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

Md. Minhazul Islam (ETE), S. M. Mahedy Hasan (CSE), Md. Rakib Hossain (ETE)

[a] *Rajshahi University of Engineering & Technology, , Rajshahi, 6204, Bangladesh*

Abstract

Exploring the complex landscape of undergraduate admissions in Bangladesh, specifically in public universities, where intense competition arises due to limited seats and many applicants, sophisticated data mining and machine learning techniques are utilized to analyze the various factors influencing admission outcomes. Careful examination of a comprehensive dataset sourced from online surveys amalgamates vital factors including academic achievements, socio-economic backgrounds, and pertinent criteria to derive valuable insights. Through the use of different machine learning classifiers like Bagging, Random Forest, Gradient Boosting and Extra Trees, and by fine-tuning hyperparameters using Grid Search and validating through 10-fold Cross-Validation, accurate predictions of admission success are made. To enhance interpretability, SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) are employed, providing insights into the contributions of individual features and offering a clear understanding of the factors driving admission decisions. The analysis highlights the crucial significance of factors such as previous academic achievements, socio-economic status, and access to preparatory resources in determining students' chances of securing coveted spots in public universities. By comprehending the key elements that contribute to successful admissions, targeted interventions can be developed to provide support for students, especially those from underrepresented backgrounds. Moreover, this research emphasizes the transformative potential of combining machine learning with explainable AI methodologies, thereby enhancing decision-making processes in the field of education.

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 classifiers, researchers can extract valuable insights from the collected data, uncovering patterns, trends, and correlations that may not be immediately evident through manual analysis [5]. Furthermore, ML techniques enable the automation of repetitive tasks, reducing the potential for human error and enhancing the accuracy of the analysis [6]. By combining data mining with ML, educators can acquire a deeper comprehension of student preferences and behaviors, enabling more personalized and effective educational interventions [7, 8]. 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. Moreover, the integration of technology in education is not confined solely to the realm of academia but extends across various sectors. In education, as in other fields, the use of advanced technologies, such as the integration of deep learning, is helping to create more efficient and effective solutions, driving progress and innovation on multiple fronts [9]. Additionally, these technologies bring about a comprehensive assessment of various factors, providing a clearer understanding of their overall impact and effectiveness [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 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 Works

Machine learning techniques, in conjunction with educational data mining, have become essential in the field of education. These techniques facilitate a comprehensive analysis of student preferences, aspirations, and academic performance. By utilizing large datasets, educators and researchers are now able to unravel the intricate dynamics that influence students' decisions regarding universities, departments, and academic pursuits. Moreover, these techniques enable a thorough understanding of the reasons behind students' inability to achieve their desired academic outcomes. Digitalization through these advanced technologies in education supports long-term economic growth by enhancing educational outcomes and aligning academic pathways with evolving industry needs [12]. The sophisticated algorithms employed in machine learning and educational data mining empower institutions to identify the key factors that influence student choices, thereby allowing for the development of more personalized recruitment strategies and efficient resource allocation. Additionally, these approaches provide nuanced insights into academic performance, uncovering patterns and correlations that are often overlooked by traditional analyses. They also shed light on students' overall attitudes and behaviors towards educational matters, revealing valuable patterns and correlations that are frequently missed by conventional methods. Armed with these valuable insights, educators can identify the underlying factors that contribute to students' success or failure and implement targeted interventions to support their progress. Predictive analytics, driven by machine learning and educational data mining, help institutions foresee potential hurdles to student success, facilitating proactive interventions to prevent academic issues from worsening. Additionally, education departments can use these tools to customize educational experiences for each student, leveraging their individual track records and preferences to enhance learning outcomes.

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] examined factors leading to the failure of undergraduate admission seekers using machine learning techniques. They prepared a dataset with ten attributes and 343 observations to simplify model complexity. Six ML techniques were used, with the combination of edited nearest neighbor (ENN) and borderline SVM-based SMOTE showing the best performance, followed by borderline SVM-based SMOTE and Adaboost. The study aimed to aid underprivileged and middle-income families in improving their children's admission prospects to public universities in Bangladesh. By leveraging data mining and ML, it provided insights into predicting exam outcomes, potentially informing the development of mobile apps for assessing admission prospects. Models combining ENN, borderline SVM-based SMOTE, and Adaboost held promise for effectively predicting admission test outcomes, with implications for educational practices and student mental health support.

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] conducted a comprehensive study to address the increasing dropout rates in Bangladeshi universities, particularly within engineering disciplines. Leveraging machine learning techniques, including neural networks and educational data mining, the research aimed to predict at-risk engineering students and uncover the factors driving dropout. Through analysis of personal, academic, and institutional data from 480 students across various universities, significant predictors of dropout, such as living conditions, involvement in student politics, distance from the university, academic performance, and institutional support programs, were identified. Their predictive model achieved an impressive accuracy of 91.5%, surpassing models based solely on personal or academic data. This multifaceted approach provides valuable insights into dropout dynamics and offers strategies for prevention, highlighting the importance of addressing both academic and personal challenges to enhance retention rates in Bangladeshi engineering education.

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. Conversely, Random Forest yielded the lowest accuracy (79.65%). 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] aimed to predict students' academic performance in a technical institution in India using educational data mining techniques. A dataset was collected through a questionnaire-based survey and academic records. Data preprocessing and factor analysis were conducted to clean the data and reduce dimensionality. Machine learning algorithms, including multiple linear regression (MLR) and support vector regression (SVR), were compared using Python. The SVR algorithm with a linear kernel yielded the best prediction accuracy (83.44%), followed closely by MLR. SVR with polynomial and radial basis function kernels showed lower accuracy. The results suggest a linear relationship between the input variables and academic performance. Overall, the study demonstrates the potential of data mining techniques in predicting students' academic outcomes and underscores the importance of past performance in forecasting future performance.

Nieto et al. [22] explored the usage of machine learning algorithms to support strategic decision-making at Higher Educational Institutions (HEIs), focusing on predicting graduation rates of undergraduate engineering students in South America. Three supervised classification algorithms were deployed and evaluated using real data from a public university in Colombia. The analysis, conducted on a dataset of 6100 engineering students over a ten-year period, revealed that Random Forest outperformed the other algorithms, achieving the highest overall accuracy of 84.11%. While Logistic Regression demonstrated a slightly higher area under the curve (AUC), Random Forest proved to be more effective in predicting student graduates, with a recall rate of 91.93%. The findings underscored the potential of machine learning in facilitating informed decision-making processes in HEIs, aiding in resource planning, curriculum design, and other related factors.

The study [23] aimed to predict students' difficulties in subsequent sessions of a digital design course using machine learning algorithms applied to data from a technology-enhanced learning (TEL) system called DEEDS. Various algorithms including artificial neural networks (ANNs), support vector machines (SVMs), logistic regression, Naïve Bayes classifiers, and decision trees were employed and evaluated. Nine significant variables were identified through statistical analysis, indicating their correlation with session grades. ANNs and SVMs demonstrated superior accuracy in predicting student difficulty, achieving up to 80% accuracy using selected features. Integration of these models into TEL systems could assist in early identification of struggling students and enhance 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.

Trivedi et al. [25] 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. [26] addressed the challenge faced by graduate students in identifying suitable universities for postgraduate studies by proposing a machine learning approach to predict the likelihood of admission based on student profiles. Three regression models, including Linear Regression, Decision Tree, and Logistic Regression, were evaluated using data from Kaggle. The Logistic Regression model emerged as the most accurate, with the lowest Root Mean Square Error (RMSE) of 0.072. The study highlighted the potential of machine learning in aiding students' university selection process, providing a tool for informed decision-making and optimizing admission chances. Future work could involve implementing the proposed algorithm as a software tool to assist students in identifying universities that align with their profiles effectively.

The combination of research efforts highlights the significant impact of machine learning and educational data mining on reshaping educational paradigms in Bangladesh. Through the use of various methodologies and complex algorithms, these studies aim to predict and analyze different aspects of students' academic journeys, such as performance predictions, admission probabilities, dropout tendencies, and stress levels. Simultaneously, the research investigates the experiences of students who turned to private universities in Bangladesh after failing to secure admission to public institutions. In a country where public universities are the norm, this investigation aims to comprehend the factors that lead to their exclusion from public universities and subsequent enrollment in private ones. By identifying the challenges these students encountered in meeting the criteria of public universities, the objective is to shed light on the factors that influence their educational paths and choices.

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. 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.

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

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

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 underwent comprehensive scrutiny and were validated as authentic, representing a diverse range of genuine perspectives and experiences related to the admission process across various educational backgrounds. These meticulously curated data have great potential for utilization in machine learning algorithms, offering deeper insights into the complexities of student admissions.

3.3. Description of Attributes

The research utilized a comprehensive dataset collected through the Google Form survey, which provided a rich source of information for analysis. This dataset comprised 15 columns and 596 rows, each row representing an individual response from students enrolled in both public and private universities. The columns in the dataset corresponded to specific questions posed in the survey, each designed to capture a range of attributes related to the admission process and student demographics. The 15 columns included various types of data, such as students' personal information, their preferences, and perceptions regarding the admission process. By analyzing this dataset, the research aimed to gain a deeper understanding of the factors influencing student decisions and experiences in the context of higher education institutions.

SSC GPA & HSC GPA: These attributes capture the academic performance of respondents during their Secondary School Certificate (SSC) and Higher Secondary Certificate (HSC) exams, respectively. The SSC GPA ranged from 2.99 to 5.00, while the HSC GPA ranged from 3.34 to 5.00 among the 595 students included in the dataset. The SSC and HSC GPAs reflect students' dedication and effort in their academic pursuits. A higher GPA demonstrates consistent hard work and effective study habits, influencing admission decisions for higher education. Conversely, a lower GPA can suggest areas of potential weakness, possibly reflecting a lack of effort or challenges in certain subjects. This could limit opportunities in the future, as GPA often plays a significant role in academic and career advancement. Thus, these GPAs play a pivotal role in shaping students' educational paths and future opportunities.

Family's Economic Condition: Reflecting the socioeconomic background of respondents, this attribute sheds light on the financial resources available to support their education. Understanding the economic condition of students' families offers insights into disparities in access to educational opportunities and the impact of financial constraints on academic success. Students from wealthier families often benefit from private tutoring, advanced technological tools, and extracurricular activities, which can enhance their academic performance. Conversely, students from less affluent backgrounds might face challenges such as the need to work part-time jobs, limited access to study materials, and fewer opportunities for enrichment activities.

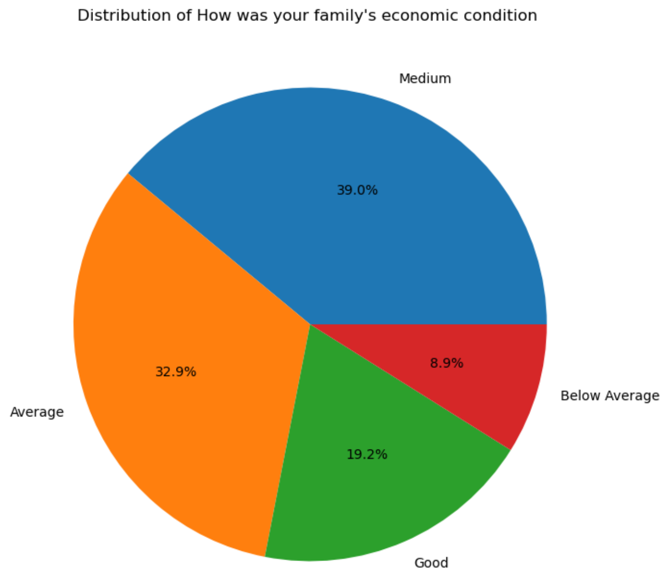


Figure 1: Distribution of Responses on Family Economic Condition.

In Fig. 1, it can be seen that responses are categorized into options like "Good," "Medium," "Average," or "Below Average. The chart reveals that the largest proportion of students, accounting for 39.0%, hail from families with a medium economic condition, followed by 32.9% from families with an average economic condition. Together, these two categories constitute the largest portion.

Where Stayed During Exam Preparation: This attribute reveals the living environment of respondents during their exam preparation, distinguishing between urban (Town) and rural (Village) settings. The differential access to resources, such as educational facilities, internet connectivity, and support networks, between urban and rural areas can significantly influence students' preparation strategies and academic outcomes. Students in urban areas often benefit from better access to libraries, tutoring services, and technology, which can enhance their study efficiency and overall academic performance. In contrast, those in rural settings might face challenges such as limited access to educational materials, fewer extracurricular support options, and potential disruptions due to infrastructural issues.

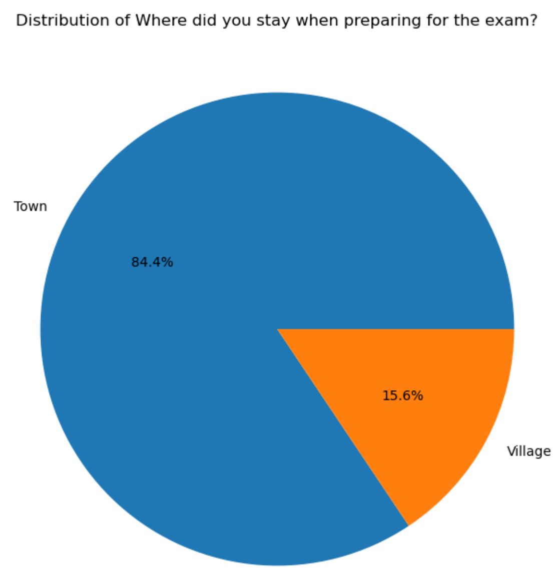


Figure 2: Distribution of Responses on Where Stayed During Exam Preparation.

This question asks respondents to select where they stayed during their exam preparation, either in a "Village" or "Town." In Figure 2, the pie chart illustrates that 84.4% of students opted to stay in town during their exam preparation, while the remaining percentage chose to stay in the village.

Educational Status of Family: Capturing the educational attainment of respondents' families provides crucial insights into the educational background and values present within their households. This attribute highlights the diversity in family educational status, which can significantly influence students' academic aspirations, expectations, and the support networks available to them. Families with higher educational attainment may foster a more supportive environment, encouraging higher academic achievement and providing resources that facilitate educational success. Conversely, students from families with lower educational attainment might face different challenges and rely more on external support systems. Understanding these variations is essential, as they shape students' approaches to education, their motivation, and their ability to navigate academic demands.

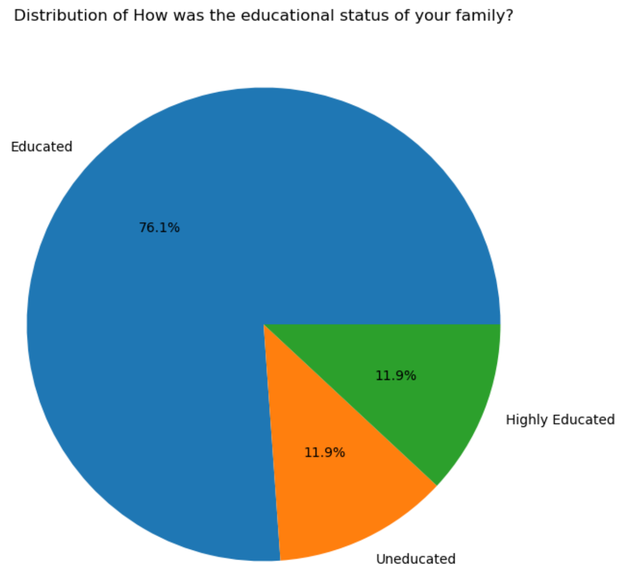


Figure 3: Distribution of Responses on Educational Status of Family.

Here, respondents are asked to choose the educational status of family from options like "Educated," "Highly Educated" or "Uneducated." In Figure 3, the pie chart illustrates that 76.1% of students belong to educated families, while 11.9% belong to both uneducated and highly educated families.

