figure 2: Confusion matrices illustrating the performance of different classifiers on the dataset

In Figure 2a and 2b, the confusion matrix for the unaltered undergraduate admission test survey dataset offers an overall comparison of the performance of all 14 classifiers. The matrix provides a detailed breakdown of predictions across various classes, where each row represents actual class labels and each column signifies predicted outcomes. True Positives (TP) denote correct positive predictions, True Negatives (TN) indicate correct negative predictions, False Positives (FP) represent incorrect positive predictions, and False Negatives (FN) highlight incorrect negative predictions. Understanding these components is pivotal for evaluating model performance, with TP and TN reflecting accurate predictions, while FP and FN revealing misclassifications. Moreover, the confusion matrix complements and validates other evaluation metrics such as accuracy, precision, and recall, thereby furnishing a comprehensive understanding of the predictive capabilities of each classifier.

The confusion matrix presented provides a comprehensive assessment of various classifiers, revealing distinct performance trends across several metrics. AdaBoost (AB), XGBoost (XGB), LightGBM (LGBM), and CatBoost (CB) stand out as leaders, consistently recording higher numbers of True Positives (TP) and True Negatives (TN). For example, Gradient Boosting (GB) records 169 TP instances, with LGBM and XGB closely following at 166 and 169 TP instances, respectively. These leading classifiers also show strong TN counts, underscoring their proficiency in accurately identifying negative instances. Conversely, classifiers such as Random Forest (RF), Support Vector Machine (SV), and Multi-Layer Perceptron (MLP) demonstrate balanced performance across TP, TN, False Positives (FP), and False Negatives (FN) counts. For instance, RF achieves 160 TP and 335 TN instances, illustrating its dependability in accurately classifying both positive and negative cases.

Logistic Regression (LR), Naive Bayes (NB), and K-Nearest Neighbors (KNN) exhibit moderate performance, with lower TP and TN counts compared to the leading classifiers; LR records 136 TP and 333 TN instances, indicating capability but lower accuracy than the top contenders. Classifiers like Extra Trees (ET) mirror Gradient Boosting with moderate TP and TN counts but higher FP and FN counts; ET registers 155 TP and 329 TN instances, showcasing competence yet a slightly elevated false classification rate. This analysis aligns with prior evaluations, reaffirming the superior performance of AdaBoost, XGBoost, LightGBM, CatBoost, and Bagging classifiers, supported by their higher TP and TN counts across various classification tasks. While RF, SV, and MLP demonstrate stable performance, they do not surpass the top classifiers, and LR, NB, and KNN consistently show moderate effectiveness.

5.1.2. ROC Curve Analysis

Figure 3 presents a visual representation of receiver operating characteristic (ROC) curves, a widely recognized method for evaluating the performance of machine learning classifiers. These curves plot the true positive rate (TPR) against the false positive rate (FPR), where TPR reflects the proportion of actual positive cases correctly identified, and FPR indicates the proportion of negative cases misclassified as positive. Each curve corresponds to a different classifier, and the curve's shape and distance from the diagonal line illustrate the classifier's discriminatory power. Alongside each curve is the Area Under the Curve (AUC) value, providing a quantitative measure for assessing classifier performance; a higher AUC value denotes better capability in distinguishing between positive and negative instances.

The ROC curve analysis reveals notable performance trends among classifiers assessed on the oversampled dataset. The Random Forest, Decision Tree, XGBoost, Extra Trees, and Bagging classifiers achieve a perfect AUC score of 1.00, indicating exceptional discrimination between positive and negative instances. Following closely are the Multi-Layer Perceptron (MLP), LightGBM (LGBM), and CatBoost classifiers, with AUC scores of 0.99 and 0.98, respectively, showcasing their strong discriminative abilities. Utilizing advanced ensemble techniques, these classifiers effectively capture complex patterns within

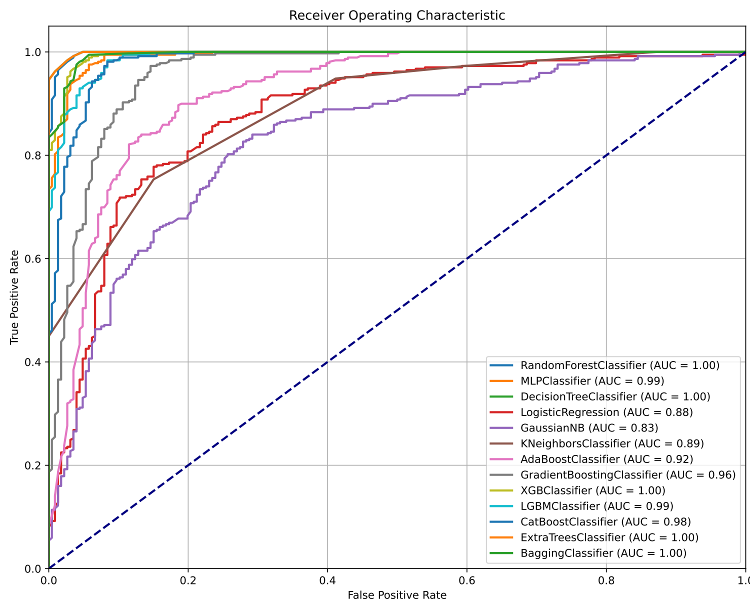


Figure 3: ROC curve comparison for various classifiers on the dataset

the dataset, making them highly reliable for classification tasks. The Gradient Boosting classifier exhibits impressive performance, achieving an AUC score of 0.96, whereas AdaBoost secures a commendable score of 0.92. In comparison, KNeighbors and Logistic Regression present lower AUC scores of 0.89 and 0.88, respectively, while Gaussian Naive Bayes lags behind with an AUC score of 0.83, indicating a diminished ability to differentiate between the two classes.

The Bagging classifier has achieved the highest accuracy of 84% and an impeccable AUC of 1.00. In comparison, the AdaBoost classifier also reaches an accuracy of 84%, but with a slightly lower AUC of 0.92. The distinction between accuracy and AUC arises from the fundamental nature of these metrics; accuracy reflects the overall correctness of predictions, whereas AUC assesses the classifier's capability to rank instances according to predicted probabilities. A high accuracy coupled with a lower AUC may indicate difficulties in effectively ranking instances, even though most instances are classified correctly. On the other hand, a high AUC signifies a strong ability to differentiate between classes, even if the overall accuracy is marginally reduced. Therefore, while accuracy offers a general evaluation of performance, AUC provides more nuanced insights into the classifier's discriminatory power. Following a thorough analysis of various metrics, including precision, recall, F1-score, accuracy, confusion matrices, and ROC curves, the Bagging classifier stands out as the most suitable option for tackling the challenges presented by the undergraduate admission test survey dataset, demonstrating excellence in both prediction accuracy and discriminatory capability.

5.2. Performance Analysis with Undersampling

This section presents the performance analysis of various classifiers on the dataset after applying undersampling technique. To address class imbalance, undersampling is applied, and its impact on performance metrics is thoroughly examined. Specifically, undersampling is achieved by reducing the number of public university instances to match the number of private university instances which is 226, given the larger sample size of public universities. This analysis provides insights into the performance of each classifier when trained on a balanced subset of the dataset, highlighting the effects of undersampling on predictive accuracy.

