IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 26, Issue 4, Ser. 1 (Jul. – Aug. 2024), PP 11-27
www.iosrjournals.org

A hybrid model for prediction of students' academic performance based on their admission pattern

Avom Patience[1], Ogala Emmanuel[1], Achir Jerome[1], Jude Akase and Nachamada Vachaku Blamah[2].
[1]*Department of Computer Science, Joseph Sarwuan Tarka University, Makurdi, Benue State, Nigeria.*
[2]*Department of Computer Science, University of Jos, Plateau State, Nigeria*

Abstract
The emergence of data mining techniques has encouraged researchers to attempt to apply them in the educational sector to discover knowledge from the students' data available to higher institutions of learning. This research work aims to develop a hybrid model that predicts academic performance based on admission pattern, thereby assessing the effectiveness of selected algorithms using metrics like Accuracy, Precision, Recall, and F1 Score. The study focuses on the prediction of students' academic performance using various data mining techniques, including Random Forest, Decision Tree, XGBoost, Support Vector Machine, K-Nearest Neighbors, Logistic Regression, Naive Bayes, and a Hybrid Algorithm which comprises Random Forest, Decision Trees, and K-Nearest Neighbors. Among the algorithms assessed, the Hybrid Algorithm emerges as the top performer across multiple key metrics. Notably, the Hybrid Algorithm demonstrates outstanding accuracy, precision, Recall, and F1 score, achieving remarkable values of 96.84%, 98.36%, 97, and 97.67%, respectively, surpassing all other algorithms. The significance of the study lies in its potential to provide early intervention and customized study schedules, thereby optimizing educational outcomes.
Keywords: *Decision tree, Random forest, K-Nearest Neighbors, Academic performance, Hybrid Algorithm*

Date of Submission: 26-06-2024 Date of Acceptance: 05-07-2024

I. Introduction

Since data mining techniques have become more popular, researchers have been more inclined to try using them in the educational sector to extract knowledge from student data that is accessible to higher education institutions (Baek & Doleck, 2022). Data mining which is historically referred to as knowledge discovery in data (KDD), entails the process of unveiling "hidden information," patterns, and knowledge within large volumes of data, hence, it is the practice of making predictions for outcomes or behaviors (Cruz & Encarnacion, 2021).

Educational institutions gather massive amounts of data that involve information regarding student application lists, admission records, as well as their examination results. Extracting such nature of data and the need to analyze the data generated from this educational ecosystem brings about the educational data mining (EDM) field. EDM is a field of research that employs data mining, machine learning, and statistical methods to analyze data obtained from educational settings, such as universities and intelligent tutoring systems (Abu, 2016; Zhang *et al.*, 2021)

In recent times, the prediction of students' academic performance has been a great area for researchers interested in EDM to explore and thus; has become an important factor in improving the educational process. The student's performance level may be affected by many factors related to the choice of course of study, the admission pattern, and average marks in the previous years (Aman *et al.*, 2019) emphasized that predicting students' academic performance in advance holds significant value for parents, higher education institution management, and the students themselves. Opting for the appropriate academic program at the right juncture can conserve time, effort, and resources for both parents and educational institutions.

Various techniques in data mining can be applied to educational data to draw out meaningful knowledge and help extract hidden and useful information. These techniques encompass Random Forest, Decision Tree, XGBoost, Support Vector Machine, K-Nearest Neighbors, Logistic Regression, Naive Bayes, and a Hybrid Algorithm.

This study utilized multiple machine learning algorithms for prediction tasks, encompassing Random Forest, Decision Tree, XGBoost, Support Vector Machine, K-Nearest Neighbors, Logistic Regression, Naive Bayes, and a Hybrid Algorithm. The hybrid algorithm employed Decision tree, Random forest and K-Nearest Neigbor for experimental purposes. The major objective of the proposed methodology is to build the hybrid model that predicts students' academic performance based on how they got admission.

II. Literature Review

Predicting students' academic performance is a multifaceted endeavor with profound implications for educational institutions and policymakers (Walia *et al*., 2020). Numerous studies conducted in this area show how different data-driven methods and models can be used to predict student performance.

One influential study by Namoun & Alshanqiti (2020) used machine learning algorithms to predict students' academic performance in Saudi Arabia. They employed data such as prior test scores, attendance records, and demographic information to build predictive models, demonstrating that machine learning techniques can provide valuable insights into academic achievement.

An innovative Students' Academic Performance Predicting (SAPP) system designed to address prevalent challenges and elevate prediction accuracy was introduced by (Kukkar *et al.*, 2023). Featuring an improved architecture, employed the SAPP system a well-integrated combination of a 4-layer stacked Long Short-Term Memory (LSTM) network, Random Forest (RF), and Gradient Boosting (GB) methods to predict whether students would pass or fail. In evaluating its effectiveness, the developed SAPP system underwent a comprehensive comparison with established prediction systems using the publicly accessible student OULAD dataset, supplemented by a self-curated emotional dataset. Performance assessment encompasses key metrics such as Accuracy, Precision, F-measure, and Recall. Comparative analysis involves benchmarking against LSTM + RF, LSTM + GB, and end-to-end deep learning models including ANN, LSTM, RNN, Convolutional Neural Networks (CNN), as well as widely utilized machine learning models like Support Vector Machine (SVM), Decision Tree (DT), Naive Bayes (NB), and RF. The SAPP system demonstrated an impressive prediction accuracy of approximately 96%, surpassing existing systems in the literature. This study encompassed three primary objectives: firstly, it comprehensively explored the contextual intricacies of rural education management; secondly, it investigated the correlation between continuing education at the upper secondary level; and thirdly, it developed a suitable predictive model for educational programs tailored to students attending high school in a rural environment.

The research carried out by (Razaque & Alajlan, 2020) investigated the use of six machine learning models to evaluate and examine the students' academic performance: Decision Tree, Random Forest, Support Vector Machine, Logistic Regression, ADA boost, and Stochastic Gradient Descent. The performance was assessed using the following metrics: f-measure, sensitivity, accuracy, and precision. The outcomes confirm that, of the models that were chosen, Stochastic Gradient Descent has a higher training dataset efficiency. Furthermore, it generates greater accuracy in comparison to other models. The goal of this contribution is to create the best model possible for determining students' academic success. That is, it maps inputs to the desired outputs.

An analysis of the variables influencing students' academic achievement was carried out as part of Ethiopia's New Medical Education Initiative (Gebru & Verstegen, 2023). Between December 2018 and January 2019, a survey was conducted involving the distribution of a structured self-administered questionnaire to students from four randomly chosen medical schools. The questionnaire encompassed inquiries regarding participants' socio-demographic and educational backgrounds. Using multiple linear regression analysis, the variables correlated with academic performance were found. To investigate qualitatively, in-depth interviews were done with fifteen important informants. Stress was linked to poorer academic performance in the multiple linear regressions. Students with other bachelor's degrees performed worse than those with prior health science education. Performance was also significantly predicted by the previous bachelor's degree cumulative grade point average and the medical school entrance exam score. Although additional variables were revealed from the qualitative interviews, its findings corroborated the survey results. Out of all the predictor variables examined in the model, only stress, prior educational degree, performance in the prior degree, and entrance exam score were significantly correlated with the performance of students in their preclinical medical engagement.

III. Methodology

This study presents a predictive model designed to predict students' academic performance by analyzing their admission patterns. The proposed methodology comprises five key phases: data collection, data pre-processing, sub-dataset generation, application of classification algorithms, and evaluation.

3.1 Data collection

The dataset to be used in this study will be obtained using a questionnaire-based survey from Taraba State University (TASU). The data will be collected for the academic sessions 2019-2021 of some randomly selected departments of the institution. It will include their course of study, choice of course, student CGPA, and their interest in the current course of study.

3.2 Data preprocessing

Pre-processing holds a crucial role in data mining, aiming to transform raw data into a suitable format usable by mining algorithms. The following tasks are executed in this phase:

i. **Data integration**: Data integration is the process of combining data from different sources into one repository. During this process, redundancy in the integrated data is a common challenge.
ii. **Data cleaning**: In this phase, missing and noisy data is handled to achieve data consistency.
iii. **Discretization**: Discretization is the process of transforming continuous data into discrete values or intervals. This technique is applied to numeric data and involves dividing a range of values into intervals or bins. The primary objective of discretization is to handle continuous data more effectively, making it suitable for certain types of analysis and algorithms that work with categorical or discrete data.

3.3 Data balancing

A data balancing approach is implemented after data pre-processing to address the challenge of class imbalance. The class imbalance problem occurs when the number of instances in one class significantly differs from the number of instances in another class or other classes

3.4 Feature selection

There could be a lot of attributes in the student performance dataset, some of which may not be suitable for classification. When a large number of student factors are included such as educational history, family information, financial position, and social demographics, all that can affect a student's performance, high-dimensional data becomes problematic. This issue can be solved by selecting the most relevant features from the dataset.

The aim of feature selection is to find a suitable subset of features that efficiently describe the input data, thereby reducing the dimensionality of the feature space and remove irrelevant data. Fisher's score is among the most commonly used supervised feature selection methods. The algorithm we employed ranks the variables based on Fisher's score.

In this study, the Fisher's score method was applied utilizing an information gain-based selection algorithm to evaluate feature ranks. The goal was to find which features are most important for building a student performance model. Each feature was assigned a rank value based on its influence on data classification during the feature selection process.Table 1, shows the list of some selected features as sample feature

Table 1. Samples of Selected Features

Course at application	Course of Study	Level of interest	Current level	CGPA	Remark
Mathematics	Mathematics	2	400	2.39	0
Mathematics	Mathematics	5	400	3.42	1
Mathematics	Mathematics	5	400	3.57	1
Mathematics	Mathematics	5	400	3.72	1
Computer science	Mathematics	2	400	2.23	0
Mathematics	Mathematics	5	400	4.59	1
Statistics	Mathematics	4	400	2.07	0

3.5 Model construction

The literature review generally recommends that there is no single classifier that works best in all contexts to provide accurate prediction.