Involvement in Politics:

Examining respondents' engagement in political activities during exam preparation provides valuable insight into the broader socio-political context impacting students' lives. Political involvement can have significant implications for students, potentially influencing their time management, stress levels, and overall well-being, all of which can, in turn, affect their academic performance. For instance, active participation in political activities may require students to allocate time and energy away from their studies, leading to potential disruptions in their exam preparation routines. Additionally, involvement in political movements or events can contribute to heightened stress and emotional strain. In contrast, students who are not engaged in political activities may avoid these particular stressors and time management challenges, allowing them to focus more directly on their academic responsibilities. By understanding the extent and nature of this engagement, researchers can better contextualize the various pressures and challenges students face.

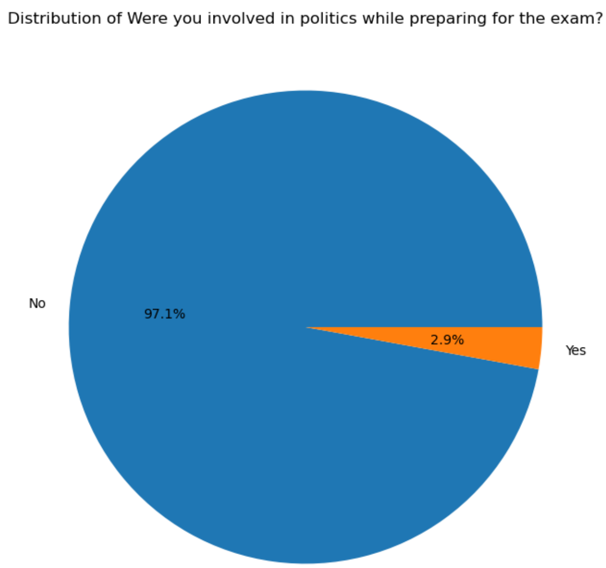


Figure 4: Distribution of Responses on Involvement in Politics.

As depicted in Figure 4, the pie chart shows that a substantial majority, 97.1%, of students were not involved in politics during their exam preparation. This suggests that most students prioritize their academic responsibilities over political engagement during this critical period. Conversely, a small fraction, 2.9%, were involved in political activities.

Time Spent on Social Media/Other Activities: This attribute quantifies respondents' engagement in non-academic activities, such as social media usage, during exam preparation. Understanding how students allocate their time between academic and leisure pursuits can provide valuable insights into potential distractions or coping mechanisms that may influence their academic performance. High levels of engagement in non-academic activities might indicate distractions that could detract from study time and focus, thereby negatively impacting academic outcomes. Conversely, these activities can also serve as important coping mechanisms, providing necessary breaks that help manage stress and maintain mental well-being during intensive study periods.

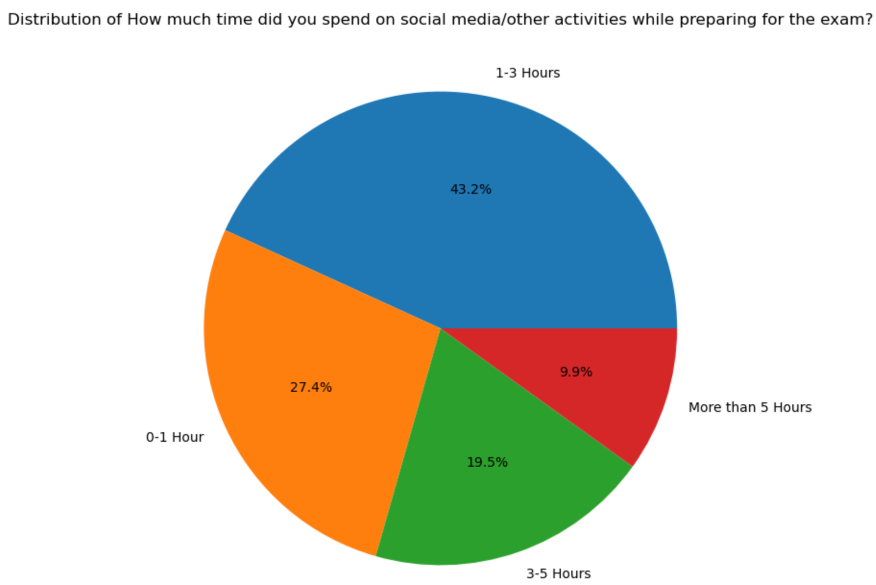


Figure 5: Distribution of Responses on Time Spent on Social Media/Other Activities.

This field asks respondents to indicate the amount of time spent on social media or other activities during exam preparation, with options ranging from "0-1 Hour" to "More than 5 Hours." Figure 5, the pie chart demonstrates that 43.2% of students spent 1-3 hours on social media during exam preparation.

Stayed with Family During Exam Preparation: This attribute indicates whether respondents resided with their families during exam preparation. Family support and living arrangements can significantly influence students' study environment, emotional well-being, and access to academic resources, all of which impact their academic performance. Living with family can provide a stable and supportive environment, where students benefit from emotional encouragement, routine meals, and potentially fewer financial burdens. This supportive atmosphere can enhance focus and reduce stress, contributing positively to academic outcomes. Conversely, those who do not live with family might face challenges such as increased responsibilities, loneliness, and limited access to familial support, which can affect their ability to concentrate and perform well academically.

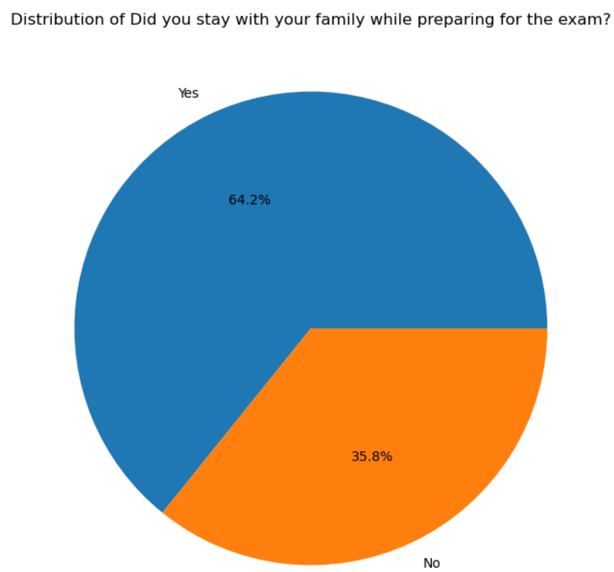


Figure 6: Distribution of Responses on Stayed with Family During Exam Preparation.

Figure 6 shows that 64.2% of students stayed with their family during exam preparation, likely benefiting from their support and stability. In contrast, 35.8% of students did not stay with their family, opting for other living arrangements.

Average Duration of Study in a Single Day: By capturing the amount of time respondents devoted to studying daily during exam preparation, this attribute offers valuable insights into their study habits and time management strategies. Variations in study duration can reveal differences in academic dedication, self-discipline, and the level of perceived academic pressure experienced by students. For instance, students who allocate more time to studying may demonstrate a higher level of commitment and discipline, potentially reflecting greater concern about academic performance or a desire for thorough preparation. Conversely, students who study less may face challenges in balancing their academic responsibilities or may not perceive the same level of pressure, influencing their overall academic experience.

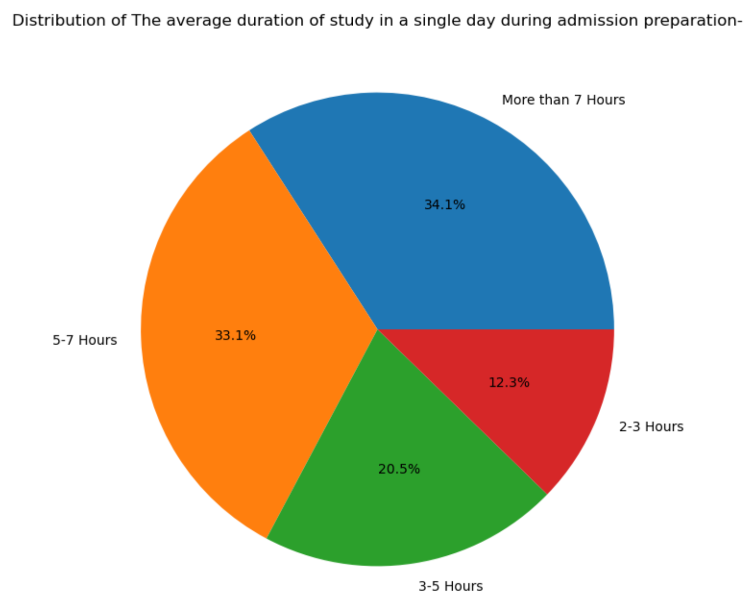
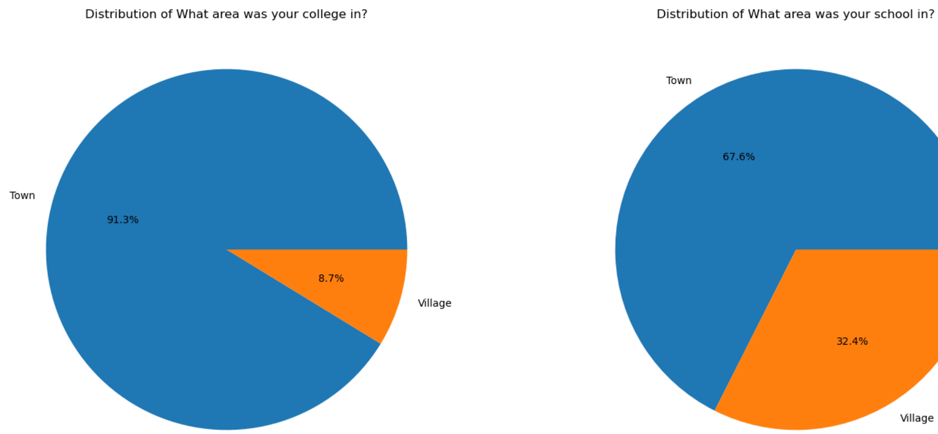


Figure 7: Distribution of Responses on Average Duration of Study in a Single Day.

This question asks respondents to select the average duration of study in a single day during their exam preparation, with options like "2-3 Hours," "3-5 Hours," "5-7 Hours," or "More than 7 Hours." Figure 7 indicates that 34.1% of students studied for more than 7 hours, while 33.1% studied for 5-7 hours.

College Area & School Area: These attributes identify the geographic location of respondents' colleges and schools, providing valuable insights into the educational infrastructure and opportunities available within their respective communities. The geographic setting, whether urban or rural, can greatly influence the quality and accessibility of educational resources, such as advanced facilities, experienced faculty, and extracurricular programs. Urban areas often offer more robust educational infrastructure and a wider range of academic and extracurricular opportunities, which can enhance students' learning experiences and academic advancement. In contrast, rural areas might face challenges such as limited access to modern facilities, fewer specialized programs, and less frequent exposure to diverse learning experiences. These differences in educational access and quality can significantly shape students' academic experiences, opportunities for growth, and overall educational outcomes.



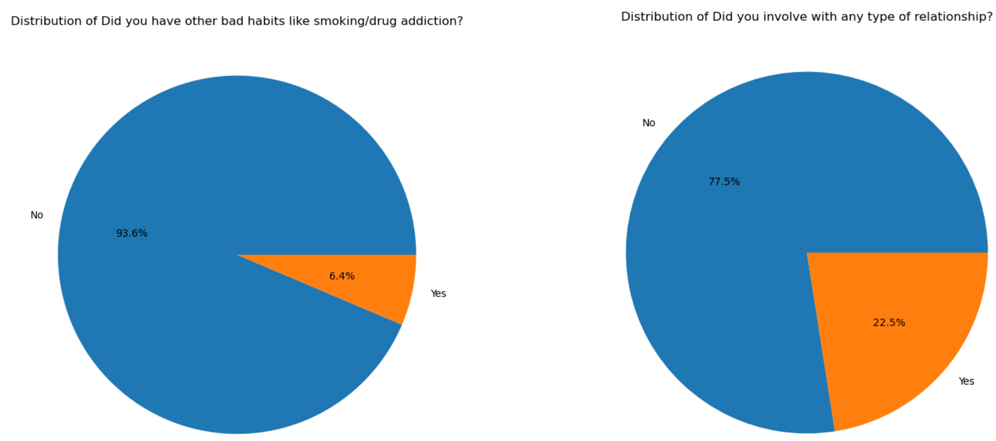
(a) College.

(b) School.

Figure 8: Distribution of Responses on College Area & School Area.

Respondents are asked about the location of their college and school, with options "Village" or "Town." In Figure 8, it is evident that 91.3% of students studied at colleges located in town, while 67.6% of students studied at schools situated in town. A noticeable trend emerges indicating that students tend to relocate from villages to towns for college education during their Higher Secondary Certificate (HSC) studies.

Bad Habits & Involvement in Relationships: These attributes examine respondents' engagement in behaviors such as smoking, drug addiction, and personal relationships during exam preparation. By understanding students' lifestyle choices and personal relationships, we can gain insights into potential distractions, coping mechanisms, or sources of stress that may impact their academic performance. Engaging in behaviors such as smoking or drug use can negatively affect cognitive function, health, and overall well-being, potentially impairing academic focus and performance. Similarly, complex personal relationships might introduce additional stress or emotional challenges that can disrupt study routines and concentration. Addressing these factors can help in developing targeted support strategies to mitigate their impact on students' academic success and overall well-being.



(a) Smoking/Drug Addiction. (b) Relationship.

Figure 9: Distribution of Responses on Bad Habits and Involvement in Relationships.

From Figure 9, it's apparent that 6.4% of students reported having bad habits such as smoking or drug addiction, while 22.5% were involved in a relationship during exam preparation.

External Factors Affecting Performance: This attribute assesses respondents' perceptions of external factors, such as personal issues, health concerns, or financial challenges, that may have influenced their exam performance. By understanding how these external stressors impact students, we gain valuable insights into the broader context affecting their academic experiences. Personal issues, such as family problems or relationship difficulties, can create emotional strain that diverts attention and energy away from studying. Health concerns, including physical or mental health issues, can directly impair students' ability to focus, study effectively, or perform well on exams. Financial challenges may lead to increased stress and preoccupation with economic stability, further detracting from academic efforts. Recognizing these external factors allows for the development of targeted support strategies that address specific challenges faced by students.

In Figure 10, it can be observed that approximately half of the respondents specifically 50.9% reported the presence of external factors that may have impacted their performance during admission preparation, while the other half did not.

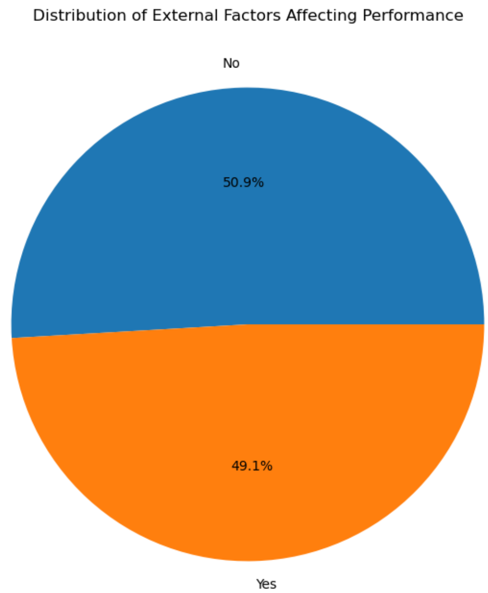


Figure 10: Distribution of Responses on External Factors Affecting Performance.

Name of Current Institution: This attribute functions as the dependent variable, shaped by a range of independent variables previously examined, such as academic performance, socioeconomic background, study habits, and other relevant factors within the dataset. By collecting the names of students' institutions, which are then categorized as either public or private, we can analyze how these institutional types influence or correlate with various aspects of student outcomes. The classification of institutions allows for a more detailed exploration of how different educational environments affect students' academic experiences and achievements. Utilizing machine learning techniques for this analysis enables us to uncover complex patterns and relationships between institutional types and the various independent variables. This approach not only enhances our understanding of how public and private institutions impact academic performance but also helps in identifying key factors that contribute to educational success and challenges within different types of institutions.