Table 5: Classification Report for Various Classifiers on the Undersampled Balanced Dataset

Classifier	Class	Precision (%)	Recall (%)	F1-Score (%)	Support
Random Forest	**Private University**	86	93	90	226
	Public University	84	91	87	226
	Accuracy (%)		89		452
	Macro Avg (%)	90	89	89	452
	Weighted Avg (%)	90	89	89	452
Support Vector	**Private University**	82	88	85	226
	Public University	87	81	83	226
	Accuracy (%)		84		452
	Macro Avg (%)	84	84	84	452
	Weighted Avg (%)	84	84	84	452
Multi-layer Perceptron	**Private University**	83	88	85	226
	Public University	88	81	84	226
	Accuracy (%)		85		452
	Macro Avg (%)	85	85	85	452
	Weighted Avg (%)	85	85	85	452
Decision Tree	**Private University**	84	86	85	226
	Public University	86	84	85	226
	Accuracy (%)		85		452
	Macro Avg (%)	85	85	85	452
	Weighted Avg (%)	85	85	85	452
Logistic Regression	**Private University**	84	79	81	226
	Public University	80	85	82	226
	Accuracy (%)		82		452
	Macro Avg (%)	82	82	82	452
	Weighted Avg (%)	82	82	82	452
Naive Bayes	**Private University**	81	68	74	226
	Public University	73	85	78	226
	Accuracy (%)		76		452
	Macro Avg (%)	77	76	76	452
	Weighted Avg (%)	77	76	76	452
K-Nearest Neighbors	**Private University**	81	85	83	226
	Public University	84	81	82	226
	Accuracy (%)		83		452
	Macro Avg (%)	83	83	83	452
	Weighted Avg (%)	83	83	83	452

Tables 5 and 6 provide a comprehensive evaluation of 14 machine learning classifiers applied to the balanced 'Undergraduate Admission Test Survey' dataset using undersampling techniques. Among these, Bagging, Random Forest, XGBoost, and CatBoost stand out with top accuracy scores between 89% and 90%, showcasing their capacity to handle complex data and capture intricate patterns effectively. Bagging achieves the highest accuracy at 90%, with notable precision rates of 86% for Private University and 94% for Public University, along with recall rates of 95% and 85%, yielding strong F1-scores of 90% and 89%, and consistent macro and weighted averages of 90%. Random Forest closely follows, reaching an accuracy of 89%, with high precision (86% for Private, 84% for Public), recall (93% for Private, 91% for Public), and F1-scores (90% for Private, 87% for Public), resulting in balanced macro and weighted averages of 89%. XGBoost and CatBoost also achieve 89% accuracy, with slightly varied precision, recall, and F1-scores. XGBoost exhibits 86% precision for Private University and 92% for Public, recall rates of 92% and 85%, and F1-scores of 89% and 88%. CatBoost's metrics closely match, with macro and weighted averages consistently at 89%. These classifiers excel due to their ensemble or boosting frameworks, enabling effective learning and generalization, which translates into superior performance on the balanced dataset.

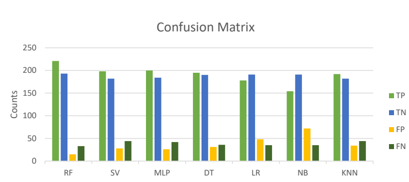
Table 6: Classification Report for Various Classifiers on the Undersampled Balanced Dataset Contd.

Classifier	Class	Precision (%)	Recall (%)	F1-Score (%)	Support
XGBoost	**Private University**	86	92	89	226
	Public University	92	85	88	226
	Accuracy (%)	89			452
	Macro Avg (%)	89	89	89	452
	Weighted Avg (%)	89	89	89	452
AdaBoost	**Private University**	87	78	83	226
	Public University	80	88	84	226
	Accuracy (%)	83			452
	Macro Avg (%)	84	83	83	452
	Weighted Avg (%)	84	83	83	452
Gradient Boosting	**Private University**	86	90	88	226
	Public University	89	85	87	226
	Accuracy (%)	88			452
	Macro Avg (%)	88	88	88	452
	Weighted Avg (%)	88	88	88	452
Light Gradient Boosting Machine	**Private University**	86	92	89	226
	Public University	92	85	88	226
	Accuracy (%)	87			452
	Macro Avg (%)	87	87	87	452
	Weighted Avg (%)	87	87	87	452
CatBoost	**Private University**	86	93	89	226
	Public University	92	85	88	226
	Accuracy (%)	89			452
	Macro Avg (%)	89	89	89	452
	Weighted Avg (%)	89	89	89	452
Extra Trees	**Private University**	84	90	87	226
	Public University	89	83	86	226
	Accuracy (%)	87			452
	Macro Avg (%)	87	87	86	452
	Weighted Avg (%)	87	87	86	452
Bagging	**Private University**	86	95	90	226
	Public University	94	85	89	226
	Accuracy (%)	**90**			452
	Macro Avg (%)	90	90	90	452
	Weighted Avg (%)	90	90	90	452

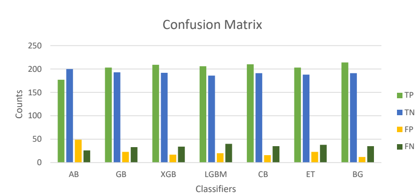
Gradient Boosting achieves an accuracy of 88%, with strong precision and recall for Private (86% precision, 90% recall) and Public Universities (89% precision, 85% recall), leading to F1-scores of 88% and 87% and balanced macro and weighted averages of 88%. LightGBM, though slightly lower at 87% accuracy, performs robustly with comparable precision and recall, maintaining macro and weighted averages at 87%. The success of these models stems from their gradient boosting techniques, which iteratively improve predictions by focusing on previously misclassified instances, enhancing overall accuracy. In contrast, Support Vector Machines (SVM), Multi-layer Perceptron (MLP), and Decision Tree classifiers achieve similar performance with accuracies around 84-85%. SVM reaches an accuracy of 84%, balancing precision and recall between 82-87% for both classes. MLP and Decision Tree attain 85% accuracy, with precision, recall, and F1-scores closely aligned between 83-88%. These models effectively manage non-linear relationships and complex decision boundaries, which are essential for accurate class predictions within this dataset.

Moderate performance is seen with Logistic Regression and K-Nearest Neighbors (KNN) classifiers, achieving accuracies of 82% and 83%, respectively. Logistic Regression has a precision of 84% for Private University and 80% for Public University, with recall rates of 79% and 85%. KNN maintains balanced precision and recall between 81-85%. These classifiers, while effective for simpler datasets and linear relationships, tend to underperform with complex or non-linear data patterns. AdaBoost also shows similar results, with 83% accuracy and balanced precision and recall across classes, leading to macro and weighted averages around 83-84%, though its sensitivity to noise and outliers can hinder generalization. Naive Bayes, however, has the lowest performance at 76% accuracy, particularly struggling with the Private University class (81% precision, 68% recall) but faring better for Public University (73% precision, 85% recall). Its assumption of feature independence often reduces effectiveness when features are correlated or relationships are complex. This comparative analysis underscores the strengths and weaknesses of each classifier, providing clear insights into their relative effectiveness on the balanced dataset. Notably, the Bagging classifier delivers superior results, similar to its performance on the unaltered dataset, but with significantly improved outcomes on the balanced dataset.

5.2.1. In-depth Confusion Matrix Analysis



(a) Confusion matrix showing classifier performance



(b) Confusion matrix showing classifier performance

Figure 4: Confusion matrices illustrating the performance of different classifiers on the undersampled dataset

The confusion matrices in Figures 4a and 4b provide a clear view of classifier performance on the balanced 'Undergraduate Admission Test Survey' dataset. Random Forest (RF) shows strong predictive accuracy, with 221 true positives

(TP) and 193 true negatives (TN), highlighting its capability to correctly identify both classes despite minor misclassifications, with 15 false positives (FP) and 33 false negatives (FN). Support Vector Machine (SVM) follows closely, yielding 198 TP and 182 TN but facing slightly more classification challenges, as seen in its 28 FP and 44 FN. Both Multi-layer Perceptron (MLP) and Decision Tree (DT) classifiers show moderate results, with TP and TN values near those of RF and SVM, yet a somewhat higher number of FP and FN. In contrast, Logistic Regression (LR) and Naive Bayes (NB) encounter more pronounced difficulties, as indicated by higher FP and FN values, reflecting limits in their predictive accuracy; LR, for instance, records 178 TP, 191 TN, with 48 FP and 35 FN, underlining notable misclassifications.