In this paper, the hybrid model for students academic performance prediction is build using ensemble method. Ensemble learning is a machine learning model called the Model Combiners or multi-classifiers. By combining the abilities of two or more learners, it generates a final prediction with better performance than any other single model (Kaviyarasi & Balasubramanian, 2020). It is admitted that the ensemble model combines the performance of the individual weak models to improve performance (Vanerio & Casas, 2017).

To explore the potential benefits of combining classifiers, a hybrid algorithm is created. The proposed Hybrid Algorithm synergistically leverages the predictive capabilities of the Decision Tree and Random Forest classifiers. This fusion is achieved by averaging their respective predictions, leading to a more robust and balanced prediction outcome. Subsequently, a novel hybrid feature matrix is constructed by combining the original feature set with the ensemble of averaged predictions. This innovative matrix is then utilized to train a KNN classifier, which harnesses the combined power of the constituent classifiers. The performance assessment of this Hybrid Algorithm is carried out using the identical set of evaluation metrics as employed for the individual classifiers. This rigorous evaluation not only establishes the efficacy of the hybrid approach but also facilitates a comprehensive comparison against the performance of the standalone classifiers.

3.6 Evaluation Metrics

The performance of each classifier is evaluated using a set of standard evaluation metrics:

- **Accuracy:** Measures the overall correctness of the classifier's predictions.
- **Precision:** Indicates the proportion of correctly predicted positive instances among all predicted positives.
- **Recall:** Measures the ability of the classifier to correctly identify positive instances among all actual positives.
- **F1-Score:** Balances precision and recall, providing a single metric to evaluate a classifier's performance.
- **Confusion Matrix:** Provides insights into the number of true positives, true negatives, false positives, and false negatives.

3.7 Architecture of the Developed System

The developed system architecture is seen in Figure 3 below, which comprises five major phases which include: the data collection phase, data pre-processing phase, feature selection phase, model construction, and the model evaluation phase.

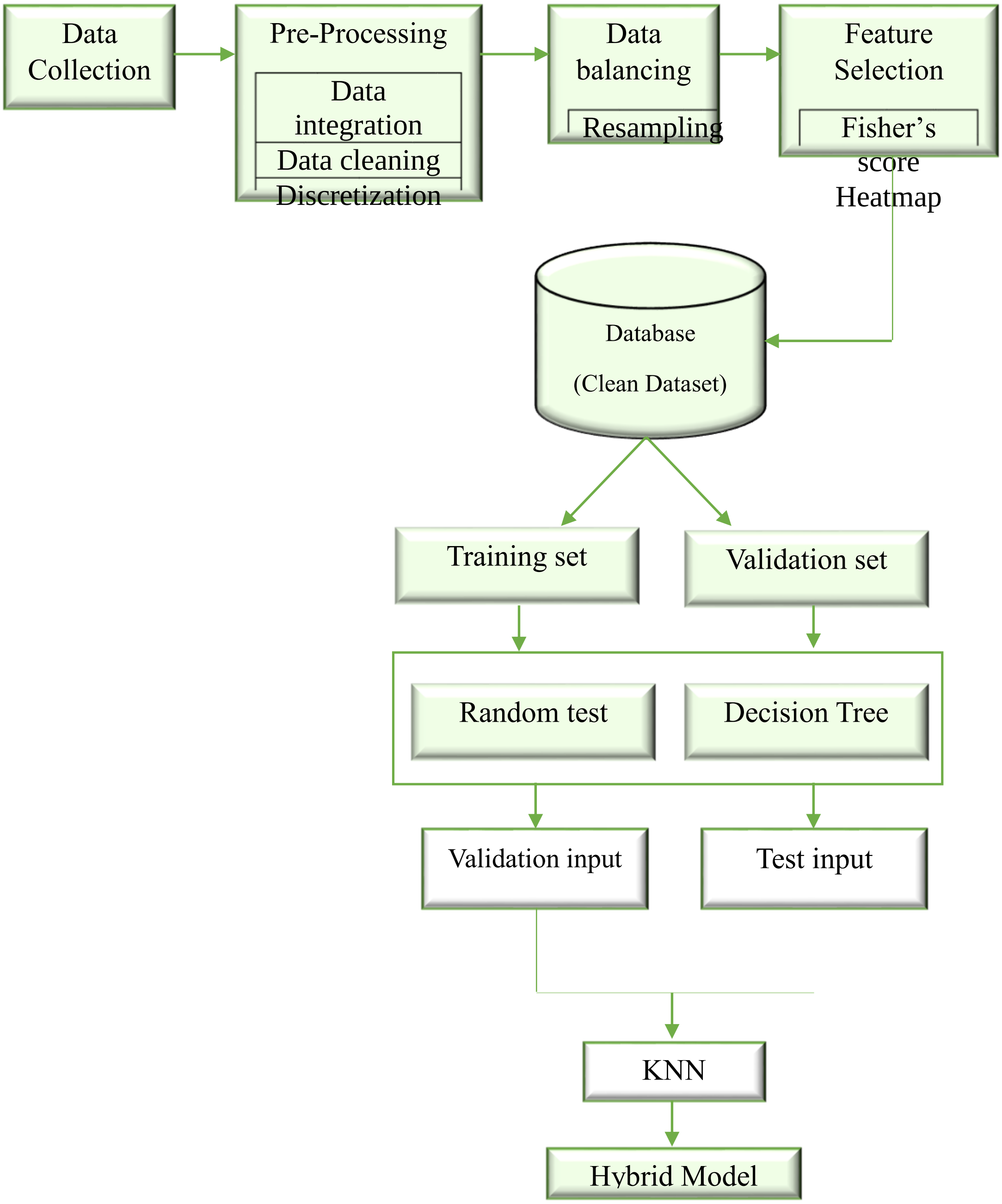

Figure 1: Architecture of the developed model

IV. Results and Discussion

4.1 Model evaluation

In this chapter, various machine learning algorithms were employed for prediction purposes, including Random Forest, Decision Tree, XGBoost, Support Vector Machine, K-Nearest Neighbors, Logistic Regression, Naive Bayes, and a Hybrid Algorithm. The performance of these algorithms was evaluated using metrics such as accuracy, precision, recall, and F1 score. Additionally, the Area Under the ROC Curve (AUC) was utilized to assess the overall prediction accuracy of each algorithm. The confusion matrix was also examined to gain insights into the false positive and false negative rates associated with each algorithm's prediction.

4.2 Evaluation Matrics

The performance of each classifier is evaluated using a set of standard evaluation metrics:

❖ **Accuracy:** Measures the overall correctness of the classifier's predictions.

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

❖ **Precision:** Indicates the proportion of correctly predicted positive instances among all predicted positives.

$$\text{Precision} = \frac{TP}{TP + FP}$$

❖ **Recall:** Measures the ability of the classifier to correctly identify positive instances among all actual positives.

$$\text{TPrate} = \frac{TP}{TP+FN}$$

❖ **F1-Score:** Balances precision and recall, providing a single metric to evaluate a classifier's performance.

❖ **Confusion Matrix:** Provides insights into the number of true positives, true negatives, false positives, and false negatives.

4.3 Result Analysis

This chapter presents the outcomes of the experiment conducted to predict student performance based on their admission pattern. Various machine learning algorithms were employed for prediction purposes, including Random Forest, Decision Tree, XGBoost, Support Vector Machine, K-Nearest Neighbors, Logistic Regression, Naive Bayes, and a Hybrid Algorithm. The performance of these algorithms was evaluated using metrics such as accuracy, precision, recall, and F1 score is presented in Figure 2.

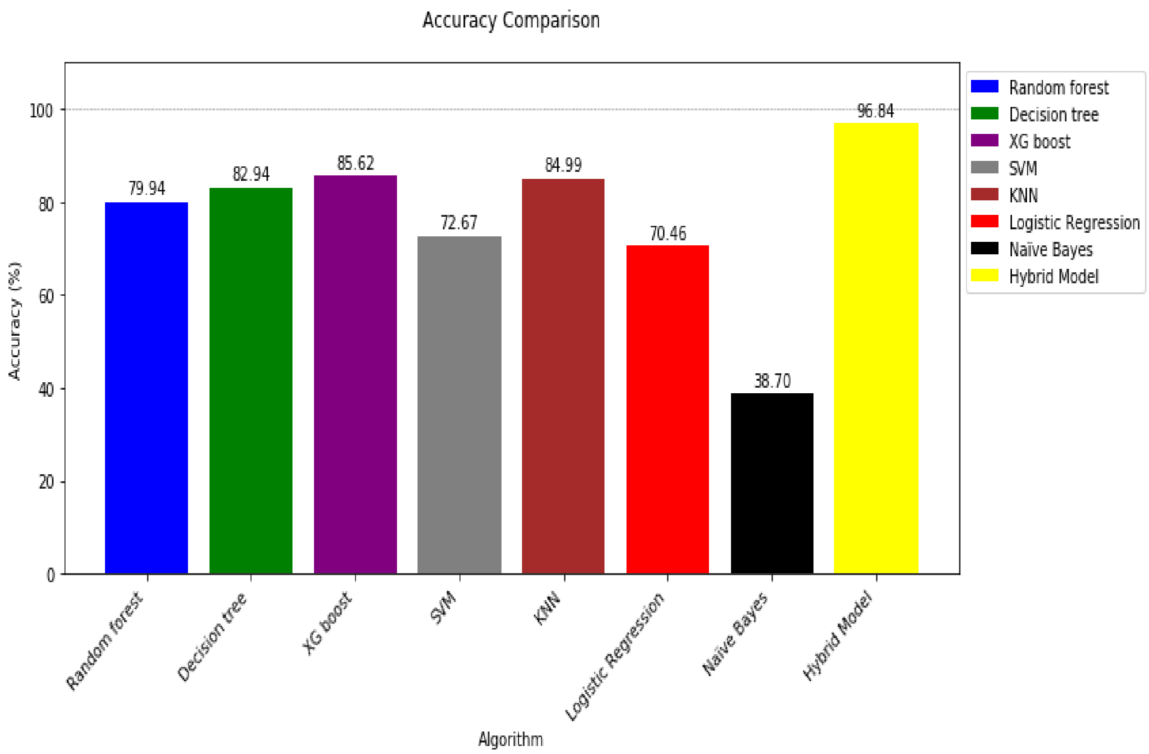

Figure 2: A bar chart showing the accuracy of Different ML Algorithms.