Figure 11 illustrates that out of the total respondents, 369 students are from public universities while 226 are from private universities, resulting in a ratio of 62% to 38%.

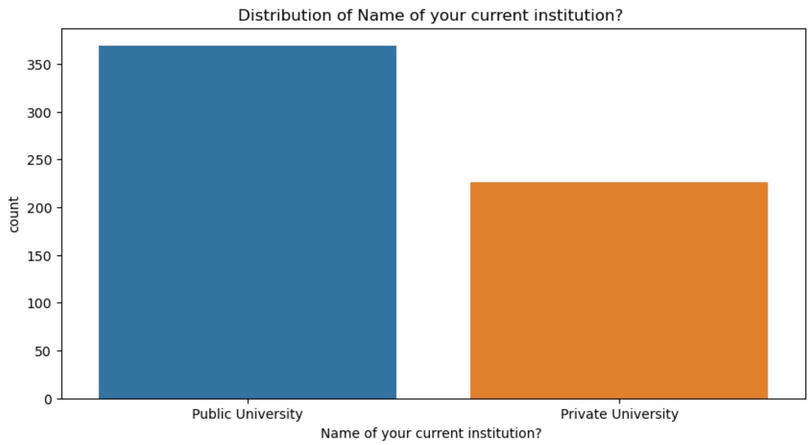


Figure 11: Distribution of Responses on Name of Current Institution.

The dataset compiled from the Undergraduate Admission Test Survey serves as a valuable resource for understanding the multifaceted factors influencing student admissions to public universities in Bangladesh. Through meticulous data collection and analysis, this dataset provides insights into the complex interplay of academic performance, socioeconomic background, and various other factors shaping students' admission outcomes. Leveraging machine learning techniques and visualization methods enables a deeper exploration of feature importance and patterns contributing to admission success or failure. This dataset contributes to the academic discourse on university admissions and offers practical implications for policymakers, educators, and stakeholders aiming to enhance the inclusivity and fairness of 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 12 illustrates the workflow of the proposed methodology.

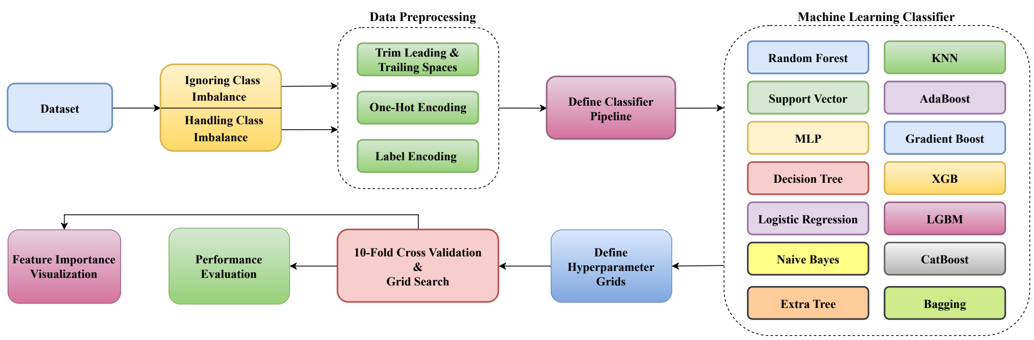


Figure 12: Workflow of the Proposed Methodology

4.1. Class Imbalance Handling

As previously mentioned, the undergraduate admission test survey dataset comprises a total of 595 samples. Within this dataset, the target column indicating the institution type reveals 226 instances for private universities and 369 for public universities. This distribution highlights a class imbalance issue, with a substantial difference in the number of samples between private and public universities. Class imbalance is undesirable because it can skew model performance and accuracy. In this context, the disproportionate representation of classes may lead the model to prioritize the majority class (public universities) at the expense of the minority class (private universities). Consequently, the model's predictions may be biased towards the majority class, resulting in lower predictive accuracy and potentially overlooking important patterns or insights present in the minority class. This imbalance can pose challenges in accurately capturing the underlying relationships between features and the target variable, thereby hindering the model's ability to generalize effectively to new data.

To address the class imbalance issue, there are two primary techniques: oversampling and undersampling. Oversampling involves increasing the number of instances in the minority class by replicating existing samples or generating new synthetic data points using methods like SMOTE (Synthetic Minority Over-sampling Technique). This helps to balance the dataset by ensuring that the minority class is well-represented, allowing the model to learn patterns from it more effectively. On the other hand, undersampling reduces the number of instances in the majority class by randomly removing samples. This technique aims to create a more balanced dataset by aligning the size of the majority class with that of the minority class, which can help prevent the model from being biased towards the majority class. Both methods strive to enhance the model's ability to generalize by providing a more balanced representation of the classes, although they must be applied carefully to avoid potential drawbacks such as overfitting in oversampling or loss of important information in undersampling.

In this research, both oversampling and undersampling techniques are employed to tackle the class imbalance issue evident in the dataset of the undergraduate admission test survey. Initially, the 'RandomOverSampler' technique is used to increase the number of samples in each class to 500, thereby balancing the representation of private and public universities. By duplicating instances from the minority class randomly, this method ensures a well-balanced dataset, reducing the bias towards the majority class seen in the original distribution. Furthermore, a nuanced strategy is implemented where the majority class (public universities) undergoes undersampling, decreasing its instances to 226 through random selection without replacement. At the same time, the minority class (private universities) is oversampled, with instances replicated to match the reduced size of the majority class. Through this meticulous blend of undersampling and oversampling, the dataset is harmonized, with both classes having 226 instances each. This integrated method not only corrects the initial class imbalance but also enriches the dataset with diverse samples, leading to a more robust and generalizable model. Both oversampling and undersampling techniques are utilized not only to assess their efficacy but also to compare their performance against the unbalanced dataset, thereby discerning which approach yields superior results in improving 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 employed to convert categorical variables, such as "Where did you stay when preparing for the exam?" and "How was the educational status of your family?", into a machine-readable format suitable for analysis by machine learning algorithms. This process is vital as algorithms often require numerical data for processing, and categorical variables in their raw form pose a challenge. One-hot encoding resolves this issue by creating binary columns for each category within a variable. For example, categories like "Town" and "Village" within the "Where did you stay when preparing for the exam?" variable are transformed into separate binary columns, ensuring each category is distinct without implying any ordinal relationship. Without this encoding, there's a risk of the model erroneously inferring ordinal relationships between categories, potentially biasing results. For instance, if "Town" and "Village" were represented numerically without one-hot encoding, the model might mistakenly perceive one category as more significant than the other. Moreover, the absence of one-hot encoding can impede the model's ability to accurately distinguish between different categories within a variable, undermining its capacity to capture and generalize patterns effectively. Therefore, one-hot encoding is pivotal for maintaining the integrity and reliability of machine learning analyses, safeguarding the categorical nature of variables while facilitating their efficient processing by algorithms.

Label encoding plays a pivotal role in preprocessing ordinal variables, such as "How was your family's economic condition?" which comprises categories like "Good," "Medium," and "Average." Through this process, each category is assigned a unique integer based on its rank, preserving the inherent order within the data. This encoding ensures that the machine learning model recognizes the hierarchy among categories; for instance, understanding that "Good" supersedes "Average," which is superior to "Medium." By leveraging the relative significance of each category during both the training and prediction phases, label encoding enables more informed decision-making. Without it, the model may mistakenly treat ordinal variables as nominal, disregarding their underlying order and potentially leading to erroneous interpretations. For instance, the absence of label encoding could result in the model erroneously considering "Medium" as equivalent to or better than "Good," distorting analysis outcomes. Additionally, the model's ability to accurately weigh the relative importance of different categories would be compromised, undermining its predictive accuracy and generalization capabilities. Hence, label encoding is essential for preserving the ordinal nature of variables and ensuring the model effectively interprets and utilizes the hierarchy within the data.

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

Developing a robust machine learning model involves constructing a thorough classifier pipeline and defining precise hyperparameter grids, which are pivotal for influencing study outcomes. The classifier pipeline delineates a structured sequence of data processing steps and model fitting procedures, ensuring a methodical approach to model creation and assessment. By encapsulating feature scaling, dimensionality reduction, and model training within a unified framework, researchers can streamline the model development process, bolster reproducibility, and enable scalability. In tandem with the classifier pipeline, specifying hyperparameter grids empowers researchers to systematically explore and optimize model performance across a spectrum of parameter settings. These hyperparameters steer the behavior of machine learning algorithms and impact their ability to discern intricate patterns and generalize to unseen data. Through meticulous definition of hyperparameter grids tailored to each classifier, researchers can conduct exhaustive searches to pinpoint the optimal combination of parameters that maximize model efficacy. This iterative hyperparameter tuning process ensures peak predictive accuracy while mitigating risks of overfitting or underfitting, thereby fortifying model robustness and generalization prowess.

In this research, a diverse array of classifiers is instantiated, encompassing 14 distinct models meticulously crafted to capture the unique nuances inherent in the dataset. These classifiers, ranging from classical techniques like Logistic Regression and Decision Trees to cutting-edge ensemble methods such as Random Forests and Gradient Boosting, are encapsulated within pipelines integrating preprocessing steps such as one-hot encoding categorical variables and standardizing numerical features. Such pipeline architecture fosters consistency in data processing and model training across various classifiers, streamlining workflow and bolstering reproducibility. Furthermore, hyperparameter grids are meticulously delineated for each classifier, specifying the range of parameter values to be explored during grid search optimization. These grids facilitate exhaustive exploration of hyperparameter space, allowing for the identification of optimal parameter settings that maximize model efficacy. Serving as critical blueprints guiding the optimization process, hyperparameter grids offer granular control over the model's behavior and performance. Through systematic evaluation of diverse hyperparameter combinations, it's possible to uncover the most effective configurations, enhancing model generalization capabilities and alleviating overfitting concerns. Additionally, the adaptability of hyperparameter grids enables tailoring of the optimization process to suit dataset characteristics and research requisites, ensuring efficient and effective model development. Overall, the combination of classifier pipeline definition and hyperparameter grid specification facilitates systematic exploration and optimization of machine learning models.

4.4. Grid Search Optimization

In this research, Grid Search serves as a fundamental tool for fine-tuning the machine learning models by systematically exploring a predefined set of hyperparameters for each classifier within the pipeline. Through exhaustive evaluation of various hyperparameter combinations within specified ranges, Grid Search endeavors to identify the optimal configuration that maximizes the models' performance metrics, including accuracy, precision, recall, and F1-score. This thorough approach enables a comprehensive assessment of the classifiers across a diverse range of hyperparameter configurations, facilitating a detailed exploration of the model's potential. By precisely specifying the hyperparameter grids tailored to each classifier, the aim is to conduct a rigorous and exhaustive search, evaluating the models' performance under varying settings. Such systematic exploration ensures that the models are finely tuned and optimized, enhancing their predictive accuracy and robustness. Additionally, Grid Search allows for valuable insights into the interplay between hyperparameters and model performance, elucidating the complex dynamics governing the algorithms' behavior. By leveraging Grid Search, the models can be iteratively refined, uncovering the hyperparameter settings that yield the most effective predictive outcomes. This iterative optimization process is essential for enhancing the models' generalization capabilities, enabling them to perform optimally and provide reliable predictions. Thus, Grid Search stands as a cornerstone of the methodology, empowering the development of highly performant and reliable machine learning models.

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.

Specifically, 10-fold Cross-Validation is employed in this research to evaluate each classifier's performance across various hyperparameter settings. By dividing the dataset into 10 equal-sized folds and conducting model training and evaluation iteratively, researchers obtain more reliable performance estimates. This approach ensures the robustness of the models and their ability to generalize well to unseen data, as they undergo evaluation on diverse subsets of the dataset. Moreover, Cross-Validation facilitates the assessment of the stability of the models' performance metrics across different data partitions, providing valuable insights into their consistency and reliability. Furthermore, Cross-Validation serves as a means to optimize model hyperparameters effectively. By repeatedly training and evaluating the models on different subsets of the data, it is possible identify hyperparameter settings that yield consistently high performance across various folds. This iterative process aids in fine-tuning the models, enhancing their generalization capabilities and ensuring their reliability in making predictions on new, unseen data. Overall, Cross-Validation stands as a crucial component of the research methodology, enabling robust evaluation of model performance and fostering confidence in the reliability and generalization prowess 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 Python-3 experiment was carried out exclusively on Kaggle Notebooks, taking advantage of its user-friendly platform for machine learning tasks. The experiment made effective use of libraries such as pandas for data manipulation, scikit-learn for machine learning algorithms, and seaborn for advanced visualization, ensuring efficient analysis. The hardware resources included 13GB of RAM and a 16GB P100 GPU, which greatly facilitated computationally intensive tasks. The Python code imported essential libraries, utilizing pandas for dataset manipulation, scikit-learn for implementing and evaluating classifiers, and seaborn for visualization purposes. A range of classifiers, from Random Forest to advanced ensemble methods like XGBoost and CatBoost, were configured with hyperparameter grids to optimize performance. The experimental setup focused on comprehensive evaluation, employing k-fold cross-validation and confusion matrix visualization techniques. Stratified k-fold cross-validation was used to address class imbalance, while seaborn aided in generating informative confusion matrices. By combining Python, Kaggle Notebooks, and a well-structured experimental setup, the research aimed to provide robust insights into the analysis of undergraduate admission test survey data.

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 2: 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

The tables 2 and 3 present a comprehensive evaluation of various machine learning classifiers applied to the unaltered 'Undergraduate Admission Test Survey' dataset. This dataset comprises pertinent features utilized in assessing university admission eligibility. The evaluation focuses on key performance metrics: precision, recall, F1-score, and accuracy, providing insights into each classifier's effectiveness in predicting class labels. Additionally, weighted and macro average metrics are considered, offering further insights into classifier performance across Private University and Public University

Table 3: 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

classes. Weighted average reflects overall classification accuracy considering class distribution, while macro average indicates classifiers' ability to balance precision and recall across all classes, regardless of size.

The evaluation encompasses a diverse range of machine learning classifiers, each employing unique methodologies to model the underlying patterns within the dataset. This includes ensemble methods such as Random Forest, Bagging, and AdaBoost, as well as sophisticated algorithms like Support Vector Machine (SVM), Multi-layer Perceptron (MLP), and Gradient Boosting. These algorithms capture and interpret the complex relationships inherent in the admission test dataset. Additionally, traditional algorithms like Decision Tree, Logistic Regression, Naive Bayes, and K-Nearest Neighbors (KNN) are included to provide a comparative analysis of classic versus state-of-the-art approaches. This comparison offers insights into the performance trade-offs between simpler, more interpretable models and more complex, black-box algorithms. In total, 14 classifiers are assessed, representing a comprehensive array of machine learning techniques. This thorough assessment ensures a detailed exploration of the diverse landscape of machine learning methodologies, enabling a nuanced understanding of their efficacy in addressing the challenges posed by the admission test dataset.

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 achieve an accuracy of 83%. These ensemble methods show robust performance, particularly in recognizing "Public University" cases, with recall rates around 89-91%. However, their performance for "Private University" is slightly lower, with recall rates between 71-75%. Despite this, their overall F1-scores and balanced averages ensure they remain strong contenders for accurately classifying the dataset. The slight drop in performance for "Private University" might be attributed to class imbalance or the complexity of patterns specific to this class. Ensemble methods like Random Forest and Gradient Boosting are effective due to their ability to capture interactions between features, which simpler models might miss.

In contrast, classifiers like Decision Tree, Support Vector Machine (SVM), Logistic Regression, Multi-layer Perceptron (MLP), Naive Bayes, and K-Nearest Neighbors (KNN) show varying degrees of performance but generally fall short compared to the top performers. Decision Tree achieves 80% accuracy, with balanced precision, recall, and F1-scores, yet it is somewhat less consistent. SVM and Logistic Regression both achieve 79% accuracy, with higher recall for "Public University" but lower for "Private University" (59-60%), indicating limitations in handling class imbalances. MLP, with 76% accuracy, shows notable differences in recall between the two classes, suggesting a need for improvement in handling "Private University" cases. Naive Bayes and KNN, the least effective classifiers with accuracies of 75% and 74% respectively, particularly struggle with the "Private University" class, showing lower precision, recall, and F1-scores. These simpler models may lack the capacity to capture the complexity of the data patterns, highlighting the advantages of more sophisticated ensemble methods in achieving better results.