In contrast, Gradient Boosting (GB), XGBoost (XGB), LightGBM (LGBM), CatBoost (CB), Extra Trees (ET), and Bagging (BG) classifiers demonstrate robust overall performance. With higher TP and TN counts and lower FP and FN rates, these classifiers exhibit strong predictive capabilities. For instance, GB records 203 TP and 193 TN with only 23 FP and 33 FN, showcasing its robust predictive capability. Bagging classifier (BG) achieves 214 TP and 191 TN, with only 12 FP and 35 FN, indicating its effectiveness in correctly identifying both positive and negative instances. Similarly, XGB, LGBM, CB, and ET classifiers exhibit strong performance metrics, making them ideal choices for accurately predicting class labels in the 'Undergraduate Admission Test Survey' dataset. This overall analysis of the confusion matrix complements the previous evaluation with evaluation metrics for all these classifiers on the undersampled dataset. By considering both the evaluation metrics and the confusion matrix, a comprehensive understanding of each classifier's predictive capabilities emerges.

5.2.2. ROC Curve Analysis

Figure 5 presents the ROC curves for various classifiers evaluated on the undersampled dataset. Notably, classifiers such as Random Forest, MLP, Decision Tree, XGBoost, LGBM, CatBoost, Extra Trees, and Bagging achieve a perfect AUC score of 1.00, demonstrating exceptional capability in distinguishing between positive and negative instances, thanks to their advanced ensemble techniques that enhance predictive accuracy and generalization. In contrast, Gradient Boosting achieves a strong AUC of 0.98, while KNeighbors and AdaBoost score 0.93, indicating solid but slightly lower performance. Logistic Regression and Gaussian NB exhibit lower AUC scores of 0.89 and 0.85, respectively, highlighting their reduced effectiveness in class distinction within this balanced dataset.

While a perfect AUC of 1.0 is achieved by several classifiers, a comprehensive assessment also requires consideration of other metrics like accuracy. For instance, the MLP Classifier, though achieving an AUC of 1.0, has an accuracy of 85%, which is lower than some other high-AUC models. This gap can be attributed to the dataset's complexity and MLP's sensitivity to parameter tuning and training data size. AUC evaluates the model's ability to prioritize true positives over false positives, while accuracy measures overall prediction correctness, which may be influenced by factors like class imbalance. Conversely, the

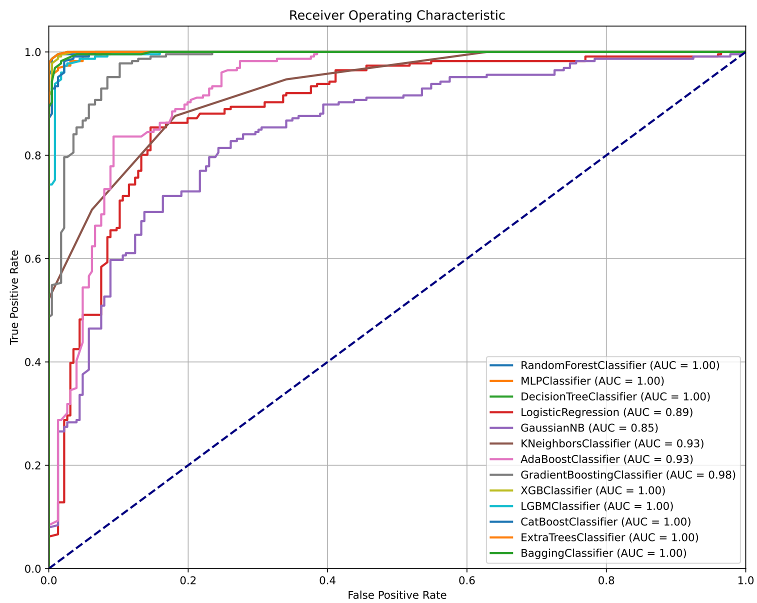


Figure 5: ROC curve comparison for various classifiers on the undersampled dataset

Bagging Classifier stands out by achieving both an AUC of 1.0 and the highest accuracy of 90%, demonstrating its robust discriminatory power and consistent accuracy in correctly classifying instances across both classes.

After evaluating all classifiers using metrics such as precision, recall, F1-score, accuracy, confusion matrices, and ROC curves, the Bagging classifier consistently stands out as the top performer. With a 90% accuracy, minimal misclassification errors, and a perfect AUC of 1.0, Bagging proves to be highly effective in predicting class labels in the 'Undergraduate Admission Test Survey' dataset. It reliably handles complex data, reducing both false positives and false negatives. Notably, Bagging was also the best performer in prior evaluations on the unaltered dataset. However, on the undersampled balanced dataset, its accuracy improved to 90%, benefiting from the reduced data imbalance, which enabled more accurate predictions across both classes.

5.3. Performance Analysis with Oversampling

This section presents the performance analysis of various classifiers on the undergraduate admission dataset after applying oversampling techniques. Addressing the issue of class imbalance through oversampling is crucial to prevent the classifiers from being biased towards the majority class and to ensure that they learn effectively from both classes. In this context, oversampling is implemented by increasing the number of private and public university instances to 500 each, thus achieving a balanced dataset. This analysis offers valuable insights into the performance of each classifier when trained on this balanced subset, emphasizing the impact of oversampling on predictive accuracy and ensuring a comprehensive and fair assessment of each model's capabilities.

Table 7: Classification Report for Various Classifiers on the Oversampled Balanced Dataset

Classifier	Class	Precision (%)	Recall (%)	F1-Score (%)	Support
Random Forest	**Private University**	90	93	92	500
	Public University	93	90	91	500
	Accuracy (%)		**92**		1000
	Macro Avg (%)	92	92	92	1000
	Weighted Avg (%)	92	92	92	1000
Support Vector	**Private University**	88	87	88	500
	Public University	88	88	88	500
	Accuracy (%)		88		1000
	Macro Avg (%)	88	88	88	1000
	Weighted Avg (%)	88	88	88	1000
Multi-layer Perceptron	**Private University**	89	91	90	500
	Public University	91	89	90	500
	Accuracy (%)		90		1000
	Macro Avg (%)	90	90	90	1000
	Weighted Avg (%)	90	90	90	1000
Decision Tree	**Private University**	88	93	91	500
	Public University	93	87	90	500
	Accuracy (%)		90		1000
	Macro Avg (%)	90	90	90	1000
	Weighted Avg (%)	90	90	90	1000
Logistic Regression	**Private University**	82	73	77	500
	Public University	75	84	80	500
	Accuracy (%)		79		1000
	Macro Avg (%)	79	78	78	1000
	Weighted Avg (%)	79	79	78	1000
Naive Bayes	**Private University**	82	54	65	500
	Public University	66	88	75	500
	Accuracy (%)		71		1000
	Macro Avg (%)	74	71	70	1000
	Weighted Avg (%)	74	71	70	1000
K-Nearest Neighbors	**Private University**	87	90	88	500
	Public University	90	86	88	500
	Accuracy (%)		88		1000
	Macro Avg (%)	88	88	88	1000
	Weighted Avg (%)	88	88	88	1000

Tables 7 and 8 provide a detailed evaluation of 14 machine learning classifiers on the oversampled, balanced 'Undergraduate Admission Test Survey' dataset. Random Forest, Gradient Boosting, XGBoost, LightGBM, CatBoost, Extra Trees, and Bagging classifiers consistently achieve high performance, with precision, recall, and F1-scores ranging from 90% to 93% and accuracy rates around 91-92%. This strong performance across metrics highlights their robustness on the balanced dataset, a result of their ensemble approaches that combine predictions from multiple base models, enhancing accuracy and reducing overfitting. These classifiers effectively handle noise and outliers, showcasing reliability in complex data environments. For example, Random Forest achieves 90% precision and 93% recall for private universities, with an overall accuracy of 92%, demonstrating its strong predictive capability and cementing its status as a top performer.