From the figure above, The Hybrid Model stands out with the highest accuracy score of 96.84%, surpassing all other algorithms. Following closely are XG boost with 85.62% and KNN with 84.99%. Logistic Regression and Naïve Bayes trail behind, demonstrating the lowest accuracy among the listed algorithms.

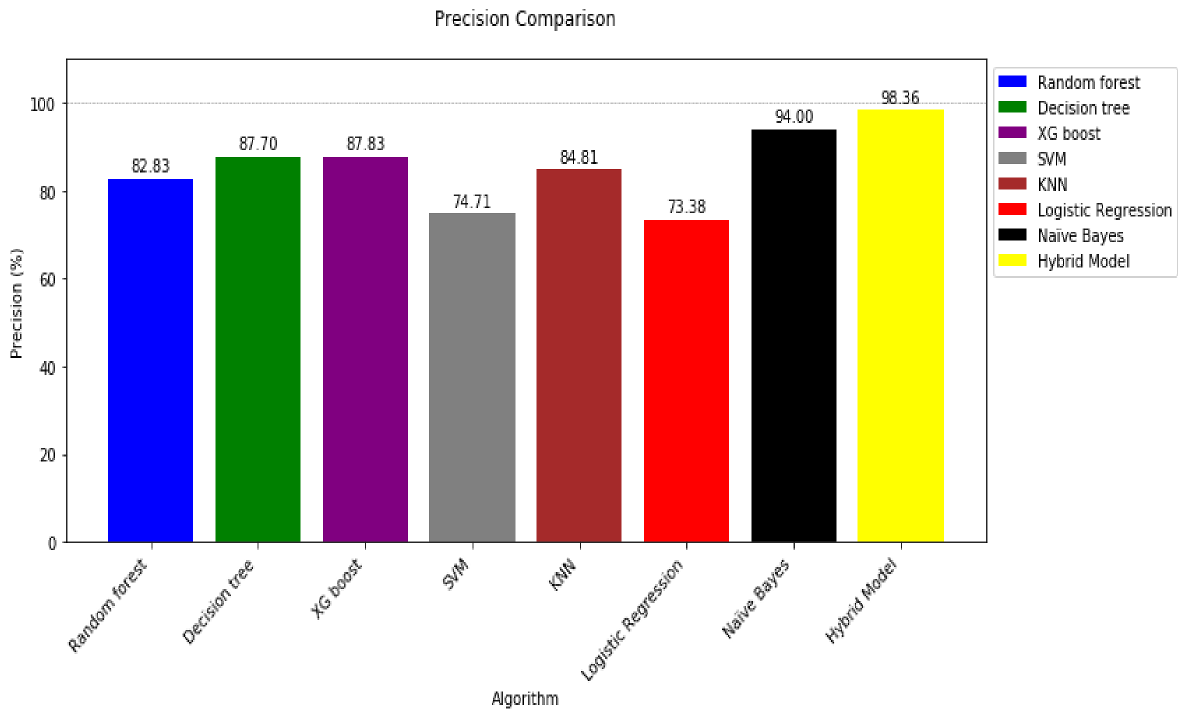

Figure 3: A bar chart showing the precision of Different ML Algorithms

From the figure above, the Hybrid Model exhibits the highest precision at 98.36%, outperforming all other algorithms. Random forest and KNN follow suit with precision scores of 82.83% and 84.81% respectively. However, Naïve Bayes and Logistic Regression display significantly lower precision scores compared to the top performers.

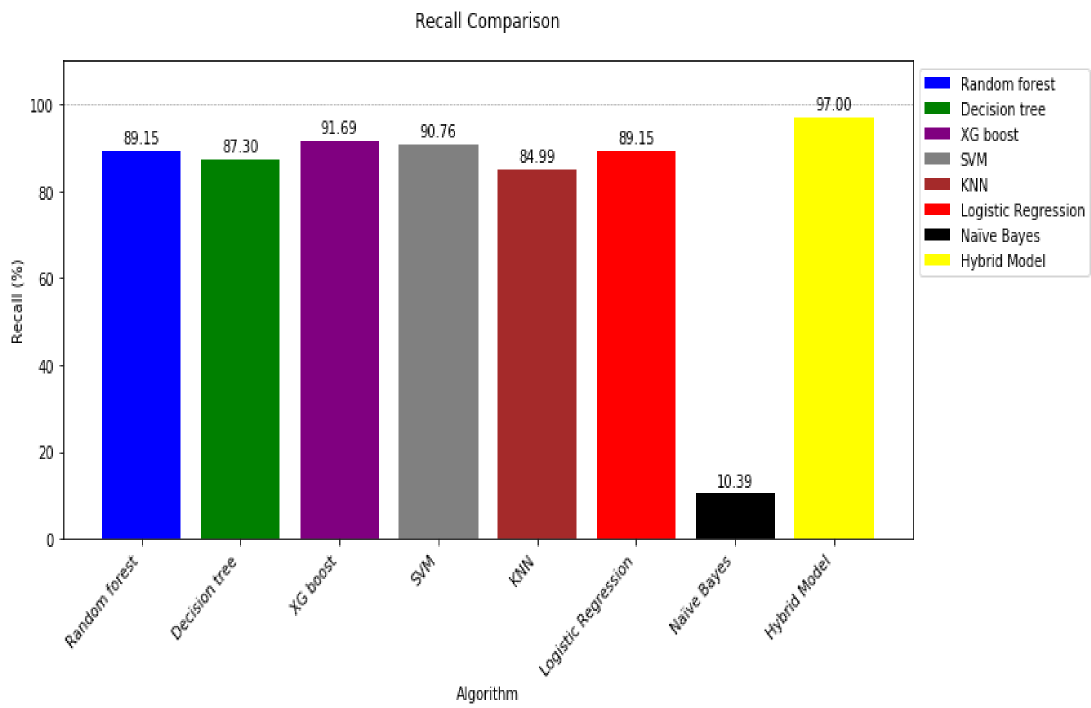

Figure 4: A bar chart showing the recall of Different ML Algorithms

From the figure above, the Hybrid Model leads in recall with a score of 97.00%, indicating its superior ability to

S/N	Algorithm	Accuracy	Precision	Recall	F1 Score
1	Random forest	79.94	82.83	89.15	85.87
2	Decision tree	82.94	87.70	87.30	87.50
3	XG boost	85.62	87.83	91.69	89.72
4	SVM	72.67	74.71	90.76	81.96
5	KNN	84.99	84.81	84.99	84.87
6	Logistic Regression	70.46	73.38	89.15	80.50
7	Naïve Bayes	38.70	94.00	10.39	18.83
8.	Hybrid Model	96.84	98.36	97.00	97.67

capture positive instances. Following closely are XG boost with 91.69% and SVM with 90.76%. Conversely, Naïve Bayes and Logistic Regression record the lowest recall scores among the listed algorithms.

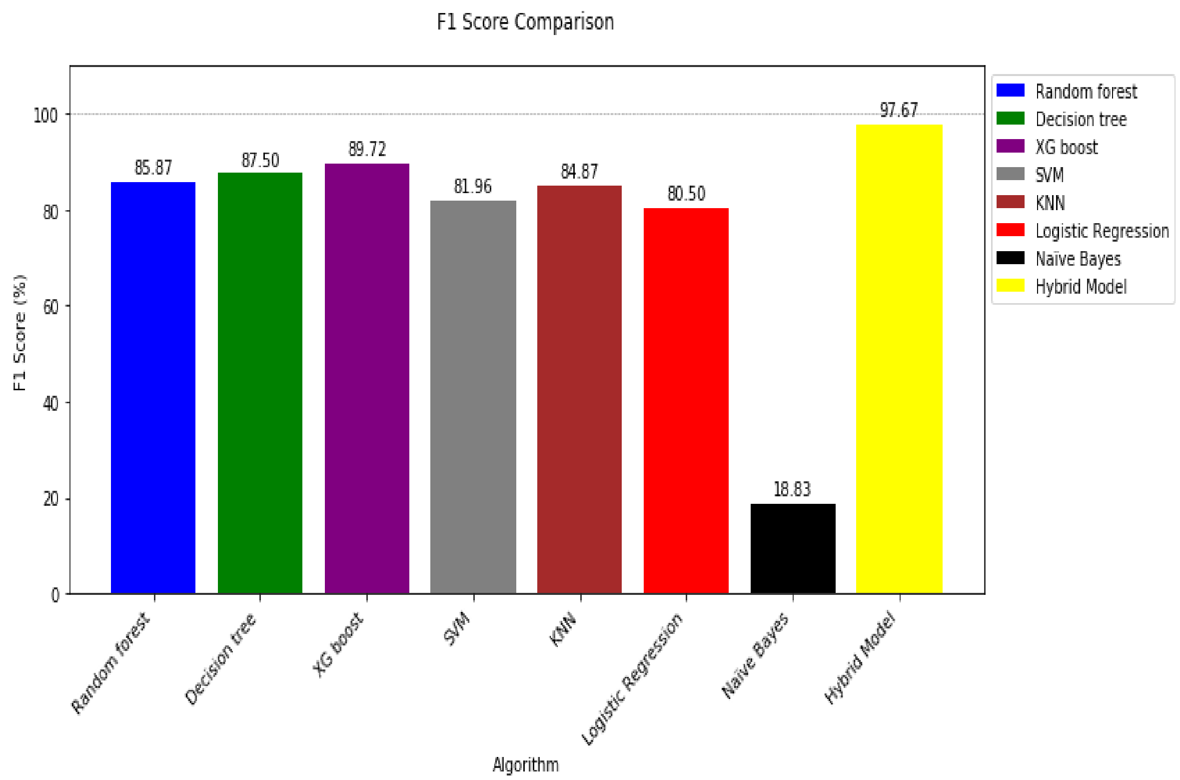

Figure 5: A bar chart showing the F1 score of Different ML Algorithms.