5.1.1. In-depth Confusion Matrix Analysis

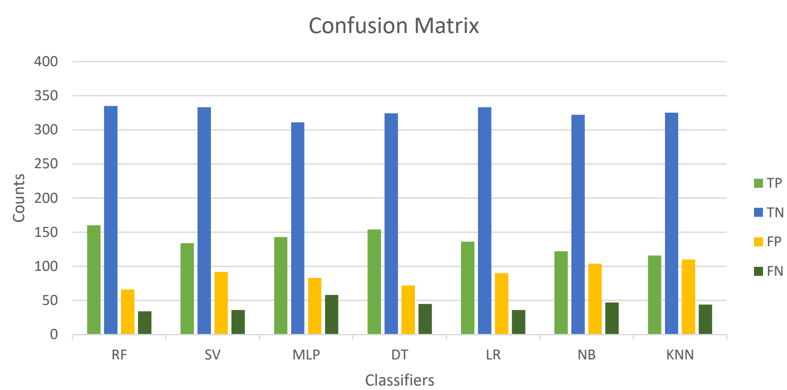


Figure 13: Confusion matrix illustrating the performance of different Classifier's on the dataset

In Figure 13 and 14, 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

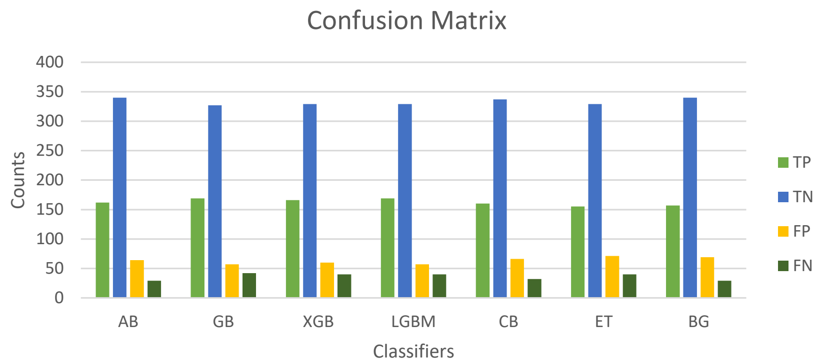


Figure 14: Confusion matrix illustrating the performance of different Classifier's on the dataset

metrics such as accuracy, precision, and recall, thereby furnishing a comprehensive understanding of the predictive capabilities of each classifier.

The provided confusion matrix offers a comprehensive evaluation of various classifiers, revealing distinct performance patterns across multiple metrics. AdaBoost (AB), XGBoost (XGB), LightGBM (LGBM), and CatBoost (CB) emerge as frontrunners, consistently demonstrating higher counts of True Positives (TP) and True Negatives (TN) compared to their counterparts. For instance, GB, XGB, and LGBM exhibit notable TP counts, with GB showcasing 169 TP instances, followed closely by LGBM and XGB with 166 and 169 TP instances, respectively. Similarly, these top-performing classifiers demonstrate robust TN counts, indicating their ability to accurately identify negative instances. Conversely, classifiers like Random Forest (RF), Support Vector Machine (SV), and Multi-Layer Perceptron (MLP) showcase a balanced performance across TP, TN, False Positives (FP), and False Negatives (FN) counts. While not excelling in any particular metric, these classifiers maintain stable performance across various classification tasks. Notably, RF achieves a respectable 160 TP instances and 335 TN instances, indicating its reliability in correctly classifying both positive and negative instances.

On the other hand, Logistic Regression (LR), Naive Bayes (NB) and K-Nearest Neighbors (KNN) classifiers demonstrate moderate performance, with slightly lower TP and TN counts compared to the top-performing classifiers. LR, for instance, records 136 TP instances and 333 TN instances, showcasing its capability in classification tasks but with a slightly lower accuracy compared to the top contenders. Lastly, classifiers such as Extra

Trees (ET) exhibit a performance profile akin to Gradient Boost (GB), with a moderate number of TP and TN instances but also higher FP and FN counts. ET, for instance, registers 155 TP instances and 329 TN instances, indicating its competence in identifying true positives and negatives, albeit with a slightly higher rate of false classifications.

This validates the prior analysis using different evaluation metrics, as both instances converge on the same conclusion. The superior performance of AdaBoost, XGBoost, LightGBM, CatBoost, and Bagging classifiers, evidenced by their higher counts of True Positives (TP) and True Negatives (TN), remains consistent across varied classification tasks. While Random Forest, Support Vector Machine, and Multi-Layer Perceptron demonstrate stable performance, they fall short of surpassing the top classifiers in TP and TN counts. Moreover, Logistic Regression, Naive Bayes, and K-Nearest Neighbors classifiers consistently exhibit moderate effectiveness. This alignment across different evaluation metrics strengthens the analysis, affirming the performance hierarchy among classifiers and its practical implications for classification tasks.

5.1.2. ROC Curve Analysis

In Figure 15, a visual representation of receiver operating characteristic (ROC) curves is depicted. These curves are a widely used method for assessing the performance of machine learning classifiers or classification models. They plot the true positive rate (TPR) against the false positive rate (FPR), where the TPR denotes the proportion of actual positive cases correctly identified, and the FPR signifies the proportion of negative cases incorrectly classified as positive. Each ROC curve corresponds to a different machine learning classifier, and the curve's shape and distance from the diagonal line indicate the classifier's discriminatory power. Alongside each curve, an Area Under the Curve (AUC) value is provided, which serves as a quantitative measure for evaluating the classifier's performance. A higher AUC value suggests that the classifier is better at distinguishing between positive and negative instances.

The ROC curve analysis illuminates distinct performance trends among the classifiers evaluated on the oversampled dataset. Notably, Random Forest Classifier, Decision Tree Classifier, XGB Classifier, Extra Trees Classifier, and Bagging Classifier achieve a perfect AUC score of 1.00, indicating impeccable discrimination between positive and negative instances. Following closely are MLPClassifier, LGBMClassifier, and CatBoostClassifier with

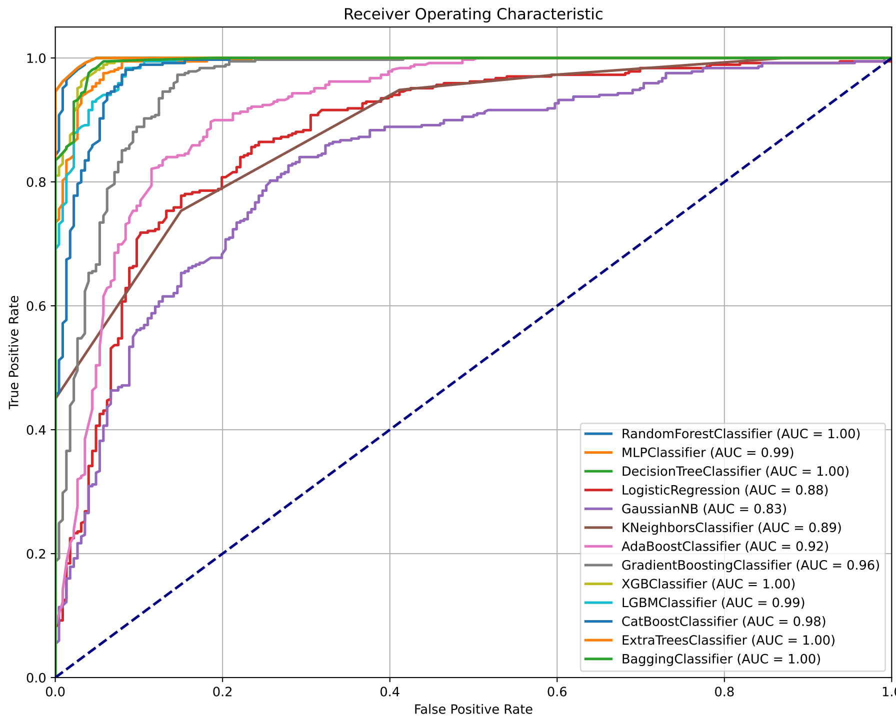


Figure 15: ROC curve comparison for various classifiers on the dataset

AUC scores of 0.99 and 0.98 respectively, showcasing exceptional discriminative capability. Leveraging sophisticated ensemble techniques, these classifiers effectively capture intricate patterns within the dataset, establishing them as highly dependable for classification tasks. In addition, Gradient Boosting Classifier demonstrates commendable performance with an AUC score of 0.96, reflecting robust discriminative power. AdaBoost Classifier achieves a respectable AUC of 0.92, while KNeighborsClassifier and LogisticRegression present relatively lower AUC scores of 0.89 and 0.88 respectively. GaussianNB trails behind with an AUC score of 0.83, indicating comparatively weaker performance in distinguishing between the two classes. Overall, the ROC curve analysis underscores the effectiveness of ensemble methods and sophisticated classifiers in accurately classifying instances in the oversampled dataset, with RandomForestClassifier and related models standing out as top performers.

However, while Bagging Classifier demonstrates the highest accuracy of 84%, alongside a perfect AUC of 1.00, the AdaBoost Classifier shares the same accuracy of 84% but with a slightly lower AUC of 0.92. These disparities between accuracy and AUC can stem from inherent distinctions in these metrics' nature. Accuracy measures the overall correctness of the classifier's predictions, whereas AUC evaluates its ability to differentiate between positive and negative instances by ranking them based on their predicted class probabilities. When a classifier attains high accuracy but a lower AUC, it may imply challenges in effectively ranking instances by their predicted probabilities despite most being correctly classified. Conversely, a high AUC suggests the classifier's proficiency in distinguishing between positive and negative instances, even if its overall accuracy is slightly lower. Hence, while accuracy offers a broad assessment of classifier performance, AUC provides deeper insights into its discriminatory capabilities.

Following a thorough evaluation spanning various metrics such as precision, recall, F1-score, accuracy, confusion matrices, and ROC curves, the Bagging Classifier emerges as the standout performer across multiple dimensions for this unaltered dataset. With an impressive accuracy of 84% and a flawless AUC of 1.0, it excels in both overall correctness of predictions and discriminatory capability in ranking instances based on their predicted probabilities. Furthermore, the Bagging Classifier demonstrates balanced performance across True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN), underscoring its reliability in accurately classifying both positive and negative instances. Thus, based on this comprehensive evaluation, the Bagging Classifier stands out as the optimal choice for effectively addressing the challenges presented by the undergraduate admission test survey dataset.

5.2. Performance Analysis with Undersampling

This section presents the performance analysis of various classifiers on the dataset after applying undersampling technique. The primary objective is to evaluate and compare the effectiveness of different machine learning classifiers using cross-validation and grid search for hyperparameter tuning. 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 4: 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 4 and 5 provide a detailed evaluation of 14 different machine learning classifiers applied to the undersampled, balanced 'Undergraduate Admission Test Survey' dataset. This dataset includes key features used to determine university admission eligibility. The evaluation emphasizes critical performance metrics: precision, recall, F1-score, and accuracy, to assess each classifier's effectiveness in predicting class labels. Additionally, both weighted and macro average metrics are considered. This evaluation offers an overview of how the classifiers' performance on the balanced dataset differs from their

Table 5: 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

performance on the original, unaltered dataset. By comparing these results, it is possible to gain insights into the impact of dataset balancing on the classifiers' ability to predict university admission outcomes accurately.

The Bagging, Random Forest, XGBoost, and CatBoost classifiers stand out with the highest accuracy rates, ranging from 89% to 90%. Their exceptional performance can be attributed to their inherent ability to handle complex datasets and capture intricate patterns effectively. Bagging achieves the top accuracy at 90%, exhibiting strong performance with a precision of 86% for Private University and 94% for Public University, and recall rates of 95% and 85% respectively. This results in high F1-scores of 90% and 89%, with consistent macro and weighted averages at 90%. Random Forest also excels with an overall accuracy of 89%, maintaining high precision (86% for

Private, 84% for Public), recall (93% for Private, 91% for Public), and F1-scores (90% for Private, 87% for Public). Both macro and weighted averages are balanced at 89%. Similarly, XGBoost and CatBoost achieve an accuracy of 89%, with precision, recall, and F1-scores showing minor variations. XGBoost has a precision of 86% for Private University and 92% for Public University, with recall rates of 92% and 85%, and F1-scores of 89% and 88%. CatBoost displays nearly identical metrics, with macro and weighted averages uniformly at 89%. These classifiers excel due to their ensemble or boosting methodologies, which enable them to effectively learn from data and generalize well to unseen instances, resulting in superior performance on the balanced dataset.

Gradient Boosting follows closely with an accuracy of 88%, maintaining high precision and recall rates for both Private (86% precision, 90% recall) and Public Universities (89% precision, 85% recall), resulting in F1-scores of 88% and 87%, and balanced macro and weighted averages at 88%. LightGBM, although slightly lower in accuracy at 87%, shows robust performance with comparable precision and recall rates, keeping its macro and weighted averages consistent at 87%. The success of these classifiers can be attributed to their gradient boosting techniques, which iteratively improve model performance by focusing on instances that were previously misclassified, thereby enhancing the overall predictive power of the model. On the other hand, Support Vector Machines (SVM), Multi-layer Perceptron (MLP), and Decision Tree classifiers exhibit similar performance with accuracies around 84-85%. SVM slightly underperforms with an accuracy of 84%, balancing precision and recall metrics around 82-87% for both classes. MLP and Decision Tree classifiers achieve an accuracy of 85%, with closely matched precision, recall, and F1-scores ranging from 83-88%. The effectiveness of these classifiers lies in their inherent capacity to handle non-linear relationships and complex decision boundaries, which are crucial for accurately predicting the class labels in the dataset.

Moderate performance is observed with Logistic Regression and K-Nearest Neighbors (KNN) classifiers, which show accuracies of 82% and 83%, respectively. Logistic Regression achieves a precision of 84% for Private University and 80% for Public University, with corresponding recall rates of 79% and 85%. KNN maintains balanced precision and recall rates around 81-85%. While these classifiers are effective in handling simpler datasets and linear relationships, their performance may suffer when dealing with more complex data structures and non-linear patterns. AdaBoost also demonstrates comparable performance with an accuracy of 83%, balancing precision and recall rates between Private and Public Universities, leading to macro and weighted averages around 83-84%. However, AdaBoost's performance is limited by its sensitivity to noisy data and outliers, which can affect its ability to generalize well to unseen instances. In contrast, Naive Bayes shows the lowest performance with an accuracy of 76%, particularly struggling with the Private University class (81% precision, 68% recall), though it performs better for Public University (73% precision, 85% recall). Naive Bayes' simplistic assumption of feature independence often leads to suboptimal performance, especially when dealing with correlated features or complex relationships within the data. 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

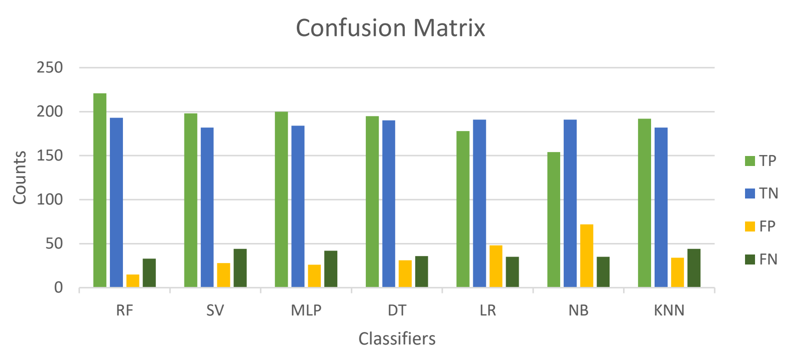


Figure 16: Confusion matrix illustrating the performance of different Classifier's on the undersampled dataset

The confusion matrices depicted in Figure 16 and 17 offer a comprehensive insight into the performance of each classifier on the undersampled balanced 'Undergraduate Admission Test Survey' dataset. In this matrix, TP (true positives) represent the instances correctly classified as positive, TN (true negatives) denote instances correctly classified as negative, FP (false positives) indicate negative instances incorrectly classified as positive, and FN (false negatives) signify positive instances incorrectly classified as negative.