The mid-range performance classifiers—Support Vector Machine (SVM), Multi-layer Perceptron (MLP), Decision Tree, and K-Nearest Neighbors (KNN)—show

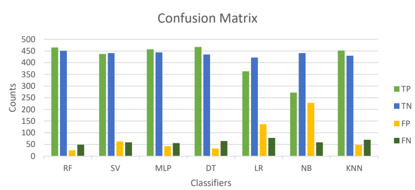
Table 8: Classification Report for Various Classifiers on the Oversampled Balanced Dataset Contd.

Classifier	Class	Precision (%)	Recall (%)	F1-Score (%)	Support
XGBoost	**Private University**	90	92	91	500
	Public University	92	90	91	500
	Accuracy (%)		91		1000
	Macro Avg (%)	91	91	91	1000
	Weighted Avg (%)	91	91	91	1000
AdaBoost	**Private University**	87	81	84	500
	Public University	82	88	85	500
	Accuracy (%)		84		1000
	Macro Avg (%)	84	84	84	1000
	Weighted Avg (%)	84	84	84	1000
Gradient Boosting	**Private University**	93	90	91	500
	Public University	90	93	92	500
	Accuracy (%)		91		1000
	Macro Avg (%)	91	91	91	1000
	Weighted Avg (%)	91	91	91	1000
Light Gradient Boosting Machine	**Private University**	90	92	91	500
	Public University	92	89	91	500
	Accuracy (%)		91		1000
	Macro Avg (%)	91	91	91	1000
	Weighted Avg (%)	91	91	91	1000
CatBoost	**Private University**	91	92	91	500
	Public University	92	91	91	500
	Accuracy (%)		91		1000
	Macro Avg (%)	91	91	91	1000
	Weighted Avg (%)	91	91	91	1000
Extra Trees	**Private University**	90	92	91	500
	Public University	92	90	91	500
	Accuracy (%)		91		1000
	Macro Avg (%)	91	91	91	1000
	Weighted Avg (%)	91	91	91	1000
Bagging	**Private University**	92	90	91	500
	Public University	90	92	91	500
	Accuracy (%)		**91**		1000
	Macro Avg (%)	91	91	91	1000
	Weighted Avg (%)	91	91	91	1000

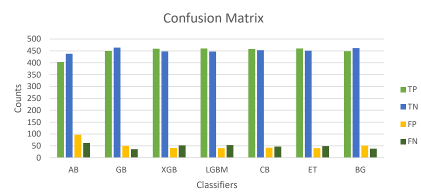
balanced metrics, with precision, recall, and F1-scores ranging from 87% to 90% and accuracy rates around 88% to 90%. SVM excels in distinguishing data points, MLP identifies complex patterns, Decision Trees offer interpretability, and KNN leverages proximity for accurate classification. For instance, MLP achieves 90% accuracy with precision rates of 89% for private universities and 91% for public ones, while the Decision Tree records 88% precision for private universities and 93% for public ones. In contrast, lower-performing classifiers—Logistic Regression, Naive Bayes, and AdaBoost—face distinct challenges. Logistic Regression shows a notable precision and recall gap between private (82% and 73%) and public universities (75% and 84%), resulting in an overall accuracy of 79%. Naive Bayes achieves high recall for public universities (88%) but struggles with precision and recall for private universities (82% and 54%), leading to an overall accuracy of 71%, due to its assumption of feature independence. AdaBoost performs relatively better with 84% accuracy and precision rates of 87% for private and 82% for public universities but can suffer from overfitting by focusing on hard-to-classify instances, explaining its lower performance compared to the top classifiers.

The evaluation reveals a clear stratification in classifier performance on the oversampled balanced dataset, with ensemble methods significantly outperforming other techniques. The robustness of Random Forest, Gradient Boosting, XGBoost, Light Gradient Boosting Machine, CatBoost, Extra Trees, and Bagging is evident from their consistently high precision, recall, F1-scores, and accuracy rates. Mid-tier classifiers such as SVM, MLP, Decision Tree, and KNN show reliable performance but do not match the top tier. Logistic Regression, Naive Bayes, and AdaBoost lag behind, struggling with the dataset's complexity and assumptions that do not hold in practice. Among all classifiers, Random Forest stands out with an impressive accuracy of 92%. This exceptional performance highlights its capability to handle the intricacies of the oversampled balanced dataset effectively.

5.3.1. In-depth Confusion Matrix Analysis



(a) Confusion matrix showing classifier performance



(b) Confusion matrix showing classifier performance

Figure 6: Confusion matrices illustrating illustrating the performance of different Classifier's on the oversampled dataset

Figures 6a and 6b display the confusion matrices for each classifier applied to the oversampled balanced 'Undergraduate Admission Test Survey' dataset. Random Forest (RF) performs strongly, with 465 true positives (TP) and 451 true negatives (TN), along with minimal errors of 25 false positives (FP) and 49 false negatives (FN). Support Vector Machine (SV) shows slightly lower performance, recording 437 TP and 441 TN, but has higher error rates of 63 FP and 59 FN. Multi-layer Perceptron (MLP) achieves 457 TP and 444 TN, yet it has more errors than RF, with 43 FP and 56 FN. Decision Tree (DT) performs well with 467 TP and 435 TN but misses more positive instances, indicated by 33 FP and 65 FN. In contrast, Logistic Regression (LR) struggles with 363 TP and 422 TN, resulting in high error rates of 137 FP and 78 FN. Naive Bayes (NB) shows limited effectiveness, achieving only 272 TP and 441 TN, alongside high FP (228) and FN (59). K-Nearest Neighbors (KNN) records decent performance with 452 TP and 430 TN, but has higher errors, with 48 FP and 70 FN.

Among the ensemble methods, Gradient Boosting (GB), XGBoost (XGB), Light Gradient Boosting Machine (LGBM), CatBoost (CB), Extra Trees (ET), and Bagging (BG) classifiers demonstrate strong performance. Gradient Boosting achieves 450 TP and 464 TN with relatively low errors of 50 FP and 36 FN. XGBoost closely follows with 459 TP and 448 TN, and errors of 41 FP and 52 FN. Light Gradient Boosting Machine shows similar results with 460 TP, 447 TN, 40 FP, and 53 FN. CatBoost and Extra Trees also maintain high accuracy with CatBoost showing 458 TP and 453 TN, and Extra Trees achieving 460 TP and 451 TN. Bagging performs well with 449 TP and 462 TN, and low errors of 51 FP and 38 FN. Overall, Random Forest stands out as the best classifier with a balance of high true positives and true negatives and relatively low false positives and false negatives.