From the figure above, the Hybrid Model achieves the highest F1 score at 97.67%, showcasing its balanced performance in terms of precision and recall. XG boost and KNN follow with F1 scores of 89.72% and 84.87% respectively. However, Naïve Bayes and Logistic Regression demonstrate notably weaker F1 scores compared to the top performers. Comparism of the Hybrid model with other Machine Learning techniques using performance metrics is presented in Table 2.

Table 2: Summary of Performance Metrics of Various Machine Learning Algorithm

The confusion matrix for the algorithms utilized in predicting students academic performance based on admission patterns was also examined to gain insights into the false positive and false negative rates associated with each algorithm's prediction. The confusion matrix is delineated in terms of true positives, true negatives, false positives, and false negatives, offering a comprehensive evaluation of algorithm performance in predicting student outcomes. In the educational landscape, where admission patterns serve as crucial predictors of academic success, predictive models play a pivotal role in understanding and anticipating student trajectories. The confusion matrix serves as a cornerstone in this endeavor, providing a structured framework to assess the accuracy, precision, recall, and other essential metrics of predictive models. By elucidating the relationship between predicted and actual student performance, the confusion matrix offers valuable insights into the efficacy and reliability of predictive algorithms. This section aims to unravel the significance of the confusion matrix in the context of predicting student performance based on admission patterns, facilitating informed decision-making and fostering a data-driven approach to educational excellence.

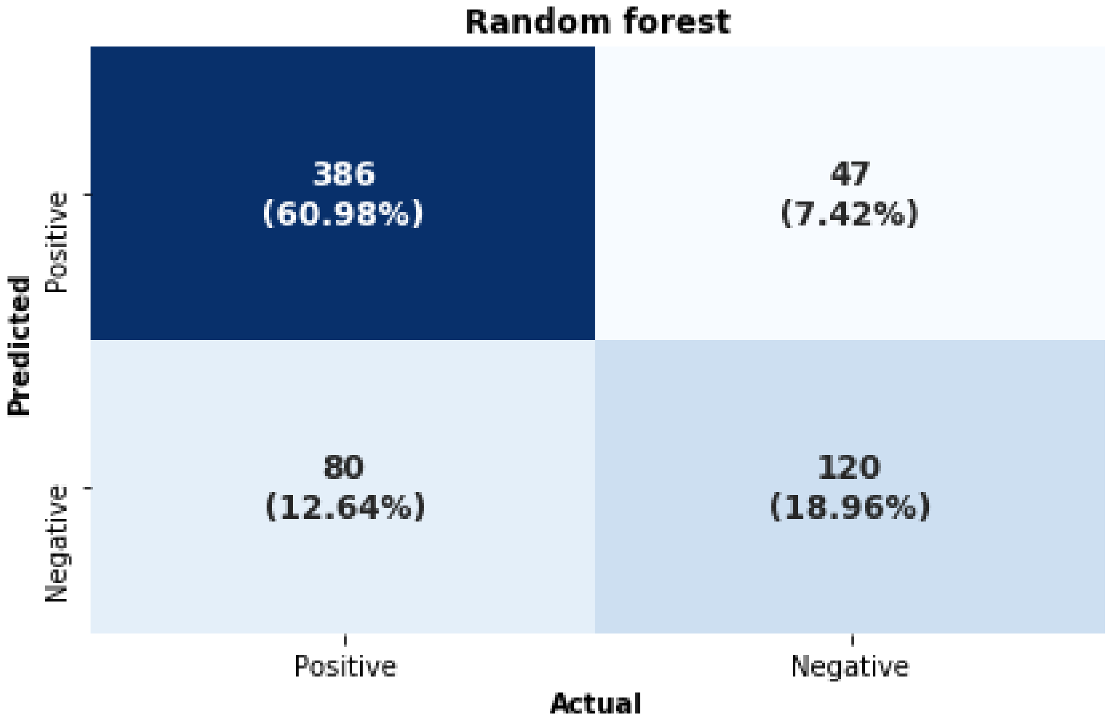

Figure 6. A Heatmap Showing the Confusion Matrix for Random Forest

From the heatmap, 386 students who chose their preferred course of study and were admitted performed well, while 120 students who were not admitted to their preferred course of choice did not perform well. Additionally, 80 students were not admitted to their preferred course but still performed well, and 47 students were admitted to study their course of choice but did not perform well.

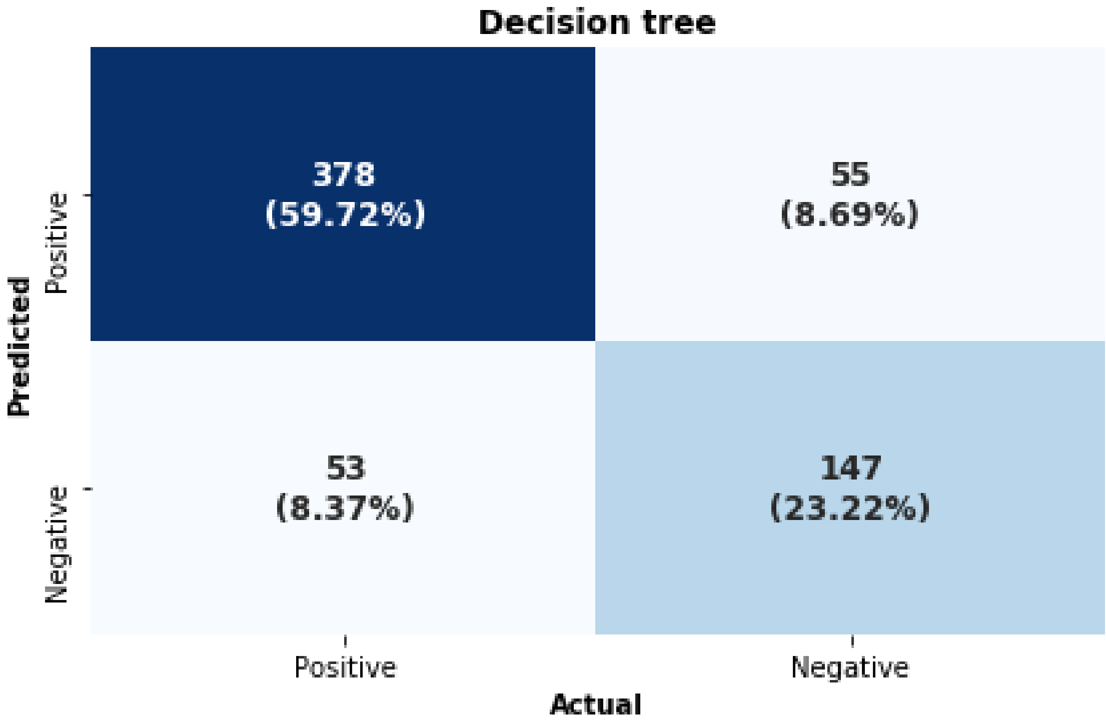

Figure 7. A Heatmap Showing the Confusion Matrix for Decision Tree

From the heatmap above, out of the students, 378 were admitted to their preferred course and performed well, whereas 147 were not admitted to study their course of choice and they did not perform well. Furthermore, 53 were not admitted to their preferred course but still performed well, while 55 were admitted to their course of choice but did not perform well.

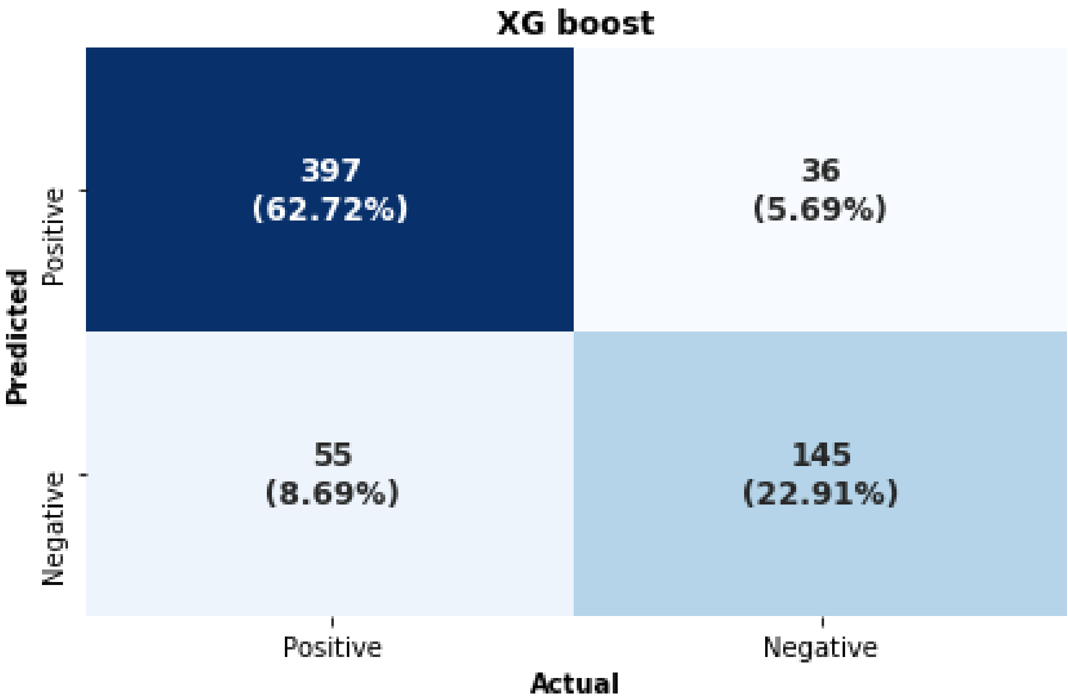

Figure 8. A Heatmap Showing the Confusion Matrix for XG Boost

As indicated in the heatmap above, 397 students were admitted to their preferred course and performed well, with 145 not admitted to admitted to their course of choice and not performing well. Additionally, 55 were not admitted to their preferred course but still performed well, while 36 were admitted to their preferred course but did not perform well.