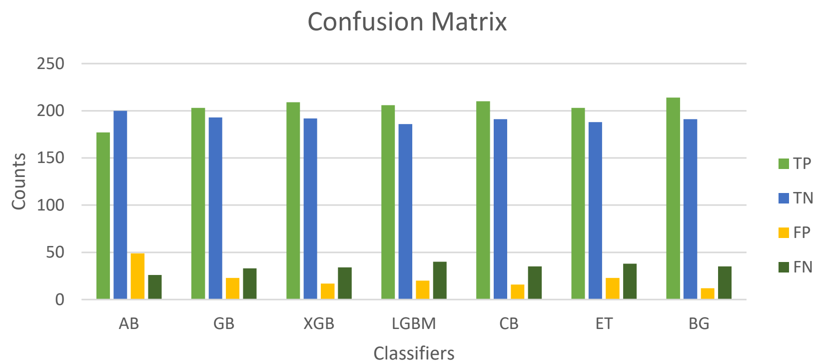


Figure 17: Confusion matrix illustrating the performance of different Classifier's on the undersampled dataset

These elements collectively offer insights into the classifier's ability to accurately predict class labels and distinguish between positive and negative instances within the dataset.

Random Forest (RF) demonstrates robust predictive power with 221 true positives (TP) and 193 true negatives (TN), indicating its proficiency in correctly identifying both positive and negative instances. Despite some misclassification errors, with 15 false positives (FP) and 33 false negatives (FN), RF maintains a strong overall performance. Similarly, Support Vector Machine (SV) exhibits a comparable pattern with 198 TP and 182 TN, albeit with a slightly higher number of FP (28) and FN (44), highlighting challenges in accurately distinguishing between classes. Multi-layer Perceptron (MLP) and Decision Tree (DT) classifiers display moderate performance, with TP and TN counts similar to RF and SV but slightly higher FP and FN rates. On the contrary, Logistic Regression (LR) and Naive Bayes (NB) classifiers face greater struggles, evident from their higher FP and FN counts, indicating limitations in their predictive abilities. For example, LR achieves 178 TP and 191 TN but registers 48 FP and 35 FN, showcasing significant misclassification errors.

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

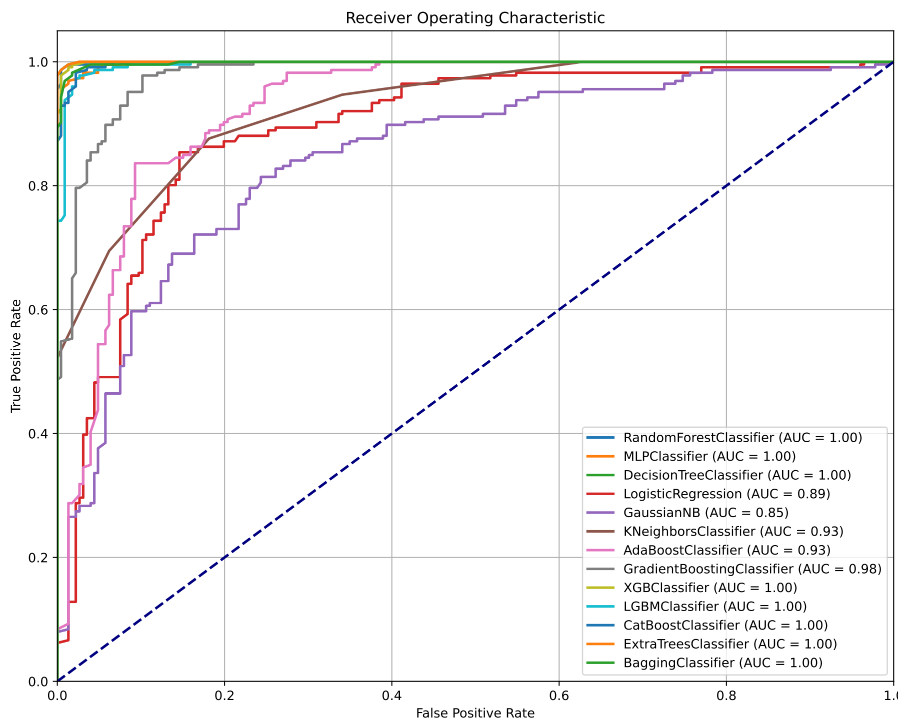


Figure 18: ROC curve comparison for various classifiers on the undersampled dataset

In Figure 18, the ROC curve depict the performance of different machine learning classifiers on the undersampled dataset. These curves show the true positive rate (TPR) against the false positive rate (FPR), assessing how well the models distinguish between positive and negative instances. Each curve represents a different classifier, with the Area Under the Curve

(AUC) indicating its overall performance. Higher AUC values signify better discrimination between positive and negative cases.

Several classifiers, such as Random Forest, MLP, Decision Tree, XGBoost, LGBM, CatBoost, Extra Trees, and Bagging, demonstrate outstanding performance with an AUC score of 1.00. This perfect AUC score indicates their exceptional ability to accurately differentiate between positive and negative instances, showcasing their robustness and high effectiveness on the balanced dataset. The remarkable success of these classifiers is largely due to their advanced ensemble methodologies, which combine multiple learning algorithms to enhance predictive accuracy and generalization capabilities. These models are particularly adept at identifying intricate patterns and relationships within the dataset, making them highly dependable for classification tasks. On the contrary, some classifiers exhibit varying levels of performance. While the Gradient Boosting Classifier achieves a strong AUC score of 0.98, indicating high discriminative capability, it falls short of perfection. Similarly, the KNeighbors Classifier and AdaBoost Classifier deliver respectable performances with an AUC of 0.93, signifying solid but not exceptional results compared to the top-performing models. Logistic Regression and Gaussian NB yield lower AUC scores of 0.89 and 0.85, respectively, underscoring their relatively diminished ability to distinguish between the two classes. These lower AUC values imply that Logistic Regression and Gaussian NB struggle more with the intricacies of the balanced dataset when compared to their more advanced counterparts.

While several classifiers achieve an AUC of 1.0, indicating flawless performance, it's essential to consider other metrics such as accuracy for a comprehensive evaluation. For instance, the MLP Classifier, despite its AUC of 1.0, demonstrates an accuracy of 85%, which is comparatively lower than other classifiers with high AUC. This discrepancy can be attributed to the dataset's complexity and the MLP Classifier's sensitivity to parameter tuning and training data size. Additionally, while AUC assesses the model's ability to rank true positives higher than false positives, accuracy evaluates the overall correctness of predictions, which might be affected by class imbalance or misclassification of minority classes. Conversely, the Bagging Classifier stands out with both the highest AUC and accuracy of 90%. This indicates not only its robust discriminatory capabilities but also its reliability in accurately classifying instances across both positive and negative classes.

After evaluating all the classifiers using various metrics, including precision, recall, F1-score, accuracy, confusion matrices, and ROC curves, the

Bagging classifier consistently emerges as the top performer. With the highest accuracy rate of 90%, minimal misclassification errors, and the highest Area Under the Curve (AUC) of 1.0, Bagging demonstrates superior predictive power and discriminatory capability. Its ability to handle complex datasets and minimize false positives and false negatives makes it the most reliable choice for accurately predicting class labels in the 'Undergraduate Admission Test Survey' dataset. Furthermore, it's worth noting that in the previous analysis on the unaltered dataset, the Bagging classifier also emerged as the best-performing one. However, in this case, its accuracy further improved to 90% when considering the undersampled balanced dataset, addressing the data imbalance issue. By employing undersampling techniques, the dataset's imbalance was mitigated, allowing classifiers like Bagging to achieve even better performance by making more accurate predictions across both positive and negative 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. The primary goal is to evaluate and compare the effectiveness of different machine learning classifiers using cross-validation and grid search for hyperparameter tuning. 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.

Classifier Evaluation and Performance Analysis on Oversampled Data Tables 6 and 7 present a detailed evaluation of 14 different machine learning classifiers applied to the oversampled, balanced 'Undergraduate Admission Test Survey' dataset. This dataset includes key features used to determine university admission eligibility. The performance of each classifier is assessed using critical metrics: precision, recall, F1-score, and accuracy, with both weighted and macro average metrics considered. This evaluation compares the classifiers' performance on the balanced dataset with their performance

Table 6: 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

on the original, unaltered dataset and an undersampled balanced dataset. This comparison provides insights into how the classifiers' ability to predict university admission outcomes is affected by the oversampling and balancing techniques used on the dataset.

Random Forest, Gradient Boosting, XGBoost, Light Gradient Boosting Machine, CatBoost, Extra Trees, and Bagging classifiers emerge as top performers, consistently demonstrating precision, recall, and F1-scores in the range of 90-93%. Their exceptional performance is underscored by high accuracy rates, typically around 91-92%. This remarkable consistency across metrics is indicative of their robustness in handling the balanced dataset.

Table 7: 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

These classifiers excel due to their ensemble nature, which leverages the strengths of multiple individual models to produce a more accurate and stable prediction. Specifically, ensemble methods aggregate predictions from multiple base estimators, effectively reducing overfitting and improving generalization. Moreover, these classifiers are less sensitive to noisy data and outliers, making them well-suited for handling the complexities inherent in a balanced dataset. For instance, the Random Forest classifier showcases 90% precision and 93% recall for private universities, culminating in an impressive overall accuracy of 92%. This exemplifies their strong predictive capability and reinforces their status as the top performers in this analysis.

In the mid-range performance tier, the Support Vector Machine (SVM), Multi-layer Perceptron (MLP), Decision Tree, and K-Nearest Neighbors (KNN) classifiers demonstrate balanced metrics, with precision, recall, and F1-scores typically ranging from 87% to 90%. These classifiers achieve overall accuracy rates of 88-90%, supported by consistent macro and weighted averages. These classifiers excel for specific reasons: SVM optimally separates data points, MLP learns complex patterns effectively, Decision Trees offer interpretability, and KNN considers data point proximity for accurate classification. These traits contribute to their mid-range performance, suitable for diverse classification tasks. The MLP classifier, for example, achieves an accuracy of 90%, with 89% precision for private universities and 91% for public universities, highlighting its dependable performance. Similarly, the Decision Tree classifier maintains a balanced precision of 88% for private universities and 93% for public universities. This collective performance underscores the adaptability and effectiveness of these classifiers in handling the nuances of the balanced dataset, positioning them in the mid-range performance tier.

The lower-performing classifiers in this analysis—Logistic Regression, Naive Bayes, and AdaBoost—face specific challenges. Logistic Regression shows a notable disparity in precision and recall between private (82% and 73%) and public universities (75% and 84%), resulting in a lower overall accuracy of 79%. This discrepancy indicates that Logistic Regression struggles with the balanced dataset's complexity. Naive Bayes, while achieving high recall for public universities (88%), suffers from significantly lower precision and recall for private universities (82% and 54%), leading to an overall accuracy of 71%. This inconsistency arises from Naive Bayes' assumption of feature independence, which often doesn't hold true in real-world data. AdaBoost, though relatively better, still lags behind top performers with an accuracy of 84%, showing a precision of 87% for private universities and 82% for public universities. AdaBoost's iterative focus on hard-to-classify instances can sometimes lead to overfitting, explaining its lower performance compared to the leading 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

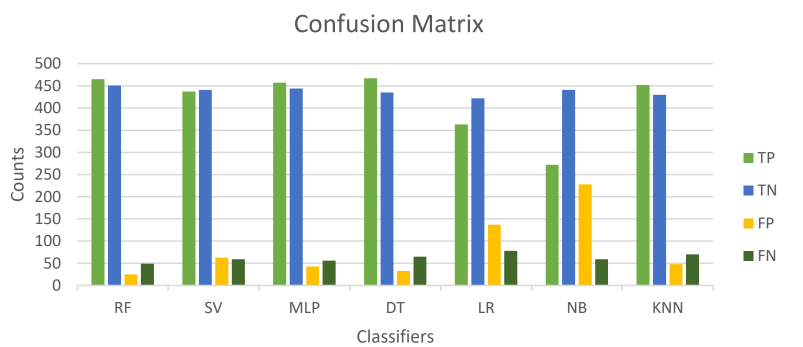


Figure 19: Confusion matrix illustrating the performance of different Classifier's on the oversampled dataset

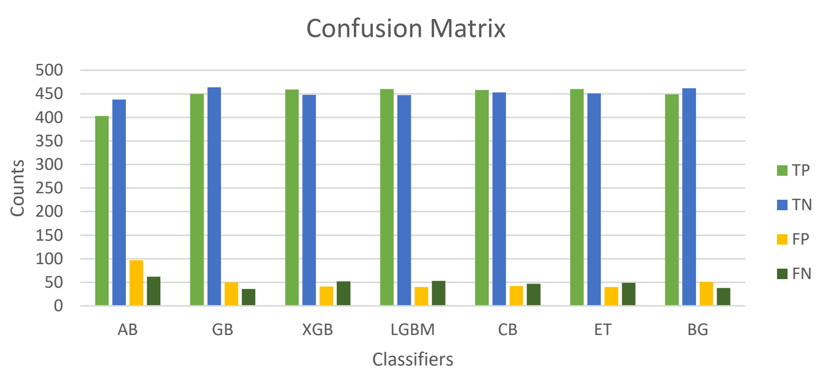


Figure 20: Confusion matrix illustrating the performance of different Classifier's on the oversampled dataset

Figures 19 and 20 illustrate the confusion matrices for each classifier applied to the oversampled balanced 'Undergraduate Admission Test Survey' dataset. These matrices provide detailed insights into classifier performance by displaying the counts of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). True positives represent instances correctly identified as positive, while true negatives are instances correctly identified as negative. False positives occur when negative instances are mistakenly classified as positive, and false negatives occur when positive instances are incorrectly classified as negative. By analyzing these components, one can assess the accuracy of the classifiers in predicting class labels and their effectiveness in differentiating between positive and negative instances in the dataset.

Random Forest (RF) shows a strong performance with 465 true positives (TP) and 451 true negatives (TN), indicating its high accuracy in correctly identifying both classes. It has 25 false positives (FP) and 49 false negatives (FN), demonstrating its robustness in prediction with minimal errors. Support Vector (SV) displays slightly lower performance with 437 TP and 441 TN, coupled with higher errors, 63 FP and 59 FN. Similarly, Multi-layer Perceptron (MLP) performs well with 457 TP and 444 TN, but has more errors than RF, with 43 FP and 56 FN. Decision Tree (DT) also shows good accuracy, with 467 TP and 435 TN, though it has 33 FP and 65 FN, indicating a tendency to miss more positive instances. Logistic Regression (LR), however, struggles significantly, with only 363 TP and 422 TN, and high error rates of 137 FP and 78 FN, suggesting it is less effective for this dataset. Naive Bayes (NB) performs poorly as well, with 272 TP and 441 TN, and particularly high FP (228) and FN (59), reflecting its limitations in handling the complexity of the dataset. K-Nearest Neighbors (KNN) shows decent performance with 452 TP and 430 TN, though it has higher errors than the top performers, 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 21 presents the ROC curves for various machine learning classifiers applied to the oversampled dataset. These curves plot the true positive rate (TPR) versus the false positive rate (FPR), providing a visual representation of each model's ability to differentiate between positive and negative instances. Each ROC curve corresponds to a different classifier, with the Area Under the Curve (AUC) serving as a key metric for overall performance evaluation. Classifiers with higher AUC values exhibit superior discrimination between the two classes, indicating more effective classification performance.