5.3.2. ROC Curve Analysis

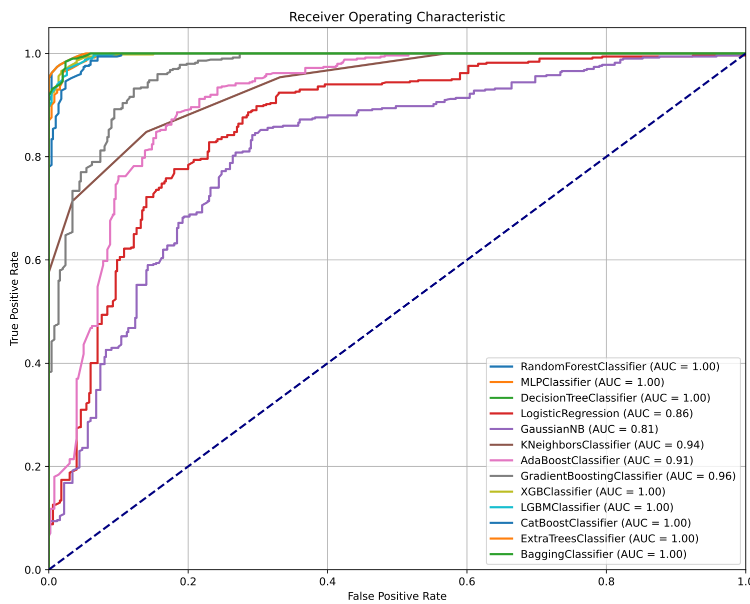


Figure 7: ROC curve comparison for various classifiers on the oversampled dataset

Figure 7 illustrates the ROC curves for various machine learning classifiers applied to the oversampled dataset, highlighting notable differences in their ability to differentiate between positive and negative instances. Random Forest, MLP, Decision Tree, XGB, LGBM, CatBoost, Extra Trees, and Bagging classifiers achieve a perfect AUC score of 1.00, indicating their flawless distinction between classes and reflecting their robustness and effectiveness in capturing data patterns. In contrast, KNeighbors and Gradient Boosting classifiers perform well with AUC scores of 0.94 and 0.96, respectively, demonstrating strong discrimination capability. AdaBoost follows with an AUC of 0.91, while Logistic Regression and Gaussian NB lag behind with scores of 0.86 and 0.81, indicating greater difficulty in accurately classifying instances compared to the top-performing classifiers.

While several classifiers achieve an AUC of 1.0, indicating flawless performance in distinguishing between positive and negative instances, it's essential to also consider accuracy for a comprehensive evaluation. Despite perfect AUC scores, classifiers like the Decision Tree and MLP Classifier have lower accuracy rates of 90%, suggesting a gap between discriminatory power and overall prediction correctness, which can be affected by class imbalance or misclassification of minority classes. In contrast, the Random Forest Classifier excels with the highest AUC and accuracy of 92%, highlighting its strong discriminatory capabilities and reliability in classifying both positive and negative instances, making it particularly effective for this classification task.

After a comprehensive evaluation across various metrics—such as precision, recall, F1-score, accuracy, confusion matrix analysis, and ROC curve assessment—Random Forest emerged as the top-performing classifier for the oversampled balanced 'Undergraduate Admission Test Survey' dataset. Its strong performance in all metrics, marked by high precision, recall, and accuracy rates, underscores its effectiveness in managing the dataset's complexities. The confusion matrix analysis further illustrates Random Forest's capability to accurately predict both positive and negative instances with minimal errors, while a perfect AUC score of 1.00 confirms its exceptional discriminative ability. Overall, Random Forest offers an optimal combination of predictive accuracy, robustness, and discriminative power for classification tasks in oversampled balanced datasets.

5.4. Overall Comparative Analysis

The table 9 offers a comprehensive comparison of classifier accuracy across three distinct scenarios: without sampled data, undersampled balanced data, and oversampled balanced data. These scenarios were examined to discern in which scenario the dataset yields the best performance. Each row corresponds to a different classifier, while each column signifies one of the three scenarios. Notably, this analysis aims to identify the optimal classifier capable of delivering robust performance across all scenarios.

Table 9: Comparison of Classifier Accuracy across 3 Scenarios

Classifier	**Without Sampled (%)**	**Undersampled Balanced (%)**	**Oversampled Balanced (%)**
Bagging	84	90	91
CatBoost	84	89	91
LightGBM	84	87	91
Random Forest	83	89	92
XGBoost	83	89	91
Support Vector	79	84	88
Logistic Regression	79	82	79
Gradient Boosting	83	88	91
Decision Tree	80	85	90
Extra Trees	81	87	91
AdaBoost	84	83	84
Multi-layer Perceptron	76	85	90
K-Nearest Neighbors	74	83	88
Naive Bayes	75	76	71

In the scenario without sampled data, the classifiers Bagging, CatBoost, and LightGBM attain the highest accuracy rate of 84%. In the scenario characterized by undersampling and balance, Bagging leads with an impressive accuracy of 90%, closely trailed by CatBoost and Random Forest, both achieving 89%. The scenario involving oversampling and balance produces the most favorable outcomes, with Random Forest achieving the highest accuracy of 92%, while Bagging, Extra Trees, CatBoost, and LGBM each reach 91%. This indicates that classifiers perform significantly better in the oversampled scenario compared to others. Nevertheless, in the undersampled scenario, Bagging maintains a commendable performance with 90% accuracy, albeit slightly lower than its performance in the oversampled context. Conversely, classifiers encounter difficulties due to data imbalance in the unaltered scenario, resulting in reduced accuracy scores, with Bagging attaining 84%. Both Random Forest and Bagging consistently exhibit strong performance, achieving a perfect AUC score of 1 across all scenarios. Random Forest stands out in the oversampled scenario, while Bagging demonstrates remarkable consistency throughout all scenarios, establishing them as the most effective classifiers for this undergraduate admission test survey dataset.

5.5. Shapley-Based Feature Significance

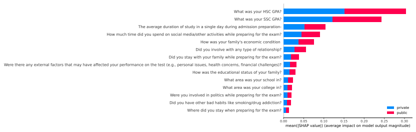
Shapley values, based on cooperative game theory and named after Lloyd Shapley [28], offer a systematic way to assess individual features' contributions to model predictions. By considering all possible feature combinations, they ensure a fair allocation of importance and accurately evaluate their impact on outputs. This measurement enhances the interpretability of complex models and clarifies the decision-making process. SHAP visualizes these values, highlighting significant features in predictions. In undergraduate admission surveys, Shapley values are crucial for identifying factors that influence a student's chances of admission, such as high school GPA, study habits, and family background. The analysis of the top five classifiers focuses on their Shapley values to reinforce the credibility of the findings.

According to the Shapley value definition [29], the Shapley value for a feature i in a model with n features is calculated as follows:

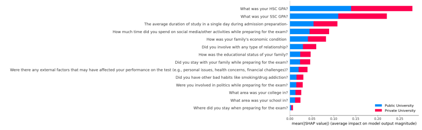
$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!\,(n - |S| - 1)!}{n!}\,[f(S \cup \{i\}) - f(S)] \tag{7}$$

where: ϕ_i is the Shapley value for feature i, N is the set of all features, S is a subset of N that does not include i, $|S|$ denotes the number of features in subset S, $f(S)$ represents the model output when only features in S are included, $f(S \cup \{i\})$ represents the model output when feature i is added to subset S.

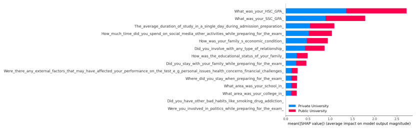
Equation 7 calculates the marginal contribution of each feature i by averaging the change in the model's output across all possible subsets S of features, weighted by the size of each subset. The Shapley value hence captures the fair contribution of feature i by considering all possible interactions with other features. This equation underpins SHAP's ability to accurately represent feature importance, making it a powerful tool for interpreting complex model predictions [30].