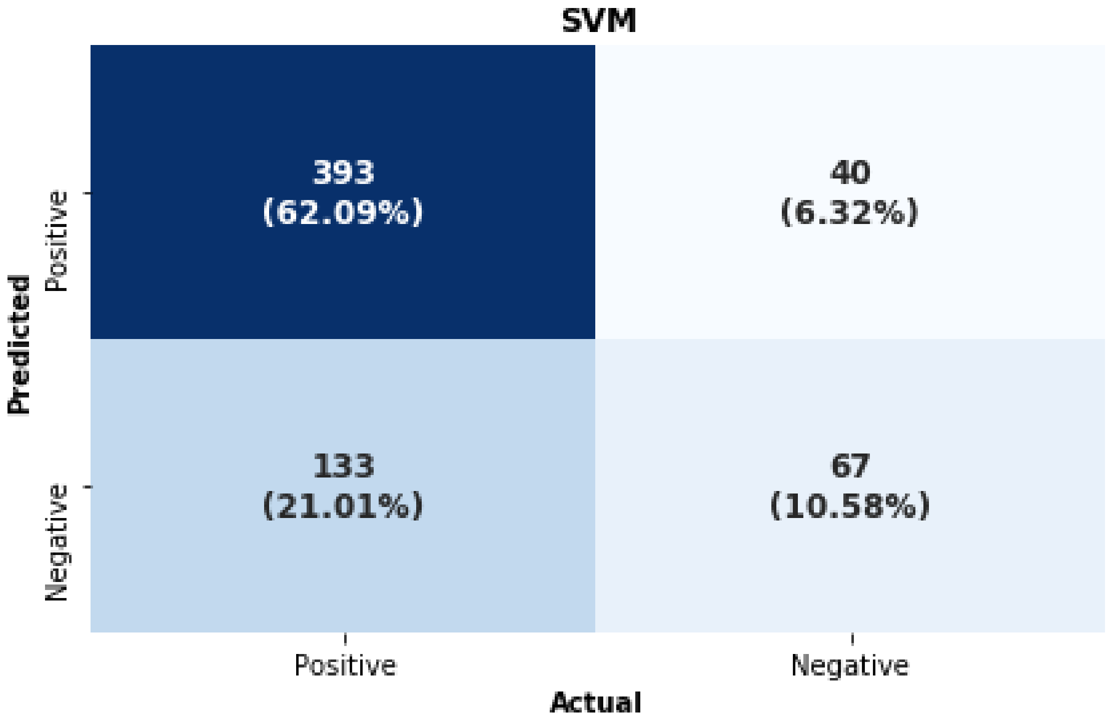

Figure 9. A Heatmap Showing the Confusion Matrix for SVM

The heatmap above shows that 393 students were admitted to their preferred course and performed well, whereas 67 were not admitted to their course of choice and did not perform well. Moreso, 133 were not admitted to their preferred course but still performed well, and 40 were admitted to their preferred course but did not perform well.

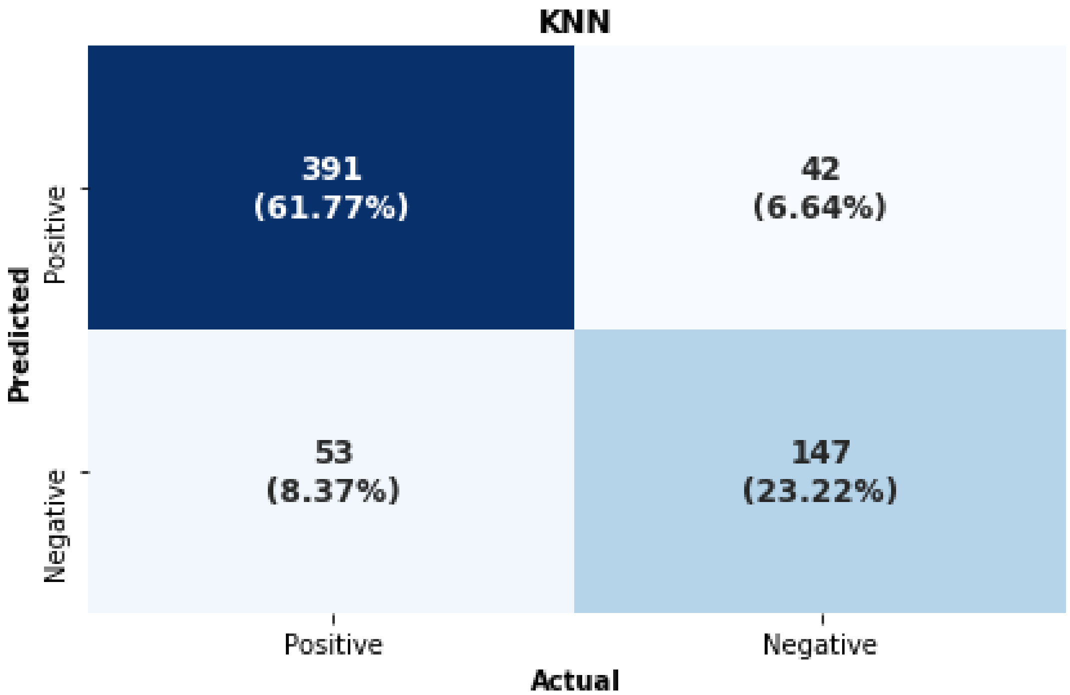

Figure 10. A Heatmap Showing the Confusion Matrix for KNN

According to the heatmap above, 391 students were admitted to their preferred course and performed well, with 147 did not get admitted to study their course of choice and so, did not perform well. Additionally, 53 were not admitted to their preferred course but still performed well, and 42 were admitted to their preferred course but did not perform well.

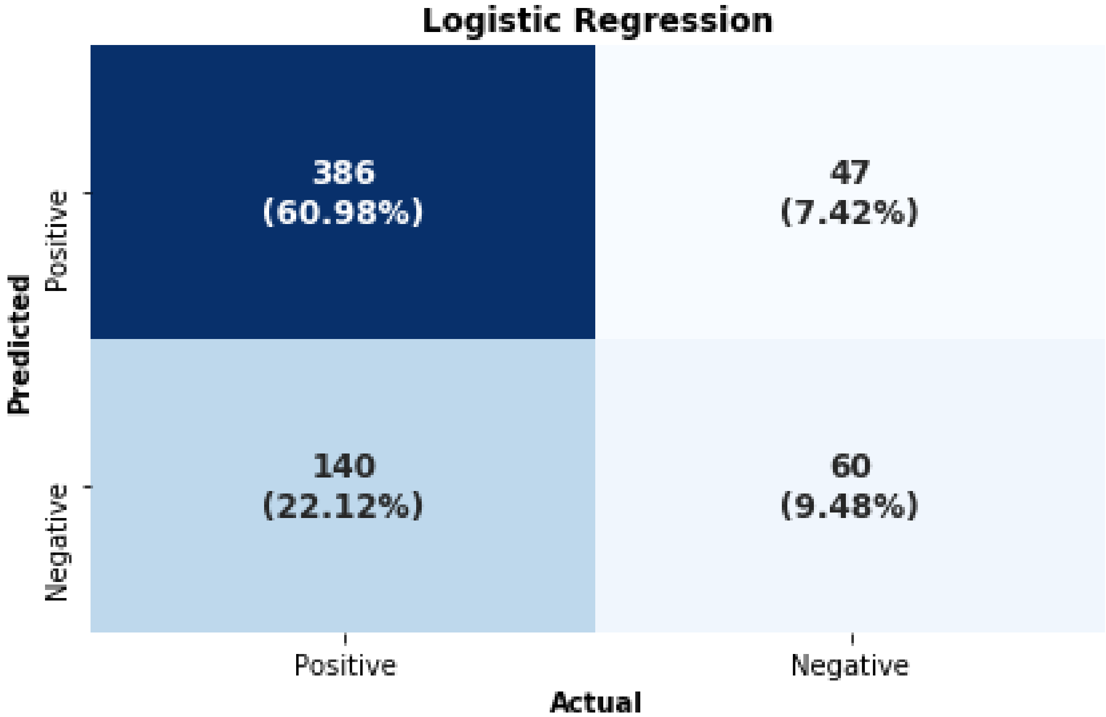

Figure 11. A Heatmap Showing the Confusion Matrix for Logistic Regression

The heatmap above indicates that among the students, 386 were admitted to their preferred course and performed well, while 60 were not admitted to their preferred course and did not perform well. Furthermore, 140 were not admitted to their preferred course but still performed well, while 47 were admitted to study their course of choice but did not perform well.