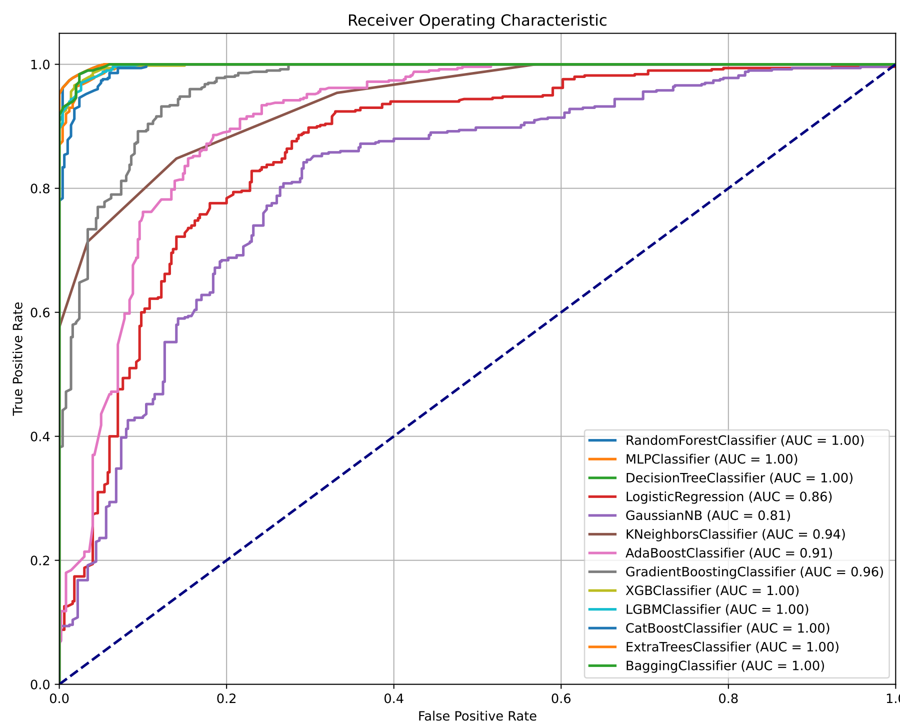


Figure 21: ROC curve comparison for various classifiers on the oversampled dataset

The ROC curve results for the classifiers on the oversampled dataset reveal significant differences in their ability to distinguish between positive and negative instances. Notably, the RandomForest Classifier, MLP Classifier, Decision Tree Classifier, XGB Classifier, LGBM Classifier, CatBoost Classifier, ExtraTrees Classifier, and Bagging Classifier all achieve a perfect AUC score of 1.00. This perfect score indicates that these models can flawlessly distinguish between the two classes without any error, underscoring their robustness and high performance on the balanced dataset. Such high AUC values reflect the effectiveness of these classifiers in capturing the underlying patterns in the data, making them exceptionally reliable for predictive tasks. On the other hand, the performance of other classifiers is more varied. The KNeighbors Classifier and GradientBoosting Classifier perform well, with AUC scores of 0.94 and 0.96 respectively, indicating strong but slightly less than perfect discrimination capability. AdaBoost Classifier follows closely with an AUC of 0.91, showing good performance though slightly lower than the top-tier models. Logistic Regression and Gaussian NB, however, exhibit lower AUC scores of 0.86 and 0.81, respectively. These lower scores suggest that these classifiers have more difficulty in accurately distinguishing between positive and negative instances compared to their higher-performing counterparts.

While several classifiers achieve an AUC of 1.0, indicative of flawless performance in distinguishing between positive and negative instances, a comprehensive evaluation must consider additional metrics such as accuracy. Despite achieving perfect AUC scores, classifiers like the Decision Tree and MLP Classifier exhibit lower accuracy rates of 90%, highlighting potential discrepancies between discriminatory power and overall correctness of predictions. Accuracy, unlike AUC, considers the overall correctness of predictions and may be influenced by factors like class imbalance or misclassification of minority classes. In contrast, the Random Forest Classifier stands out with both the highest AUC and accuracy of 92%, underscoring its robust discriminatory capabilities and reliability in accurately classifying instances across both positive and negative classes. This dual strength positions the Random Forest Classifier as a particularly effective model for this classification task.

After thorough evaluation across various metrics, including 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 exceptional performance across all metrics, with consistently high precision, recall, F1-scores, and accuracy rates, highlights its robustness and effectiveness in handling the complexities of the dataset. Moreover, the detailed analysis of the confusion matrix underscores Random Forest's ability to accurately predict both positive and negative instances, with minimal errors. The perfect AUC score of 1.00 further validates its impeccable discriminative capability, indicating flawless differentiation between positive and negative cases. Overall, Random Forest demonstrates a harmonious balance of high predictive accuracy, robustness, and discriminative power, making it the most suitable choice for classification tasks on oversampled balanced datasets.

5.4. Overall Comparative Analysis

The table 8 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 8: 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, Bagging, CatBoost, and LightGBM classifiers exhibit the highest accuracy, all achieving a score of 84%. Conversely, in the undersampled balanced scenario, the Bagging classifier outperforms others, attaining a notable accuracy of 90%, closely trailed by CatBoost and Random Forest, both achieving 89%. In the oversampled balanced scenario, Random Forest stands out as the top performer, boasting the highest accuracy of 92%. From the table, it is apparent that the dataset yields the best results in the oversampled scenario, where Random Forest achieves the highest accuracy of 92%. This is closely followed by Bagging, Extra Trees, CatBoost, and LGBM, all achieving an accuracy of 91%. In this case, where the dataset is balanced through oversampling, the classifiers demonstrate superior performance compared to the other scenarios. However, in the undersampled scenario, where data imbalance is addressed by reducing the majority class, the performance is slightly lower than in the oversampled case but still notably high, as seen with Bagging achieving 90% accuracy. Alternatively, in the unaltered scenario without any sampling adjustments, classifiers encounter challenges stemming from data imbalance, leading to diminished accuracy scores. Notably, Bagging achieves the highest accuracy of 84%, which is lower when compared to the balanced scenarios. This underscores the importance of addressing data imbalance issues to enhance classifier performance and emphasizes the significance of preprocessing techniques such as oversampling and undersampling for achieving optimal results.

Both Random Forest and Bagging classifiers consistently demonstrate robust performance across scenarios, excelling not only in accuracy but also achieving a perfect Area Under the Curve (AUC) score of 1 for all three scenarios discussed previously. While Random Forest emerges as the top performer in the oversampled balanced scenario, achieving 92% accuracy, it maintains strong performance with accuracies of 83% and 89% in the other two scenarios. On the other hand, Bagging exhibits remarkable consistency across all scenarios, securing the top position in the without sampled and undersampled scenarios with accuracies of 84% and 90%, respectively, and achieving 91% accuracy in the oversampled scenario. Both of these classifiers can be considered as optimum classifiers for this undergraduate admission test survey dataset, given their consistent and high performance across diverse scenarios.

5.5. Shapley-Based Feature Significance

Shapley values, derived from cooperative game theory, provide a robust and theoretically sound method for attributing the contribution of each feature to the predictions made by a machine learning model. Named after Lloyd Shapley, these values distribute the total gain (or prediction) among the features based on their contribution to different combinations of features. By considering all possible permutations of features, Shapley values ensure a fair allocation of feature importance. This method helps in determining the significance of each feature by quantifying how much each one contributes to the model's output, making it easier to interpret complex models and gain insights into the underlying decision-making process. Using SHAP (SHapley Additive exPlanations), these values can be visualized, providing clear and actionable insights into which features are most influential in driving model predictions. In the context of undergraduate admission test survey research, Shapley values are instrumental in identifying which attributes significantly influence whether a student is likely to get into a public or private university. By applying SHAP (SHapley Additive exPlanations) values, it is possible to determine the contribution of each feature—such as high school GPA, study habits, family background, and other personal factors—to the model's prediction. This detailed understanding allows for a more transparent and interpretable analysis of the factors that play a crucial role in university admission. Here, the top five classifiers that provide the best results have been taken into account to analyze their Shapley values and understand which features they are giving more significance. This comparative analysis helps in identifying the most reliable predictors across different models and enhances the robustness of the findings.

Random Forest Classifier:

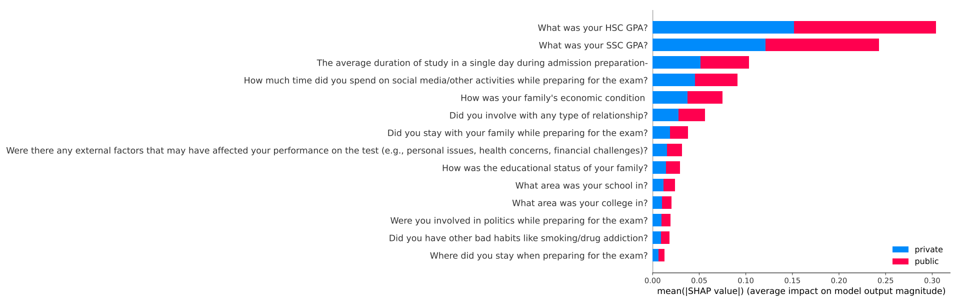


Figure 22: Feature Importance by Mean SHAP Value

The plot 22 illustrates the average impact of each feature on the output using SHAP (SHapley Additive exPlanations) values for the random forest classifier, which outperformed other classifiers. The x-axis represents the mean absolute SHAP value, indicating the average magnitude of a feature's impact on the predictions. Features are sorted by importance from top to bottom, revealing which attributes the random forest deems most influential in determining the outcome. Blue bars indicate the impact on predictions for private universities, while red bars denote the impact on predictions for public universities. The figure shows that 'HSC GPA' and 'SSC GPA' are the most influential features for predicting both public and private university admissions, with high SHAP values. This is expected, as GPA requirements are stringent for public universities in Bangladesh, while private universities have more flexible criteria. Other significant features include 'The average duration of study in a single day during admission preparation,' 'Time spent on social media/other activities while preparing for the exam,' 'Family's economic condition,' and 'Involvement in relationships.' These factors highlight the importance of study habits, social distractions, and socioeconomic background in university admissions. Less influential features, as indicated by lower SHAP values, are 'Where did you stay when preparing for the exam?,' 'Other bad habits like smoking/drug addiction,' and 'Involvement in politics while preparing for the exam?' These have a minor impact on predicting university admissions compared to academic performance and study habits. Overall, the figure underscores the critical role of academic performance in university admissions, while also highlighting the varying significance of study habits, economic conditions, and personal backgrounds for private versus public universities.

Bagging Classifier:

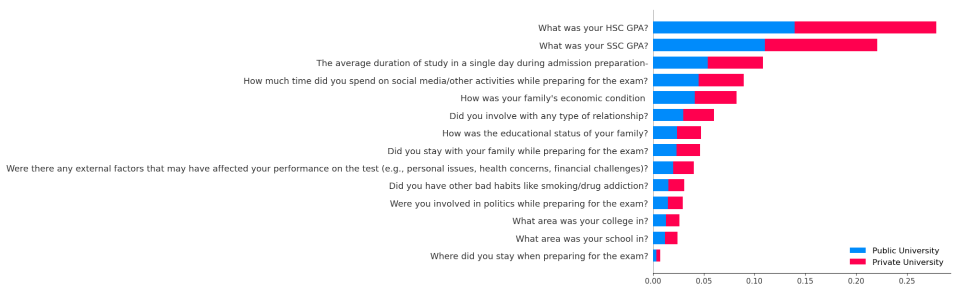


Figure 23: Feature Importance by Mean SHAP Value

The figure 23 illustrates the average influence of each feature on the output using SHAP (SHapley Additive exPlanations) values for a bagging classifier. The analysis shows that 'HSC GPA' and 'SSC GPA' are the most influential factors for predicting both public and private university admissions, similar to the results from the random forest classifier. These high SHAP values underscore the importance of GPA in the competitive admission landscape of Bangladeshi universities. Other significant features include 'The average daily study duration during admission preparation,' 'Time allocated to social media/other activities during exam preparation,' and 'Family's economic condition.' These factors consistently impact admission outcomes for both public and private universities, reflecting their importance in the student's overall preparedness and socio-economic background. This observation mirrors findings from the random forest classifier, highlighting the consistent significance of these factors across both classifiers. Less influential features, indicated by lower SHAP values, are 'What area was your college in?,' 'What area was your school in?,' and 'Where did you stay when preparing for the exam?' These factors have a relatively minor impact on predicting university admissions compared to academic performance and study habits, suggesting that geographical and environmental factors play a lesser role in the context of this analysis.

LGBM Classifier:

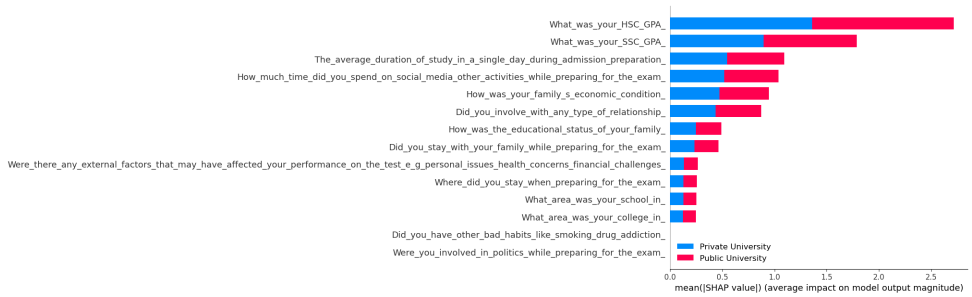


Figure 24: Feature Importance by Mean SHAP Value

The SHAP summary plot in fig. 24 for an LGBM (LightGBM) classifier illustrates the average impact of each feature on the model's output, distinguishing between predictions for private and public universities. "What was your HSC GPA?" has the highest impact, especially for public universities, followed by "What was your SSC GPA?" which significantly impacts both private and public universities, slightly more for the latter. "The average duration of study in a single day during admission preparation" and "How much time did you spend on social media/other activities while preparing for the exam" also have notable impacts, with a higher influence on public universities. "How was your family's economic condition" and "Did you involve with any type of relationship" have moderate to small impacts on both university types, again with a slightly higher influence on public universities. Conversely, features like "Were you involved in politics while preparing for the exam?" and "Did you have other bad habits like smoking/drug addiction?" have almost no impact on the model's predictions. These findings are consistent with results from two previous classifiers that demonstrated similar performance.

Gradient Boosting Classifier:

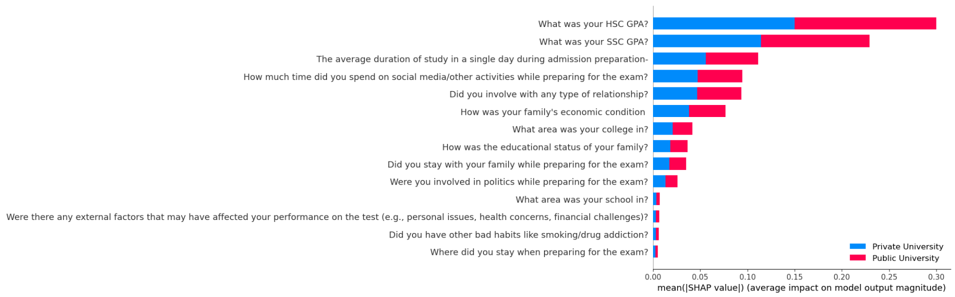


Figure 25: Feature Importance by Mean SHAP Value

Figure 25 illustrates the average impact of each feature on the model's output, determined by SHAP (SHapley Additive exPlanations) values for a gradient boosting classifier. The results are consistent with previous findings from other classifiers. The most influential features include "What was your HSC GPA?", "What was your SSC GPA?", "The average duration of study in a single day during admission preparation", "How much time did you spend on social media/other activities while preparing for the exam?", "How was your family's economic condition?", and "Did you involve with any type of relationship?". These features have the highest impact on the classifier's predictions, emphasizing the importance of academic performance and study habits in university admissions in Bangladesh. Conversely, features such as "Where did you stay when preparing for the exam?", "Did you have other bad habits like smoking/drug addiction?", and "Were there any external factors that may have affected your performance on the test (e.g., personal issues, health concerns, financial challenges)?" have minimal impact on the classifier's predictions. While these factors may influence individual circumstances, they are less significant in the broader context of university admissions decisions.