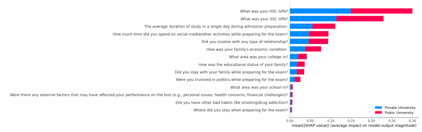
(a) **Random Forest Classifier**



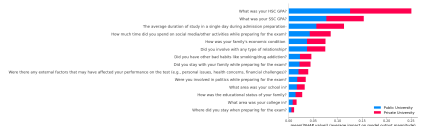
(b) **Bagging Classifier**



(c) **LGBM Classifier**



(d) **Gradient Boosting Classifier**



(e) **Extra Trees Classifier**

Figure 8: Feature Importance by Mean SHAP Value for Various Classifiers

5.5.1. Feature Importance Across Classifiers

In analyzing university admission factors, Figures 8a through 8e highlight a comparative examination of SHAP values across the top classifiers—Random Forest, Bagging, LGBM, Gradient Boosting, and Extra Trees—demonstrating consistent trends in feature importance, particularly underscoring the critical role of academic performance in predicting university admissions outcomes. All classifiers reveal that "HSC GPA" and "SSC GPA" are the most influential features, particularly for public universities, where higher GPA requirements are standard. Study habits also emerge as significant factors across classifiers, with features like "average duration of daily study" and "time spent on social media or other activities during exam preparation" showing a strong impact on predictions. This uniformity indicates that academic commitment and effective time management are essential for achieving admission success in both private and public institutions.

Economic and social factors exhibit moderate yet notable influence, as indicated by "family's economic condition" and "involvement in relationships," both of which affect outcomes to varying degrees across classifiers. LightGBM and Gradient Boosting classifiers particularly demonstrate a marginally greater sensitivity to economic conditions in the context of public university predictions. In contrast, Bagging and Extra Trees classifiers exhibit comparable trends but attribute somewhat less importance to these non-academic factors. This discrepancy suggests that although socioeconomic background and personal relationships play a role, their impact may differ slightly based on the specific model employed. Conversely, certain features such as 'Where did you stay during exam preparation?' and 'Involvement in politics or bad habits like smoking/drug addiction' show low SHAP values across all classifiers, indicating a minor impact on admissions predictions.

5.5.2. Key Determinants of University Admissions in Bangladesh

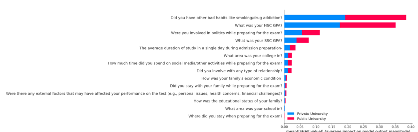
An analysis of the top five classifiers—Random Forest, Bagging, LGBM, Gradient Boosting, and Extra Trees—offers valuable insights into the factors influencing university admissions in Bangladesh, as illustrated by the feature importance shown in Figure 8. Key features consistently identified include academic performance metrics, particularly "HSC GPA" and "SSC GPA," which show the highest SHAP values across all classifiers, highlighting their critical role in admission decisions. The strict GPA standards set by public universities highlight the significance of academic achievement in gaining entry to these prestigious institutions. Conversely, private universities tend to have more flexible GPA requirements, allowing students with lower GPAs to gain admission. Consequently, those who fail to meet the demanding criteria of public universities frequently turn to private institutions for their educational pursuits. High GPAs in HSC and SSC examinations reflect robust academic capabilities, further establishing the connection between GPA and successful admissions, particularly as "HSC GPA" plays a vital role in entrance examinations following HSC.

In Bangladesh, university admissions are influenced by factors beyond GPA scores, including study habits, family economic status, and social activities during exam preparation. Key metrics like "average daily study duration" and "time spent on social media" are crucial, alongside considerations of "family economic condition" and "relationship involvement" in this competitive landscape. Consistent study habits reflect a student's preparedness and dedication to achieving academic success. Those who allocate adequate time for concentrated study generally excel in entrance examinations. In contrast, excessive use of social media can disrupt study schedules and adversely affect exam results. Furthermore, a student's financial situation plays a crucial role in shaping their educational journey, as insufficient resources may limit access to vital study materials and tutoring, thereby influencing performance in public entrance exams. Additionally, romantic relationships during the period of exam preparation can distract from academic obligations, potentially undermining exam results. These interrelated factors illustrate the intricate connection between socioeconomic status, personal relationships, and academic performance, emphasizing the necessity for strong support systems to provide equitable educational opportunities for all students.

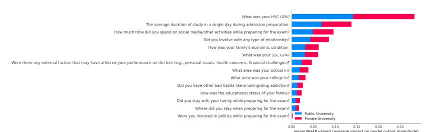
Factors like geographical location ("Where did you stay when preparing for the exam?"), personal habits ("Did you have other bad habits like smoking/drug addiction?"), and political involvement ("Were you involved in politics while preparing for the exam?") show relatively low SHAP values across all classifiers, indicating their limited impact on university admissions in Bangladesh. While these factors may influence individual situations, their overall impact on admission results is limited, underscoring the primacy of academic achievement and socioeconomic considerations. In Bangladesh, cultural norms tend to discourage early political participation, and instances of smoking or drug addiction among the youth are rare, largely due to religious and cultural values that advocate for healthy living. As a result, the emphasis on academic performance in university admissions tends to overshadow the effects of factors such as geographical location, personal behaviors, and political involvement. Although these aspects can enhance a candidate's profile, their significance in the admissions process remains minimal, highlighting the intricate interplay between cultural values, academic criteria, and socioeconomic factors in shaping the landscape of university admissions in Bangladesh.

5.5.3. Discrepancies in Feature Importance Across Classifiers and Their Impact

Figures 9a and 9b present an analysis of feature importance through mean Shapley values for the Naive Bayes and logistic regression classifiers, both of which demonstrate lower accuracy compared to top-tier classifiers. In Figure 9a, examining the Naive Bayes classifier reveals that the feature "Did you have other bad habits like smoking/drug addiction?" ranks as the most influential, despite being the second least important for the high-performing Random Forest classifier. This discrepancy illustrates the variability in feature importance across different models. Additionally, while "What was your HSC GPA?" is the second most critical feature for Naive Bayes, it, along with "Were you involved in politics while preparing for the exam?", ranks low for top classifiers, highlighting Naive Bayes's tendency to undervalue key predictors that drive better results.



(a) **Naive Bayes Classifier**



(b) **Logistic Regression Classifier**

Figure 9: Comparison of Feature Importance by Mean SHAP Value

In Figure 9b, logistic regression also presents notable findings, with "What was your HSC GPA?" emerging as the most crucial factor, consistent with expectations. However, it contrasts sharply with the top-tier classifiers, where "What was your SSC GPA?" consistently ranks as the second most important feature, yet falls to sixth place for logistic regression. This variation indicates a potential reason for the diminished accuracy of logistic regression, highlighting the significant importance of feature selection in the effectiveness of classifiers. The examination reveals that emphasizing pertinent features is crucial for achieving optimal classifier performance, especially in intricate domains such as university admissions. Failing to adequately prioritize key features may result in less favorable results, as evidenced by the variations among classifiers. SHAP serves as a valuable tool for examining these essential characteristics, enhancing the overall comprehension of their impact on classifier performance and facilitating well-informed decision-making.

5.6. Local Interpretability with LIME

LIME, or Local Interpretable Model-agnostic Explanations, is an interpretability tool that offers insights into machine learning model predictions at the local level. Unlike global methods that explain model behavior across an entire dataset, LIME focuses on individual predictions, which is crucial for understanding specific inputs' influence—especially in high-stakes scenarios. In this study, LIME complements the global interpretability of SHAP (SHapley Additive exPlanations) values. While SHAP provides an overview of feature importance, LIME delivers detailed analysis for individual predictions, enhancing the understanding of factors affecting university admissions and validating critical features identified by SHAP. This integrated approach improves transparency and reliability, illustrated by visual representations of the top five classifiers.