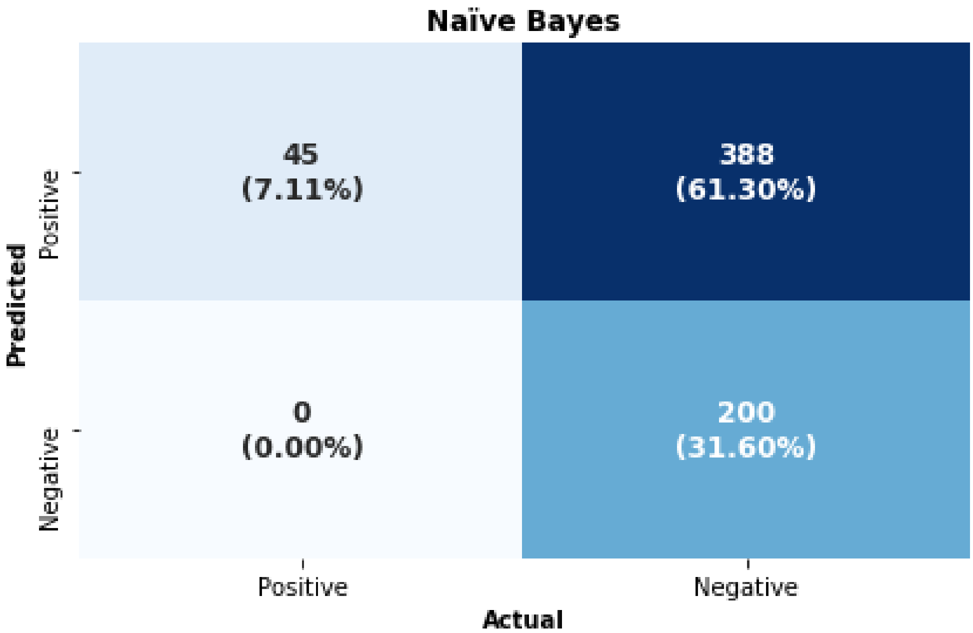

Figure 12. A Heatmap Showing the Confusion Matrix for Naïve Bayes

The heatmap illustrates that 45 students successfully gained admission to their preferred courses and demonstrated commendable performance, whereas 200 were not admitted to study their preferred course and so did not perform well. Additionally, no students were falsely admitted, although 388 were admitted to study their course of choice but did not perform well.

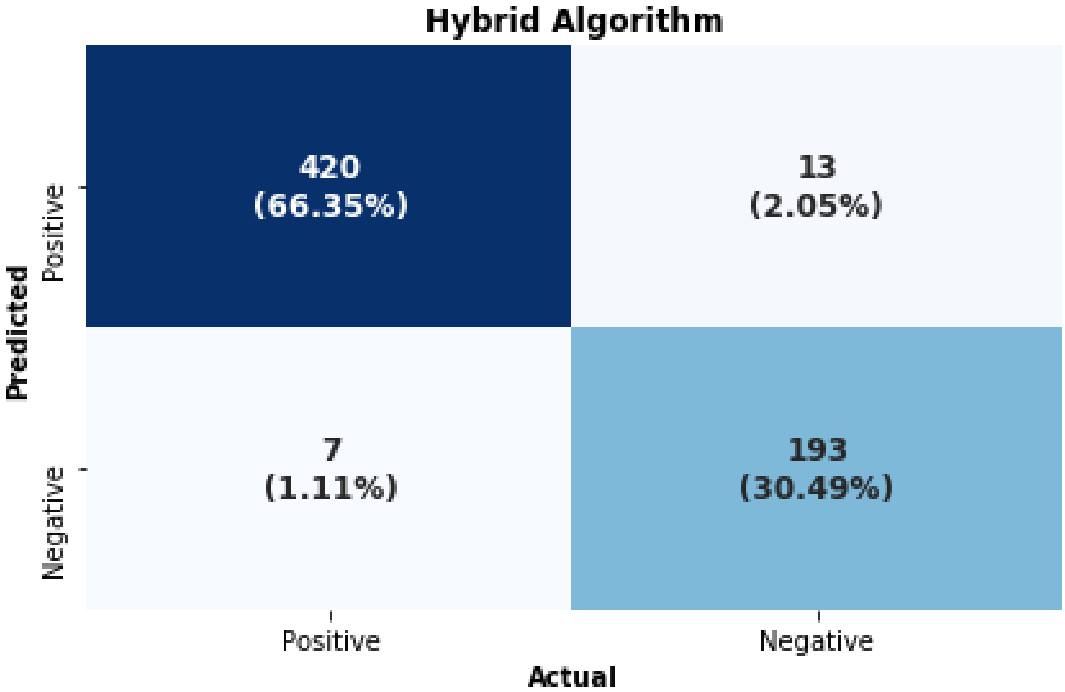

Figure 13. A Heatmap Showing the Confusion Matrix for Hybrid Model

The heatmap depicted above reveals that 420 students admitted to their preferred courses performed well. In contrast, 193 students who were not admitted to study their preferred courses did not perform satisfactorily. Moreso, 7 students were not admitted to their desired courses but still achieved commendable performance, while 13 students were admitted to study their preferred course but did not meet performance expectations.Comparism of the Hybrid model with other ML techniques using confusion matrix is presented in Table 3.

Table 3: Summary of Confusion Matrix of Various Machine Learning Algorithm

S/N	Algorithm	TP	FN	FP	TN
1	Random forest	386	47	80	120
2	Decision tree	378	55	53	147
3	XG boost	397	36	55	145
4	SVM	393	40	133	67
5	KNN	391	42	53	147
6	Logistic Regression	386	47	140	60
7	Naïve Bayes	45	388	0	200
8.	Hybrid Algorithm	420	13	7	193

Additionally, the Area Under the ROC Curve (AUC) was utilized to assess the overall prediction accuracy of each algorithm. This section presents the Receiver Operating Characteristic (ROC) curves along with their corresponding Area Under the Curve (AUC) values. ROC curves are widely used in binary classification tasks to visualize the performance of a model across different decision thresholds. The AUC metric provides a single value summarizing the performance of the classifier, with higher AUC scores indicating better discrimination between positive and negative classes. By examining ROC curves and AUC values, we gain insights into the predictive power and overall effectiveness of the classification models being evaluated.

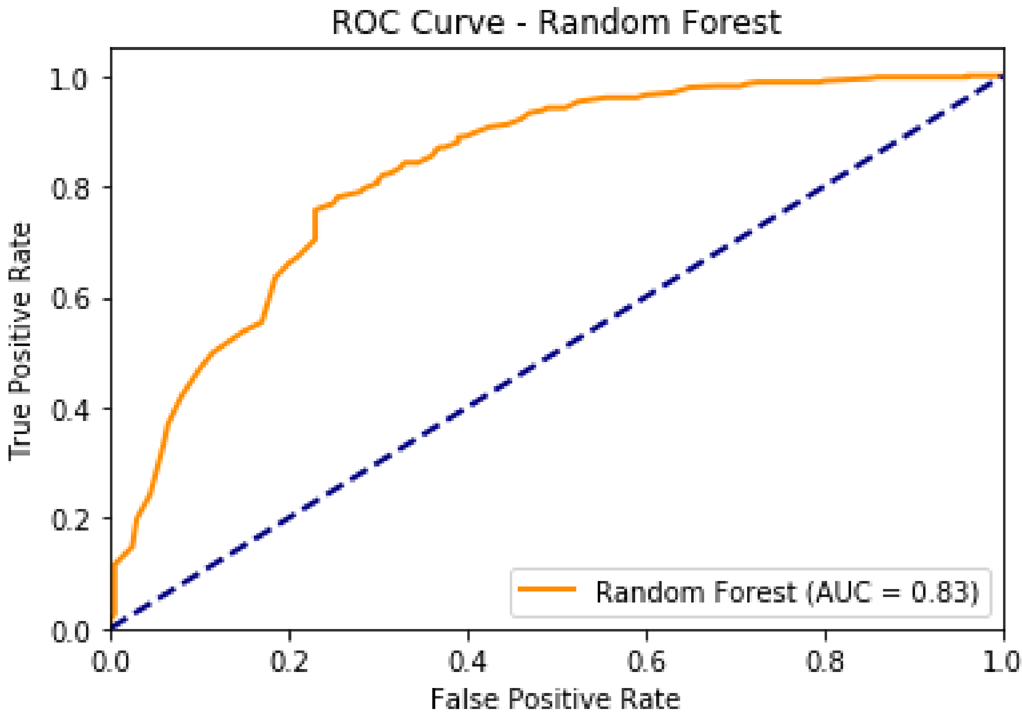

Figure 18. A Graph Showing ROC Curve for Random Forest

From the figure above, Random Forest exhibits strong predictive capability with an AUC of 0.83, indicating its effectiveness in forecasting student performance based on admission patterns. This suggests its utility in guiding institutions to make informed admissions decisions and optimizing student success rates.

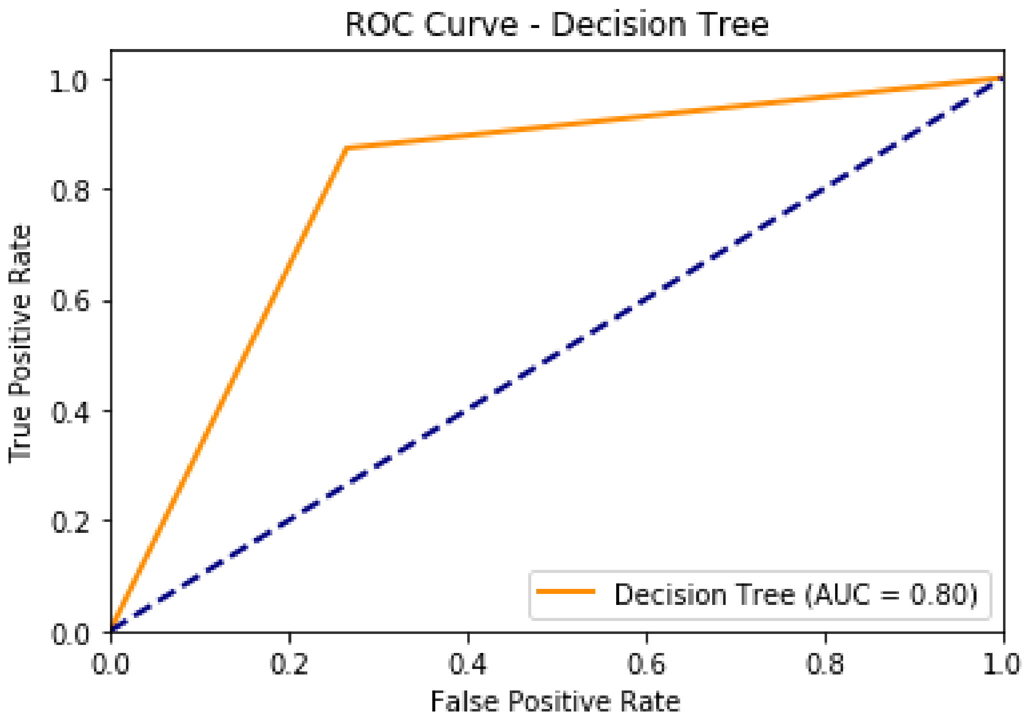

Figure 19. A Graph Showing ROC Curve for Decision Tree

The figure above shows that Decision Tree achieves a moderate AUC of 0.80, signifying its ability to assess student performance from admission patterns. This implies its potential in assisting institutions to refine admission criteria for better student outcomes.