Extra Trees Classifier:

The figure 26 illustrates the collective influence of each feature on the model's output, calculated using SHAP values, with a focus on an Extra Trees classifier. From the depicted insights, it's apparent that this classifier, much like the other top-performing classifiers discussed earlier, draws its predictions primarily from a consistent set of features. Notably, "What was your HSC GPA?", "What was your SSC GPA?", "The average duration of

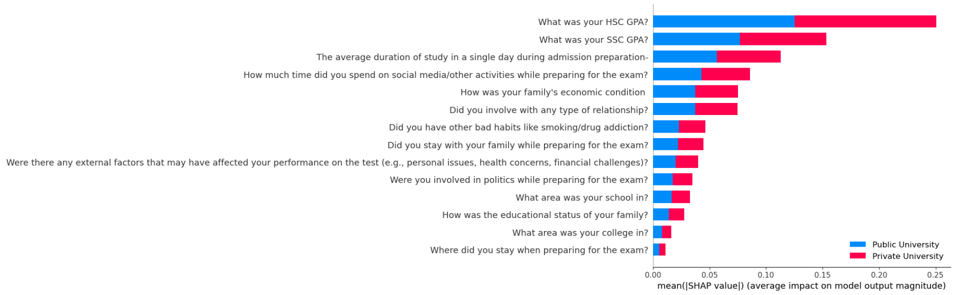


Figure 26: Feature Importance by Mean SHAP Value

study in a single day during admission preparation", "How much time did you spend on social media/other activities while preparing for the exam?", "How was your family's economic condition?", and "Did you involve with any type of relationship?" emerge as the most influential factors shaping the model's predictions. On the other hand, features like "Where did you stay when preparing for the exam?" and "What area was your college in?" are notably less influential in shaping the model's output. This indicates that while these factors may have some impact on individual circumstances, they play a relatively minor role in the broader context of university admissions predictions. This observation is consistent with the behavior of other top classifiers, which also assign lower importance to these features.

5.5.1. Key Determinants of University Admissions in Bangladesh

A comprehensive analysis of the top five classifiers, including Random Forest, Bagging, LGBM, Gradient Boosting, and Extra Trees, yields profound insights into the determinants of university admissions in Bangladesh. Across these classifiers, certain features consistently emerge as pivotal factors influencing admission outcomes, illuminating their critical role in shaping students' prospects of securing seats in public universities. Chief among these influential factors are academic performance metrics, notably "What was your HSC GPA?" and "What was your SSC GPA?", which consistently exhibit the highest SHAP values across all classifiers. This underscores their paramount importance in admission decisions. The stringent GPA prerequisites for public university admission underscore the pivotal role of academic excellence in clinching coveted spots in these institutions. Furthermore, within Bangladesh's fiercely competitive university admissions landscape, characterized by elite public institutions with rigorous GPA thresholds, academic performance emerges as a decisive determinant of admission outcomes. In contrast, private universities often offer more lenient GPA criteria, rendering them accessible to students with lower GPAs. Consequently, students falling short of the demanding GPA thresholds for public universities frequently turn to private institutions where GPA requirements are less stringent. Moreover, attaining such exemplary GPAs in HSC and SSC exams often signifies exceptional academic prowess, reinforcing the correlation between GPA and admission success. Notably, "HSC GPA" holds particular significance, given that post-HSC, students typically encounter admission tests. This significance is reflected in the figure by the highest SHAP value attributed to this attribute.

In Bangladesh, university admissions hinge on factors beyond mere GPA scores, encompassing elements such as study habits, family economic status, and social engagements during the exam preparation phase. Alongside academic metrics like "The average duration of study in a single day during admission preparation" and "Time spent on social media/other activities while preparing for the exam," considerations such as "Family economic condition" and "Involvement in any type of relationship" emerge as pivotal determinants, reflecting the country's fiercely competitive admissions landscape. Diligent study habits, gauged by the daily study duration, serve as a litmus test for a student's readiness and commitment to academic excellence. Those investing ample time in focused study sessions tend to exhibit superior performance in entrance exams, enhancing their prospects of securing coveted spots in prestigious public universities or gaining admission to private institutions. Conversely, excessive engagement in social media or non-academic pursuits can disrupt effective study routines, potentially undermining exam performance. In Bangladesh's hyper-competitive environment for public university seats, every hour devoted to meaningful study holds significant sway over one's admission prospects.

The economic background of a student's family profoundly influences their educational journey. Limited financial resources can create barriers to accessing quality study materials, enrolling in coaching classes, and even meeting basic needs. Students from economically disadvantaged backgrounds often face additional challenges in their academic endeavors, which can impact their performance in public entrance exams and their overall prospects for university admission. Moreover, involvement in romantic relationships during the exam preparation phase can serve as a significant distraction from studies. Consequently, students who are actively engaged in relationships during this critical period may find themselves allocating less time and energy to academic pursuits, potentially jeopardizing their performance in entrance exams and subsequent university admissions. These intertwined factors underscore the complex interplay between socioeconomic status, personal relationships, and academic achievement, highlighting the need for comprehensive support systems to address the diverse needs of students and promote equitable educational opportunities for all.

Factors such as geographical location ("Where did you stay when preparing for the exam?" and "What area was your college in?"), 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?") exhibit relatively lower SHAP values across all classifiers, suggesting their limited influence on university admissions decisions in Bangladesh. While these factors may impact individual circumstances, their broader significance in shaping admission outcomes appears marginal. This observation underscores the dominance of academic performance metrics and socioeconomic factors in the admissions process. In Bangladesh's cultural context, societal norms discourage early political involvement, and instances of smoking/drug addiction among youth are rare due to cultural and religious influences promoting healthy lifestyles. Furthermore, the emphasis on academic excellence in securing university admission overshadows the role of extraneous factors such as geographical location, personal habits, and political engagement. While personal habits and extracurricular activities may enrich a candidate's profile, their contribution to the admissions decision-making process seems comparatively minimal. This insight highlights the intricate interplay between cultural values, academic standards, and socio-economic factors in shaping the dynamics of university admissions in Bangladesh.

5.5.2. Discrepancies in Feature Importance Across Classifiers and Their Impact

In Figures 27 and 28, the examination of feature importance through mean Shapley values for Naive Bayes and logistic regression classifiers, which exhibit lower accuracy, unveils captivating insights, especially when juxtaposed against the top-tier classifiers with superior accuracy. Examining Figure 27, which depicts the Naive Bayes classifier, a notable observation emerges: the feature "Did you have other bad habits like smoking/drug addiction?" emerges as the most influential, despite being the second least important for the Random Forest classifier, which yields the best results. This

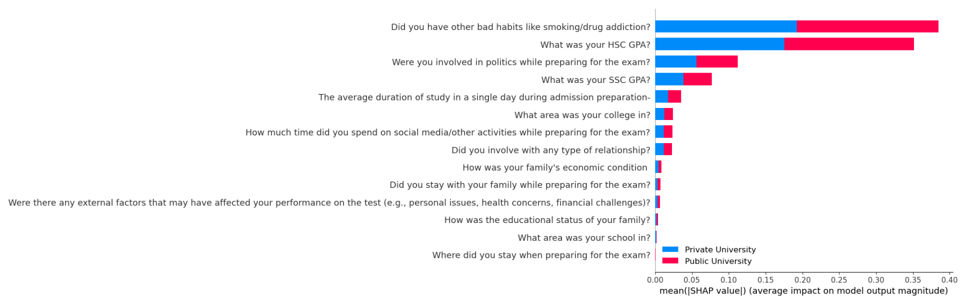


Figure 27: Feature Importance by Mean SHAP Value

discrepancy underscores the nuanced nature of feature importance across different models. Furthermore, the second most critical feature for the Naive Bayes classifier is "What was your HSC GPA?", while "Were you involved in politics while preparing for the exam?" ranks third in importance, contrasting with their positions as the least important features for the top classifiers. This discrepancy highlights a fundamental reason behind the poor performance of Naive Bayes—it fails to prioritize features that yield better results. By assigning undue importance to certain features that are less influential in determining admission outcomes, Naive Bayes neglects to focus on the critical predictors identified by top-tier classifiers. Consequently, its predictive accuracy suffers as it fails to leverage the most salient features for decision-making.

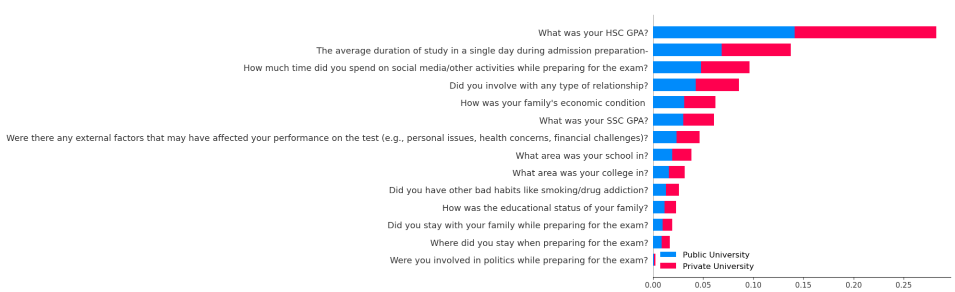


Figure 28: Feature Importance by Mean SHAP Value

In Figure 28, illustrating the results for logistic regression, a prominent observation arises: the pivotal feature "What was your HSC GPA?" emerges as the most crucial factor, aligning with expectations. However, a striking contrast becomes apparent when comparing it to the top-tier classifiers:

while "What was your SSC GPA?" consistently ranks as the second most important feature for these classifiers, it curiously occupies the sixth position in importance for logistic regression. This discrepancy sheds light on a potential contributing factor to the lower accuracy of logistic regression compared to its counterparts. The divergent weighting of features highlights the nuanced decision-making processes of different classifiers and underscores the significance of feature selection in classifier performance.

Ultimately, the comprehensive analysis underscores the critical importance of prioritizing relevant features to attain optimal classifier performance. Across various classifiers and contexts, the selection and weighting of features play a pivotal role in determining the efficacy of predictive classifiers, especially in complex domains such as university admissions. Neglecting to assign proper significance to pertinent features can result in suboptimal outcomes, as evidenced by the disparities observed between different classifiers. To achieve desirable results, it is crucial to focus more on important features that yield better results. SHAP (SHapley Additive exPlanations) provides a means to delve into these important features and conduct a comprehensive evaluation, enabling a deeper understanding of their impact on classifier performance and aiding in informed decision-making processes.

5.6. Local Interpretability with LIME Visualization

LIME, or Local Interpretable Model-agnostic Explanations, serves as a tool for interpretability that offers insights into the predictions made by machine learning models at a local level. In contrast to global interpretability techniques that seek to clarify the behavior of the model across the entire dataset, LIME focuses on elucidating individual predictions. This localized approach is essential for understanding how specific inputs impact a particular prediction, proving especially valuable in high-stakes scenarios where individual decisions must be carefully examined. In this study, LIME is employed to complement the global interpretability provided by SHAP (SHapley Additive exPlanations) values. While SHAP values present a comprehensive perspective on feature importance across various classifiers, LIME provides a more detailed analysis by explaining the contribution of each feature to individual predictions. This dual methodology offers a more comprehensive and nuanced understanding of the factors influencing the outcome of university admissions. By utilizing LIME, it becomes feasible to pinpoint and validate the crucial features highlighted by SHAP values, thereby gaining deeper insights into specific cases where the model's decisions may be questioned or necessitate further scrutiny. This integrated approach ensures a more transparent and interpretable analysis, ultimately enhancing the reliability and credibility of the results. The visual representations of the top six classifiers through LIME are showcased in this section.

Random Forest Classifier:

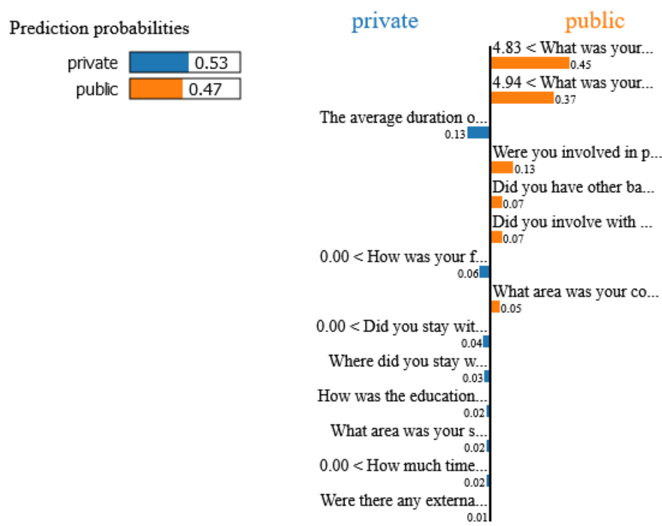


Figure 29: Random Forest Prediction Interpretation with LIME

The provided figure, Figure 29, presents a LIME (Local Interpretable Model-agnostic Explanations) visualization, which serves as a tool for elucidating individual predictions made by a Random Forest classifier. This visualization offers a detailed breakdown of the prediction process, commencing with the assessment of prediction probabilities. In this instance, the classifier allocates a probability of 0.53 (53%) to the "private" class and 0.47 (47%) to the "public" class, indicating a slight inclination towards the former category. It's worth noting that in this context, "public" and "private" denote public universities and private universities, respectively.

Delving deeper, the visualization displays the feature contributions, which highlight the significance of various attributes in determining the prediction. Features contributing to the "private" class are depicted in blue, while those impacting the "public" class are shown in orange. The length of each bar represents the magnitude of a feature's contribution, offering insights into its relative importance. Notably, for the "private" class, features such as "The average duration of study in a single day during admission preparation" and "Were you involved in politics?" hold considerable sway, each contributing 0.13 to the prediction. Conversely, for the "public" class, attributes like "What was your HSC GPA?" and "What was your SSC GPA?" exert significant influence, with contributions of 0.45 and 0.37, respectively. Despite the moderate certainty reflected in the close probabilities between private and public universities (0.53 and 0.47, respectively), the interpretation of feature contributions sheds light on the model's decision-making process. The top features contributed significantly to both classes, albeit with close margins, suggesting a balanced prediction. However, this also indicates potential uncertainty and less reliability for making clear-cut decisions in this specific instance. Thus, the interpretability provided by LIME aids in understanding the model's behavior and its implications for decision-making.

Despite achieving high accuracy, the Random Forest classifier can still face potential uncertainty and reduced reliability due to the complexity of its ensemble approach. The classifier aggregates predictions from multiple decision trees, each trained on a subset of the data. While this technique often results in robust performance, it can also create complex decision boundaries that are difficult to interpret. As a result, even though the classifier may have high overall accuracy, individual predictions can still be ambiguous, especially when dealing with closely balanced classes. Therefore, while Random Forest classifiers effectively capture intricate relationships within the data, their interpretability may be compromised. This highlights the importance of supplementary interpretability tools like LIME, which provide deeper insights into the decision-making process.

Bagging Classifier:

Figure 30 depicts a LIME visualization tailored for a Bagging classifier, offering detailed insights into individual predictions made by the model. This visualization provides a thorough breakdown of the prediction process, commencing with the assessment of prediction probabilities. In this instance, the Bagging classifier assigns a probability of 0.68 (68%) to the "Private University" class and 0.32 (32%) to the "Public University" class, indicating a distinct preference for the former category.