LIME approximates the complex model f around a specific instance x by creating a simpler, interpretable model g (often linear) within a local neighborhood of x. The local model g is trained on perturbed samples from x, weighted by their proximity. The objective function for the interpretable model $g$is s defined as follows (Ribeiro et al. [31]):

$$\arg\min_{g \in G} \mathcal{L}(f, g, \pi_x) + \Omega(g) \tag{8}$$

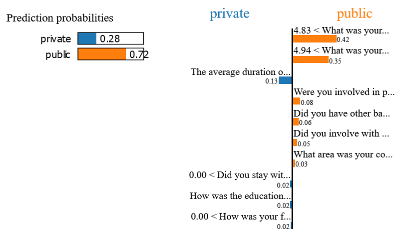
where: $\mathcal{L}(f, g, \pi_x)$ is a loss function assessing g's fidelity in approximating f around x, π_x assigns higher weights to instances closer to x, $\Omega(g)$ is a regularization term that penalizes model complexity, promoting interpretability.

Equation 8 enables LIME to create a local model g that approximates f's predictions near x, allowing for a nuanced understanding of individual feature contributions. This methodology complements the global insights provided by SHAP values.

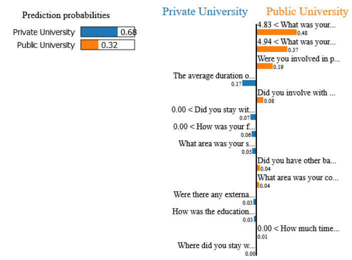
5.6.1. LIME Predictions Across Classifiers

The LIME (Local Interpretable Model-agnostic Explanations) visualizations in Figures 10a to 10e provide a comparative analysis of predictions across various classifiers—Random Forest, Bagging, LightGBM, Gradient Boosting, and Extra Trees—in the context of university admissions outcomes. The Random Forest classifier shows a probability distribution (0.28 for Private University and 0.72 for Public University), indicating a higher level of confidence in its predictions. Its emphasis on key features, such as "average daily study duration" for the private class and HSC/SSC GPAs for the public class, reflects a stronger tendency toward predicting the Public University class. In contrast, the Bagging classifier shows slightly lower certainty (68% for Private University), with "average study duration" as a prominent feature. While Bagging demonstrates a clear decision-making preference, Random Forest outperforms it with greater confidence, leading to more definitive predictions in individual cases.

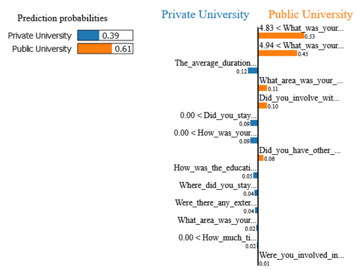
The classifiers LightGBM and Gradient Boosting exhibit a distinct inclination towards the Public University category, with probabilities of 0.61 and 0.67, respectively, suggesting a higher level of confidence compared to the Extra



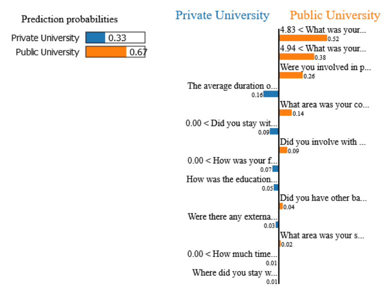
(a) **Random Forest Classifier**



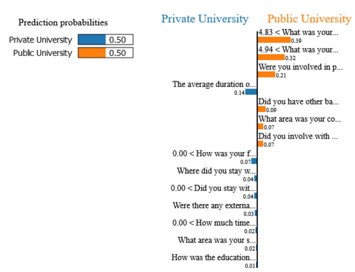
(b) **Bagging Classifier**



(c) **LGBM Classifier**



(d) **Gradient Boosting Classifier**



(e) **Extra Trees Classifier**

Figure 10: Prediction Interpretations with LIME for Various Classifiers

Trees classifier. LightGBM's predictions are predominantly aligned with the public category, placing considerable weight on GPA-related features, whereas Gradient Boosting reveals an even stronger assurance, influenced by similar features in addition to political engagement. Conversely, the Extra Trees classifier stands out by presenting an equal probability of 0.50 for both categories, indicating a lack of clarity in its predictions. This even distribution underscores a degree of uncertainty, which is in stark contrast to the elevated confidence levels demonstrated by Random forest, Bagging, LightGBM, and Gradient Boosting.

The visualizations presented collectively demonstrate that classifiers utilizing ensemble techniques, such as Bagging, Gradient Boosting, and Random Forest, generally exhibit greater predictive confidence. In contrast, Extra Trees show reduced certainty, likely due to the varied and intricate decision boundaries formed by the multiple tree configurations. LIME, which elucidates individual predictions, offers valuable insights into the specific features that the model prioritizes when making a decision for a particular instance, rather than providing a generalized overview across all predictions. The primary factors identified through SHAP values—such as HSC and SSC GPA, average study duration, and family economic conditions—are effectively emphasized and corroborated by LIME, thereby clarifying the influence of each feature on individual predictions. This analysis underscores the importance of integrating both local (LIME) and global (SHAP) interpretability tools to enhance transparency and reliability in critical areas such as university admissions, ultimately increasing trust in the model's predictions. Additionally, this LIME analysis reinforces the earlier assertion that both Random Forest and Bagging consistently exhibit strong performance across all analytical scenarios. In this context, both classifiers stand out, demonstrating the highest confidence in their decision-making processes.

6. Threats to Validity

This study, although strong in its methodological approach, encounters several potential validity threats that require thorough examination. A significant issue is the possibility of selection bias affecting internal validity, as the data was gathered through online surveys that may not accurately reflect the entire demographic of university applicants in Bangladesh. There is a risk that respondents may systematically differ from those who did not participate, resulting in selection bias. Additionally, the integrity of the data is paramount; any inaccuracies in self-reported information or instances of missing data could undermine the reliability of the findings, even with preprocessing efforts in place. Expanding the amount of data collected could improve the validity of the results by offering a more complete representation of the population.

Construct validity plays a crucial role, as the attributes under examination are pertinent and meaningful; however, the study may overlook other critical factors influencing admission outcomes. Furthermore, the validity of the statistical analysis is undermined by the disparity in sample sizes between students from public and private universities, which may diminish the robustness of the findings, even with the application of various sampling techniques. To enhance the interpretability of the model, alternative explainable AI approaches such as Integrated Gradients, DeepLIFT, and Anchors could be utilized in conjunction with SHAP and LIME. Although the methodology is robust, these issues underscore potential areas for enhancement and present opportunities for future research to address these limitations.

7. Conclusion

This study provides an in-depth analysis of the factors influencing undergraduate admission success in Bangladesh's public universities, utilizing advanced machine learning techniques and interpretability tools. By utilizing classifiers like SVM, Random Forest, and GBM, along with thorough validation techniques such as Grid Search and 10-fold Cross-Validation, the effectiveness and dependability of predictive models are guaranteed. The inclusion of SHAP and LIME enhances the clarity of results, providing a thorough analysis of the roles played by different features in admission results. This research identifies the optimal classifiers and highlights the most influential features in the admission process, offering valuable insights for future strategies in higher education.

The findings highlight the crucial influence of academic achievement, socioeconomic status, and access to preparatory resources in shaping students' prospects of being accepted into public universities. Identifying the main factors can guide the creation of focused interventions aimed at assisting disadvantaged students, thereby fostering a sustainable and more inclusive educational system. Furthermore, the utilization of machine learning and explainable AI techniques in this context showcases the potential of these approaches in tackling intricate challenges within educational environments. The transparency provided by SHAP and LIME not only enhances comprehension of model predictions but also clarifies the reasoning behind admission choices, thereby cultivating trust and confidence in the selection process.