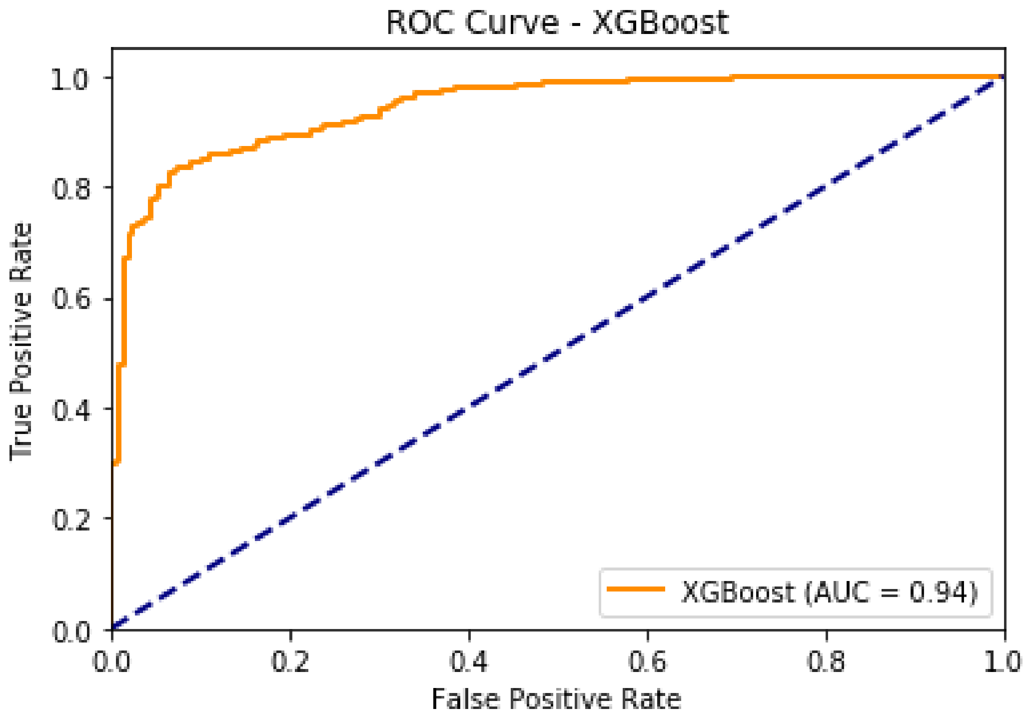

Figure 20. A Graph Showing ROC Curve for XG Boost

XG Boost, as depicted in the figure above, boasts an impressive AUC of 0.94, showcasing its exceptional predictive accuracy in analyzing admission patterns to predict student performance. This underscores its pivotal role in aiding institutions to optimize admissions strategies for fostering student success.

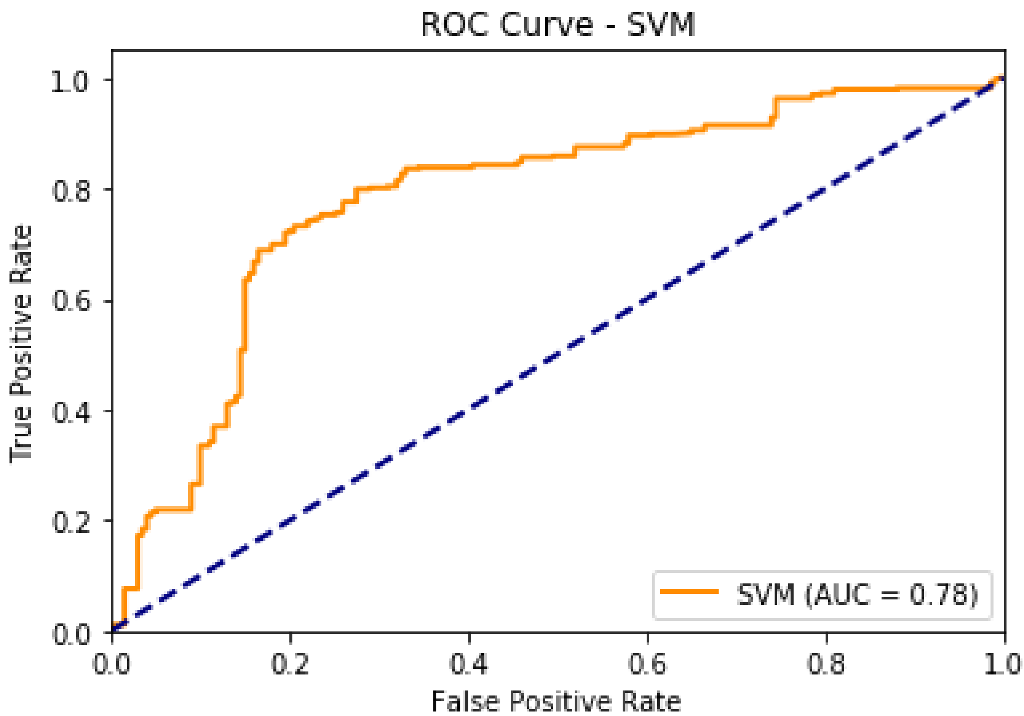

Figure 21. A Graph Showing ROC Curve for SVM

From the figure above, SVM demonstrates moderate predictive capability with an AUC of 0.78, indicating its usefulness in assessing student performance based on admission patterns. This suggests its potential to guide institutions in making data-informed decisions to enhance student outcomes.

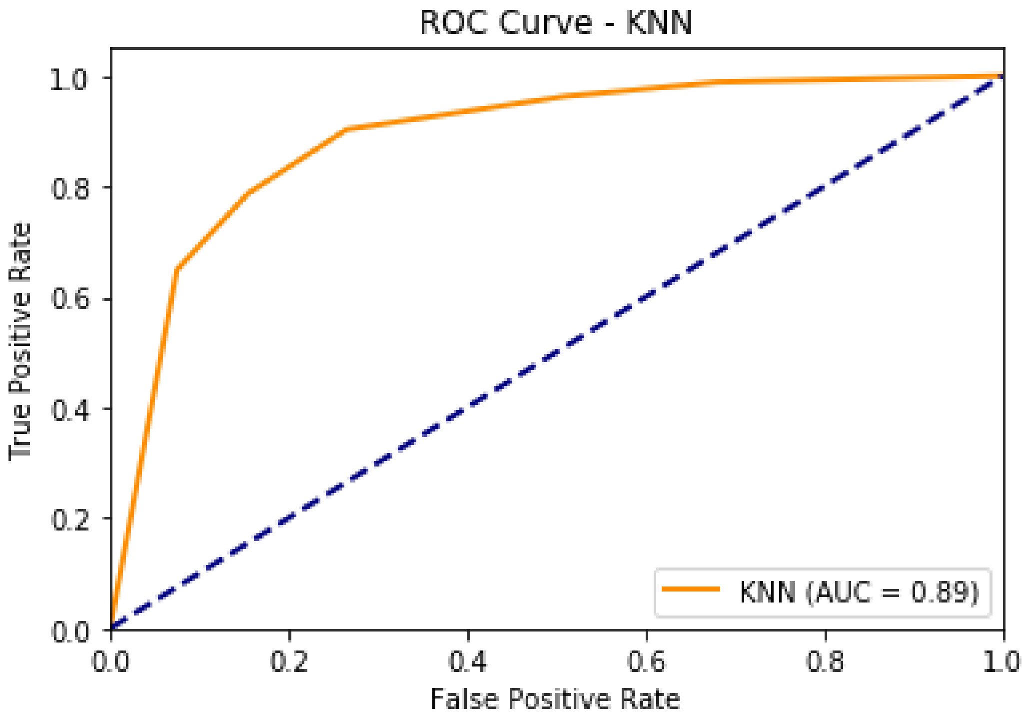

Figure 22. A Graph Showing ROC Curve for KNN

The figure above illustrates KNN's strong predictive accuracy with an AUC of 0.89, highlighting its effectiveness in forecasting student performance from admission patterns. This implies its significance in providing valuable guidance for admissions optimization.