Further examination reveals the feature contributions, elucidating the significance of various attributes in shaping the prediction outcome. The length of each bar visually represents the magnitude of a feature's contribution, offering valuable insights into its relative importance. Notably, for the

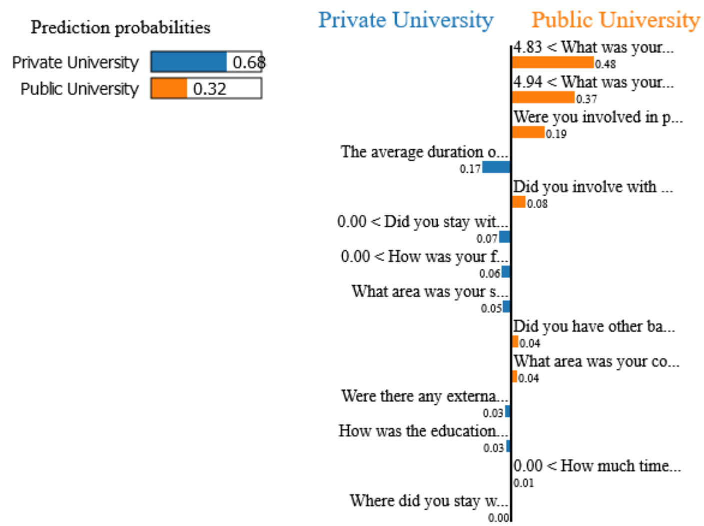


Figure 30: Bagging Prediction Interpretation with LIME

"Private University" class, features such as "The average duration of study in a single day during admission preparation" emerge with a substantial contribution of 0.17, followed by additional features contributing to a lesser extent. Conversely, for the "Public University" class, attributes like "What was your HSC GPA?" and "What was your SSC GPA?" exert significant influence, with contributions of 0.48 and 0.37, respectively. The cumulative effect of these feature contributions, which are additive in nature, ultimately determines the final prediction score. The Bagging classifier demonstrates a higher level of certainty in its prediction, with a clear preference for the "Private University" class, thereby affirming the robustness of its decision-making process.

LGBM Classifier:

Figure 31 showcases a LIME (Local Interpretable Model-agnostic Explanations) visualization for a LightGBM (LGBM) classifier, designed to elucidate the individual predictions made by the classifier. The visualization commences with a detailed analysis of the prediction probabilities. In this case, the LightGBM classifier assigns a probability of 0.39 (39%) to the "Private University" class and 0.61 (61%) to the "Public University" class, indicating a clear preference for the latter. This preference is evident in the bar chart located at the top left corner. Ultimately, the classifier predicts

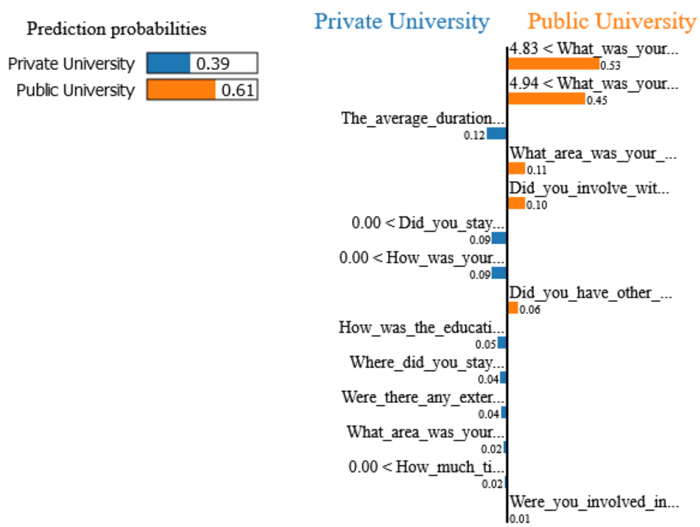


Figure 31: LGBM Prediction Interpretation with LIME

the instance as belonging to the "Public University" class with a probability of 0.61, demonstrating a high level of confidence in this classification.

Upon further investigation of the visualization, one can uncover the feature contributions, which emphasize the importance of different attributes in influencing the prediction outcome. For the "Private University" class, notable features include "The average duration of study in a single day during admission preparation" with a contribution of 0.12, "Did you stay with your family while preparing for the exam?" contributing 0.09, and "How was your family's economic condition?" also contributing 0.09. On the other hand, for the "Public University" class, significant contributions come from "What was your HSC GPA?" with 0.53, "What was your SSC GPA?" with 0.45, and "What area was your" contributing 0.11. The cumulative effect of these feature contributions, which are additive in nature, determines the final prediction score. The model's strong inclination towards the "Public University" class is underscored by the greater cumulative effect of the orange bars over the blue bars. The LightGBM classifier exhibits a high level of certainty in its prediction, with a clear preference for the "Public University" class, much like the Bagging classifier but with a different class preference.

Gradient Boosting Classifier:

In Figure 32, the LIME representation of a Gradient Boosting classifier

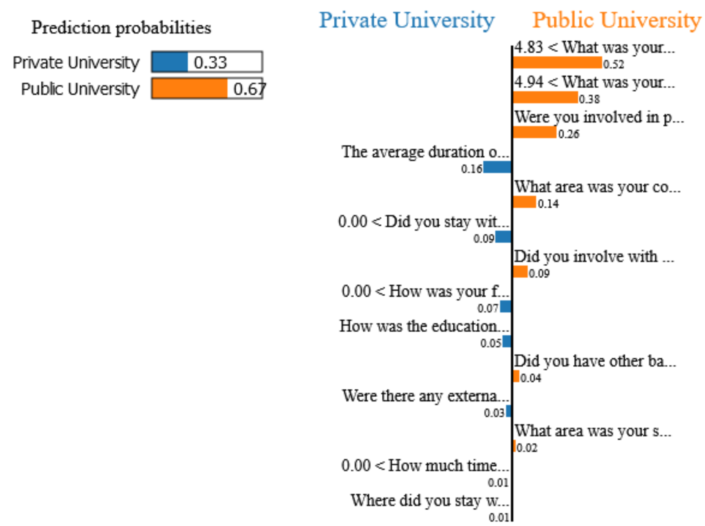


Figure 32: Gradient Boosting Prediction Interpretation with LIME

provides valuable insights into individual predictions. The analysis commences by showcasing the prediction probabilities, with a probability of 0.33 (33%) for the "Private University" category and 0.67 (67%) for the "Public University" category. This significant contrast in probabilities highlights a clear preference for the latter category, laying the foundation for a more profound comprehension of the classifier's decision-making process. Noteworthy features for the "Private University" category include "The average duration of study in a single day during admission preparation" contributing 0.16, followed by "Did you stay with your family while preparing for the exam?" at 0.09, and "How was your family's economic condition" at 0.07. In contrast, the "Public University" class is significantly influenced by features such as "What was your HSC GPA?" contributing 0.52, "What was your SSC GPA?" contributing 0.38, and "Were you involved in politics while preparing for the exam?" contributing 0.26. The Gradient Boosting classifier exhibits a high level of certainty in its prediction, with a probability of 0.67 for the "Public University" class. In comparison to other classifiers, the Gradient Boosting classifier shows strong performance with high certainty, similar to Bagging (68% for Private University) and LightGBM (61% for Public University), and higher certainty than Random Forest (53% for Private University).

Extra Trees Classifier:

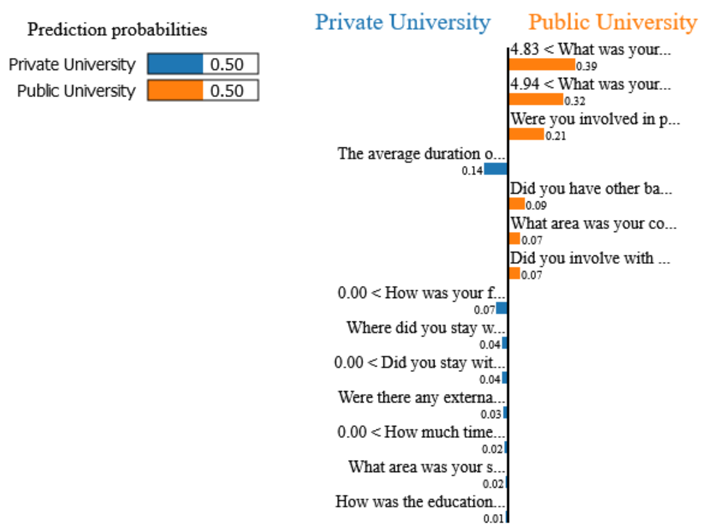


Figure 33: Extra Trees Prediction Interpretation with LIME

The Extra Trees classifier is visualized using a LIME visualization in Figure 33. In the top left corner of the visualization, a bar chart displays the prediction probabilities for the classes "Private University" and "Public University." In this instance, the classifier assigns equal probabilities of 0.50 (50%) to each class, indicating a lack of distinct preference. This balance is evident in the overall class prediction, suggesting that the classifier is uncertain and perceives the decision between "Private University" and "Public University" as evenly matched. When considering the "Private University" class, the feature "The average duration of study in a single day during admission preparation" contributes 0.14, while "How was your family's economic condition" adds 0.07, with other features playing a less significant role. Conversely, for the "Public University" class, notable contributions are made by "What was your HSC GPA?" with 0.39, "What was your SSC GPA?" with 0.32, and "Were you involved in politics while preparing for the exam?" with 0.21. Compared to other classifiers, the Extra Trees and Random Forest classifiers both show less certainty in their predictions. On the other hand, the Bagging, gradient boosting and LightGBM classifiers demonstrated higher confidence in their predictions. Despite the high accuracy, Extra Trees classifiers exhibit less certainty in individual predictions due to the ensemble nature, potential conflicts among decision trees, and the complexity of decision boundaries. This contrasts with the higher confidence seen in Bagging, gradient boosting, and LightGBM classifiers.

CatBoost Classifier:

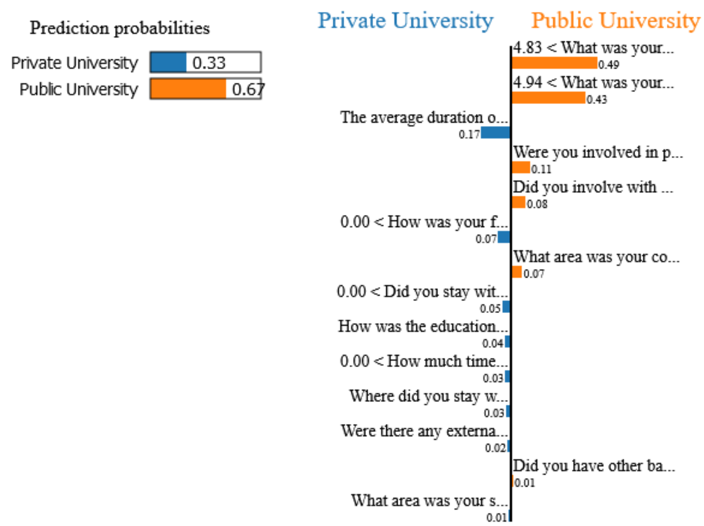


Figure 34: CatBoost Prediction Interpretation with LIME

Figure 34 presents a LIME visualization for the CatBoost classifier, starting with an intricate breakdown of prediction probabilities. It reveals a 33% likelihood for "Private University" and a 67% probability for "Public University." Notably, the classifier tends towards predicting the instance as belonging to the "Public University" class, evident from its higher 0.67 probability. Further analysis reveals the top contributing features for each class, with significant contributions highlighted. For the "Public University" class, features such as "What was your HSC GPA?" and "What was your SSC GPA?" stand out, emphasizing a clear preference for this class. Conversely, for the "Private University" class, features like "The average duration of study in a single day during admission preparation" and "How was your family's economic condition?" contribute less significantly. While the visualization aids in debugging models and understanding feature influences, it's noteworthy that CatBoost demonstrates high certainty in predicting the "Public University" class, aligning with the strong performance observed in other classifiers like Bagging, LightGBM, and Gradient Boosting. This certainty, coupled with clear feature contributions, positions CatBoost as one of the top-performing classifiers.

Across several classifiers, including Random Forest, Bagging, LightGBM, Gradient Boosting, Extra Trees, and CatBoost, the application of LIME (Local Interpretable Model-agnostic Explanations) offers a thorough and nuanced understanding of individual prediction dynamics in the context of university admissions outcomes. In high-stakes decision-making scenarios, transparency and accountability are crucial. LIME addresses this need by enabling a detailed analysis of how individual features impact each prediction, complementing the global interpretability provided by SHAP (SHapley Additive exPlanations) values. The essential characteristics indicated by SHAP values—such as HSC and SSC GPA, average study time for admission preparation, and family economic circumstances—can be precisely identified and validated by LIME visualizations, allowing for a deeper understanding of their influence on individual forecasts. By using a dual approach, the results are more credible and resilient, offering previously unknown insights into the behavior of the model. The visual aids highlight that classifiers such as Bagging, LightGBM, Gradient Boosting, and CatBoost show greater confidence in their predictions, whilst Random Forest and Extra Trees classifiers display greater ambiguity. This is probably because of the intricacy of their ensemble techniques. In the end, utilizing LIME makes it easier to go deeper into the various classifiers' decision-making processes, guaranteeing a more transparent and understandable study. By combining local and global interpretability tools, the forecasts become more reliable and the overall understanding of the factors influencing university admissions outcomes improves, which in turn fosters confidence in the decisions made.

6. Threats to Validity

This research, despite its strong approach and methodology, faces various potential threats to validity that require thorough examination. One significant concern is selection bias in terms of internal validity. The data was gathered through online surveys, which might not accurately reflect the entire pool of university applicants in Bangladesh. There is a possibility that the respondents who took part in the survey differ in a systematic way from those who did not, leading to potential selection bias. Moreover, the quality of the data is essential; the precision and completeness of the survey responses greatly influence the validity of the conclusions. Inaccurate self-reporting or missing data could impact the reliability of the outcomes, and although steps were taken to preprocess the data to address these issues, some inaccuracies may still persist. Increasing the volume of data collected could improve the validity of the results by offering a more comprehensive portrayal of the population.

Construct validity is a critical area of consideration. The attributes chosen for examination were derived from existing data and pertinent literature; nevertheless, there may be other significant factors affecting admission outcomes that were not included in the survey. The exclusion of such variables could potentially affect the findings of the study. Additionally, there are statistical validity concerns due to the sample size and the imbalance in the dataset between admitted students and those who were not. This imbalance could impact the statistical strength of the analysis. Despite the implementation of various sampling techniques to tackle this issue, the inherent imbalance may still have an impact on the results. Moreover, the heavy reliance on quantitative data analysis may overlook qualitative aspects that are essential for comprehending the complete context of admission obstacles. Incorporating qualitative research methods could offer a more comprehensive perspective.

Furthermore, alternative explainable AI approaches like Integrated Gradients, DeepLIFT, and Anchors can be utilized to validate the outcomes of machine learning models employed in the study, in addition to SHAP and LIME. By incorporating a wide range of explainable AI techniques, the interpretations of the model can be strengthened, offering more profound insights into the factors that impact admission outcomes. Although the study's methodology is solid, these potential validity concerns underscore the need for further improvement 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.

The findings highlight the crucial influence of academic achievement, socio-economic 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 fairer 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.

Subsequent research endeavors should build upon this study by utilizing longitudinal data to evaluate the lasting effects of admission policies and interventions. Furthermore, investigating the implementation of these strategies in various educational settings in Bangladesh and beyond could strengthen and broaden the conclusions. To counteract potential threats to the study's validity and bolster the robustness of the results, future investigations should concentrate on enlarging the dataset to encompass a more diverse sample of university applicants in Bangladesh, thereby reducing selection bias and enhancing data precision. By incorporating additional pertinent variables and employing a mixed-methods approach that integrates qualitative research, a more thorough comprehension of admission obstacles can be achieved. Additionally, delving into a wider range of explainable AI techniques can authenticate the machine learning models and provide deeper insights into the factors influencing admission results. These actions will refine the methodology of the study, leading to more dependable and comprehensive outcomes, and ultimately facilitating the advancement of data-informed and equitable decision-making processes in higher education.

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] 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.

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

[8] 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.

[9] M. Coccia, Deep learning technology for improving cancer care in society: New directions in cancer imaging driven by artificial intelligence, Technology in Society 60 (2020) 101198.

[10] A. Szymkowiak, B. Melović, M. Dabić, K. Jeganathan, G. S. Kundi, Information technology and gen z: The role of teachers, the internet, and technology in the education of young people, Technology in Society 65 (2021) 101565.

[11] F. Chien, A. Anwar, C.-C. Hsu, A. Sharif, A. Razzaq, A. Sinha, The role of information and communication technology in encountering environmental degradation: proposing an sdg framework for the brics countries, Technology in Society 65 (2021) 101587.

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

[26] 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.