Future investigations should build upon this study by utilizing longitudinal data to evaluate the enduring impacts of admission policies and interventions. Analyzing the application of these strategies in various educational settings within Bangladesh and other regions could further enrich the results. To mitigate potential validity issues, subsequent research should focus on expanding the dataset to encompass a more varied sample of university applicants, thus reducing selection bias and enhancing data precision. The inclusion of additional pertinent variables and the adoption of a mixed-methods framework that combines qualitative research will facilitate a more holistic understanding of admission challenges. Moreover, examining a wider array of explainable AI methodologies can substantiate the machine learning models and provide deeper insights into the determinants affecting admission outcomes. These initiatives will refine the study's methodology, leading to more dependable and comprehensive findings, ultimately fostering sustainable data-driven and equitable decision-making in higher education.

Data Availability

The dataset used in this study is available at the following link: https://www.kaggle.com/datasets/miinhaz/undergrad.

Financial Disclosure

This study has no financial interests or relationships to disclose.

Declaration of competing interest

The authors declare no conflict of interest.

References

[1] The Business Standard, Hsc results 2023: Girls continue to lead the race, https://www.tbsnews.net/bangladesh/education/hsc-results-2023-girls-continue-lead-race-746514, accessed: May, 2024 (2023).

[2] New Age, 30pc seats in universities, colleges in bangladesh to remain vacant, https://www.newagebd.net/article/194025/30pc-seats-in-universities-colleges-to-remain-vacant, accessed: May, 2024 (2024).

[3] Skilling for employment: Tvet in bangladesh, https://theconfluence.blog/skilling-for-employment-tvet-in-bangladesh/, accessed: May, 2024.

[4] C. Romero, S. Ventura, Educational data mining: A survey from 1995 to 2005, Expert systems with applications 33 (1) (2007) 135–146.

[5] M. I. Jordan, T. M. Mitchell, Machine learning: Trends, perspectives, and prospects, Science 349 (6245) (2015) 255–260.

[6] R. K. Subramaniam, S. D. Samuel, M. Seera, N. Alam, Utilising machine learning for corporate social responsibility (csr) and environmental, social, and governance (esg) evaluation: Transitioning from committees to climate, Sustainable Futures (2024) 100329.

[7] K. C. A. Khanzode, R. D. Sarode, Advantages and disadvantages of artificial intelligence and machine learning: A literature review, International Journal of Library & Information Science (IJLIS) 9 (1) (2020) 3.

[8] M. Yağcı, Educational data mining: prediction of students' academic performance using machine learning algorithms, Smart Learning Environments 9 (1) (2022) 11.

[9] K. S. Rawat, I. Malhan, A hybrid classification method based on machine learning classifiers to predict performance in educational data mining, in: Proceedings of 2nd International Conference on Communication, Computing and Networking: ICCCN 2018, NITTTR Chandigarh, India, Springer, 2019, pp. 677–684.

[10] A. M. P. Chacón, I. S. Ramírez, F. P. G. Márquez, K-nearest neighbour and k-fold cross-validation used in wind turbines for false alarm detection, Sustainable Futures 6 (2023) 100132.

[11] E. Hirata, K. X. Li, D. Watanabe, Exploring decarbonization priorities for sustainable shipping: A natural language processing-based experiment, Sustainable Futures (2024) 100358.

[12] F. Habibi, M. A. Zabardast, Digitalization, education and economic growth: A comparative analysis of middle east and oecd countries, Technology in Society 63 (2020) 101370.

[13] M. Protikuzzaman, M. K. Baowaly, M. K. Devnath, B. C. Singh, Predicting undergraduate admission: A case study in bangabandhu sheikh mujibur rahman science and technology university, bangladesh, International Journal of Advanced Computer Science and Applications 11 (12) (2020).

[14] M. A. A. Walid, S. M. Ahmed, M. Zeyad, S. S. Galib, M. Nesa, Analysis of machine learning strategies for prediction of passing undergraduate admission test, International Journal of Information Management Data Insights 2 (2) (2022) 100111.

[15] M. S. Zulfiker, N. Kabir, A. A. Biswas, P. Chakraborty, M. M. Rahman, Predicting students' performance of the private universities of bangladesh using machine learning approaches, International Journal of Advanced Computer Science and Applications 11 (3) (2020) 672–679.

[16] A. Sattar, R. Bairagi, S. Miah, R. Zannat, Prediction of finest department assortment on undergraduate admission through machine learning approach.

[17] S. A. Ahmed, S. I. Khan, A machine learning approach to predict the engineering students at risk of dropout and factors behind: Bangladesh perspective, in: 2019 10th international conference on computing, communication and networking technologies (ICCCNT), IEEE, 2019, pp. 1–6.

[18] S. H. Hemal, M. A. R. Khan, I. Ahammad, M. Rahman, M. A. S. Khan, S. Ejaz, Predicting the impact of internet usage on students' academic performance using machine learning techniques in bangladesh perspective, Social Network Analysis and Mining 14 (1) (2024) 1–41.

[19] M. R. Islam, A. M. Nitub, M. A. Al Mamunc, M. A. Marjand, M. P. Uddine, M. I. Afjalf, An educational data mining system for predicting students' programming performance (2023).

[20] M. J. Mia, A. Biswas, A. Sattar, M. T. Habib, Registration status prediction of students using machine learning in the context of private university of bangladesh, International Journal of Innovative Technology and Exploring Engineering (IJITEE) 9 (01) (2019) 2594–2600.

[21] P. Dabhade, R. Agarwal, K. Alameen, A. Fathima, R. Sridharan, G. Gopakumar, Educational data mining for predicting students' academic performance using machine learning algorithms, Materials Today: Proceedings 47 (2021) 5260–5267.

[22] Y. Nieto, V. Gacía-Díaz, C. Montenegro, C. C. González, R. G. Crespo, Usage of machine learning for strategic decision making at higher educational institutions, Ieee Access 7 (2019) 75007–75017.

[23] M. Hussain, W. Zhu, W. Zhang, S. M. R. Abidi, S. Ali, Using machine learning to predict student difficulties from learning session data, Artificial Intelligence Review 52 (2019) 381–407.

[24] R. Rois, M. Ray, A. Rahman, S. K. Roy, Prevalence and predicting factors of perceived stress among bangladeshi university students using machine learning algorithms, Journal of Health, Population and Nutrition 40 (2021) 1–12.

[25] N. Abuzinadah, M. Umer, A. Ishaq, A. Al Hejaili, S. Alsubai, A. Eshmawi, A. Mohamed, I. Ashraf, Role of convolutional features and machine learning for predicting student academic performance from moodle data, Plos one 18 (11) (2023) e0293061.

[26] S. Trivedi, Improving students' retention using machine learning: Impacts and implications, ScienceOpen Preprints (2022).

[27] A. AlGhamdi, A. Barsheed, H. AlMshjary, H. AlGhamdi, A machine learning approach for graduate admission prediction, in: Proceedings of the 2020 2nd International Conference on Image, Video and Signal Processing, 2020, pp. 155–158.

[28] A. E. Roth, Introduction to the shapley value, The Shapley value 1 (1988).

[29] L. S. Shapley, A value for n-person games, Contribution to the Theory of Games 2 (1953).

[30] S. Lundberg, A unified approach to interpreting model predictions, arXiv preprint arXiv:1705.07874 (2017).

[31] M. T. Ribeiro, S. Singh, C. Guestrin, " why should i trust you?" explaining the predictions of any classifier, in: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, 2016, pp. 1135–1144.