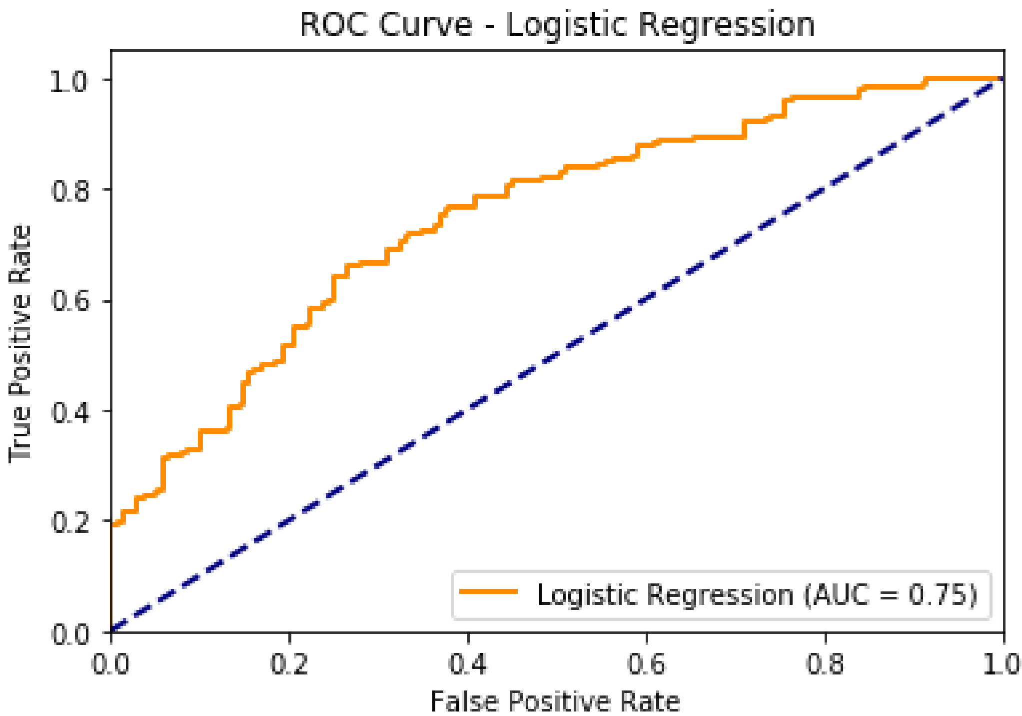

Figure 23. A Graph Showing ROC Curve for Logistic Regression

From the figure above, Logistic Regression displays moderate predictive capability with an AUC of 0.75, suggesting its utility in assessing student performance based on admission patterns. This indicates its potential to aid institutions in making informed decisions regarding admissions to improve overall student outcomes.

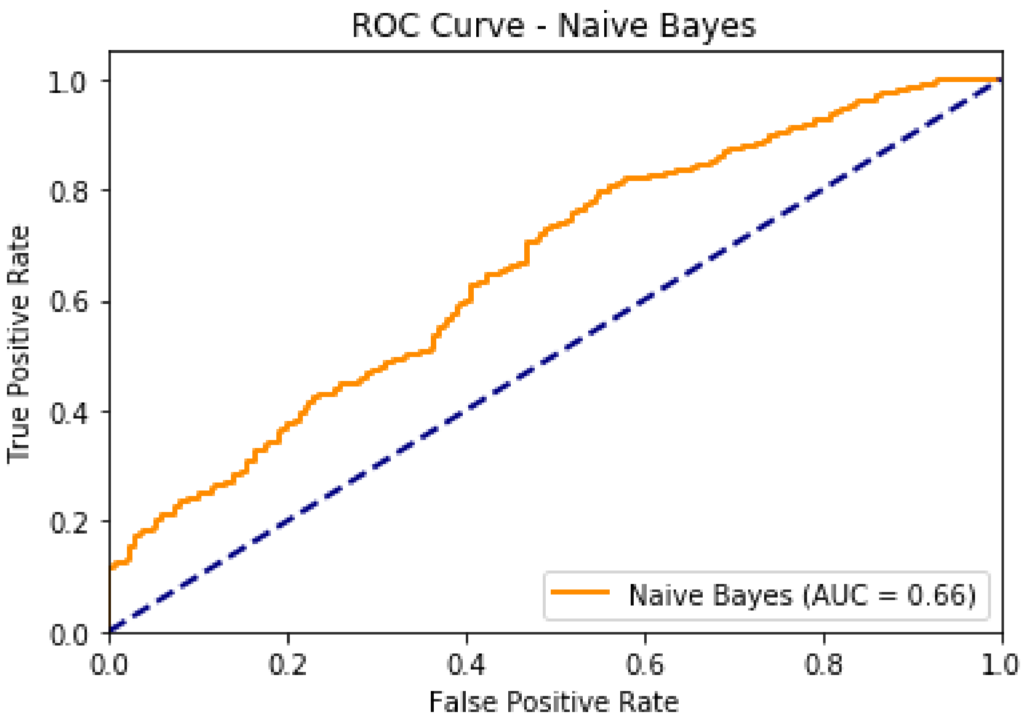

Figure 24. A Graph Showing ROC Curve for Naïve Bayes

As depicted in the figure above, Naïve Bayes demonstrates modest predictive accuracy with an AUC of 0.66, suggesting limited effectiveness in forecasting student performance from admission patterns. Despite this, it still offers insights that can aid institutions in refining admissions processes to better align with student success metrics.

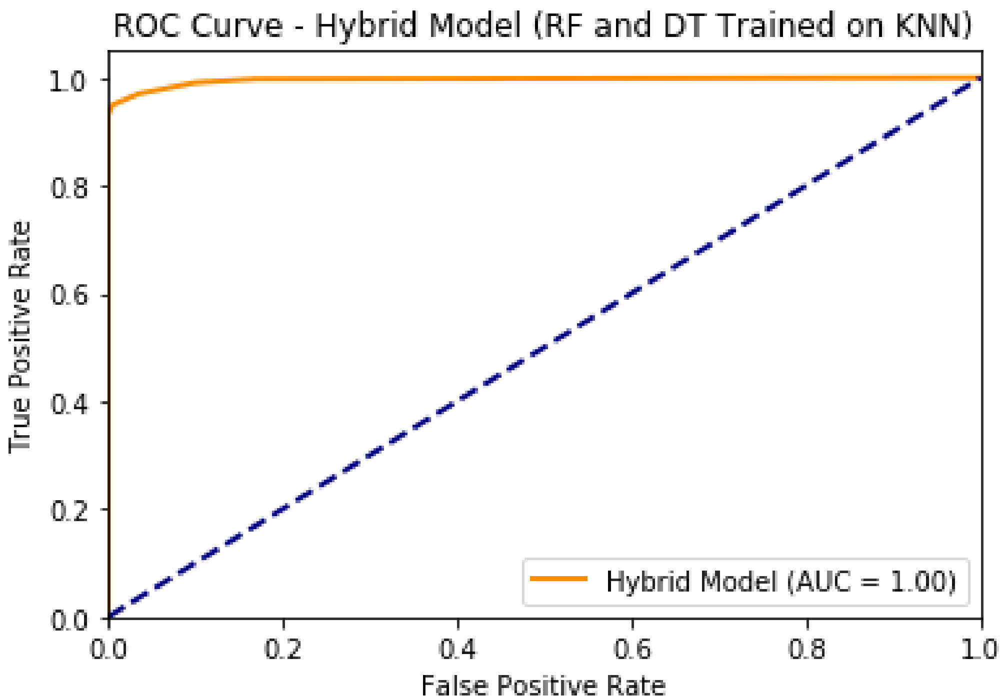

Figure 25. A Graph Showing ROC Curve for Hybrid Model

From the figure above, the Hybrid Algorithm showcases exceptional predictive accuracy with a perfect AUC of 1.00, indicating its unparalleled effectiveness in analyzing admission patterns to predict student performance. This underscores its pivotal role in optimizing admissions strategies for maximum student success.

V. Conclusion

In conclusion this study demonstrates the effectiveness of machine learning algorithms in predicting student performance based on admission patterns. Among the algorithms evaluated, the Hybrid Algorithm emerges as the top performer, exhibiting exceptional accuracy, precision, recall, and AUC. Its robustness in minimizing misclassifications and accurately identifying both high-performing and at-risk students makes it a highly reliable solution for educational predictive modeling tasks. While other algorithms show promise in certain aspects, the Hybrid Algorithm consistently outperforms them across various metrics, highlighting its superiority in predicting student outcomes.

References

[1]. Aman, F., Rauf, A., Ali, R., Iqbal, F., & Khattak, A. M. (2019, July). A predictive model for predicting students academic performance. In 2019 10th International conference on information, intelligence, systems and applications (IISA) (pp. 1-4). IEEE.

[2]. Baek, C., & Doleck, T. (2022). Educational data mining: A bibliometric analysis of

[3]. an emerging field. IEEE Access, 10, 31289-31296.

[4]. Cruz, M. E. L. T., & Encarnacion, R. E. (2021). Analysis and prediction of students'

[5]. academic performance and employability using data mining techniques: A research travelogue. The Eurasia Proceedings of Science Technology Engineering and Mathematics, 16, 117-131.

[6]. Gebru, H. T., & Verstegen, D. (2023). Assessing predictors of students' academic performance in Ethiopian new medical schools: a concurrent mixed-method study. BMC Medical Education, 23(1), 448.

[7]. Kaviyarasi, R., & Balasubramanian, T. (2020). Predictive analysis of academic performance of college students using ensemble stacking. Kongunadu Research Journal, 7(2), 94-98.

[8]. Kukkar, A., Mohana, R., Sharma, A., & Nayyar, A. (2023). Prediction of student academic performance based on their emotional wellbeing and interaction on various e-learning platforms. Education and Information Technologies, 28(8), 9655-9684.

[9]. Namoun, A., & Alshanqiti, A. (2020). Predicting student performance using data mining and learning analytics techniques: A systematic literature review. Applied Sciences, 11(1), 237.

[10]. Razaque, A., & Alajlan, A. M. (2020). Supervised machine learning model-based approach for performance prediction of students. Journal of Computer Science, 16(**8**), 1150-1162.

[11]. Vanerio, J., & Casas, P. (2017). Ensemble-learning approaches for network security and anomaly detection. In Proceedings of the workshop on big data analytics and machine learning for data communication networks (pp. 1-6).

[12]. Walia, N., Kumar, M., Nayar, N., & Mehta, G. (2020,). Student's academic performance prediction in academic using data mining techniques. In Proceedings of the International Conference on Innovative Computing & Communications (ICICC).

[13]. Zhang, Y., Yun, Y., An, R., Cui, J., Dai, H., & Shang, X. (2021). Educational data mining techniques for student performance prediction: method review and comparison analysis. Frontiers in psychology, 12, 698490.