Enhancing Student Performance and Outcome Prediction Using Hybrid Machine Learning Models

Supervised By

Dr. Marzia Ahmed

Assistant Professor

Department of Software Engineering

Daffodil International University

Submitted By

Akib Hasan

Id: 221-35-894

Department of Software Engineering

Daffodil International University

This thesis report has been submitted in fulfilment of the requirements for the Degree of Bachelor of Science in Software Engineering.

© All right Reserved by Daffodil International University

APPROVAL

APPROVAL

This thesis titled on Enhancing Student Performance and Outcome Prediction Using Hybrid Machine Learning Models, submitted by **Akib Hasan ID: 221-35-894** to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering and approval as to its style and contents.

BOARD OF EXAMINERS

Dr. Fazla Ealhe
Assistant Professor & Associate Head
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Chairman

Dr. Marzia Ahmed
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Internal Examiner 1

Dr. Shabnom Mustary
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Internal Examiner 2

Md. Rajib Mia
Lecturer (Senior Scale)
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Internal Examiner 3

Mohammod Abul Kashem, PhD
Professor
Department of Computer Science and Engineering
DUET, Bangladesh

External Examiner

Enhancing Student Performance and Outcome Prediction Using Hybrid Machine Learning Models

Akib Hasan

Id: 221-35-894

Bachelor of Science

DAFFODIL INTERNATIONAL UNIVERSITY

©Daffodil International University

SUPERVISOR'S DECLARATION

SUPERVISOR'S DECLARATION

I hereby declare that I have reviewed this thesis entitled " **Enhancing Student Performance and Outcome Prediction Using Hybrid Machine Learning Models**", and in my opinion, it is adequate in terms of scope and quality for the award of the degree of Bachelor of Science in Software Engineering.

(Supervisor's Signature)

Full Name : Dr. Marzia Ahmed

Position : Assistant Professor

Date : 23 December 2025

STUDENT'S DECLARATIONS

STUDENT'S DECLARATION

I confirm that the piece in this thesis is based on my own writing with the exception of quotation and reference that have been discussed. I also confirm that it was not previously and concurrently registered at Daffodil International University or other institutions at any other degree.

Akib Hasan

(Student's Signature)

Full Name : Akib Hasan

ID Number : 221-35-894

Date : 27 November 2025

Enhancing Student Performance and Outcome Prediction Using Hybrid Machine Learning Models

Akib Hasan

Id: 221-35-894

Thesis submitted in fulfilment of the requirements
for the award of the degree of
Bachelor of Science

Department of Software Engineering

DAFFODIL INTERNATIONAL UNIVERSITY

DECEMBER 2025

©Daffodil International University

ACKNOWLEDGEMENTS

I would like to extend my sincere thanks and appreciation to my supervisor, Dr. Marzia Ahmed, Assistant Professor for her supervision, constructive feedback and constant support at every stage of the research process. Her knowledge, her patience, and her supportive supervision have been instrumental in this thesis taking its final form and adding value to my work. I am very grateful for her time, commitment and valuable suggestions, which had helped me to improve my ideas and encouraged academic trustworthiness. Without her stupendous support, I would not have been to finalize this research as well, I would like to say thank you, my teachers, classmates and everyone who stood by me and helped make this path paved or shown. They have inspired and cooperated much to the completion of this study. Lastly, I would like to thank my family for their unwavering support, understanding, and persistent encouragement throughout the course of my studies.

©Daffodil International University

DEDICATION

To my parents This thesis is dedicated, for their encouragement, support 1 and belief in me, which I know they will always offer to me. It is also my honor to dedicate this effort to the teachers and mentors, whose enlightenment and wisdom still light up my academic pursuit Lastly, I would like to dedicate this thesis to all the people who had faith in me and encouraged me while I work towards my goals with strength and honor.

ABSTRACT

Predicting student academic outcomes has emerged as a critical feature of contemporary educational data mining to help institutions identify at-risk students, and increase student retention, and promote their academic performance. In this work, we introduce a new hybrid stacking ensemble model (OutcomeHyX) to predict three major student-resultant labels viz., Dropout, Enrolled and Graduate using a dataset that is rich in demographic, socioeconomic and academic performance features. The work also considers data quality through preprocessing, feature engineering and SMOTE-based balancing to enhance classification fairness of the proposed approach. Several baseline machine learning models such as KNN, SVM, Random Forest and XGBoost were compared to determine their performance baselines. Empirical study shows that the performance of traditional models is mediocre and there exist large differences on prediction accuracy across outcome classes, especially in minority Enrolled class. The proposed OutcomeHyX model, comprising Support Vector Machine and Random Forest as base learners, while using Logistic Regression as meta-learner outperforms other models with a Test accuracy of 87.46% along with remarkable class-wise F1-scores. With the ROC-AUC>0.95, its great discrimination capability is also substantiated. The results show that the hybrid stacking scheme dominates all standalone classifiers in exploiting non-linear and intricate structures in academic dataset. This study presents a reliable predictive model that universities can use as tool for early warning, progress monitoring and data-informed decision-making. The model also provides a scalable recipe for next-generation studies aimed at improving student success and retention.

Keywords: Student Performance Prediction, Academic Outcome Classification, Educational Data Mining (EDM), Machine Learning, Hybrid Stacking Model, OutcomeHyX, Support Vector Machine (SVM), Random Forest, XGBoost, SMOTE, Ensemble Learning, Predictive Analytics, Dropout Prediction, Model Evaluation Metrics.

©Daffodil International University

TABLE OF CONTENTS

©Daffodil International University

LIST OF FIGURES

©Daffodil International University

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Full Form
EDM	Educational Data Mining
ML	Machine Learning
SVM	Support Vector Machine
RF	Random Forest
KNN	K-Nearest Neighbors
XGBoost	Extreme Gradient Boosting
SMOTE	Synthetic Minority Oversampling Technique
AUC	Area Under the Curve
ROC	Receiver Operating Characteristic
LR	Logistic Regression
DT	Decision Tree
NB	Naive Bayes
EDA	Exploratory Data Analysis
GDP	Gross Domestic Product
RBF	Radial Basis Function
TPR	True Positive Rate
FPR	False Positive Rate
HEI	Higher Education Institution

©Daffodil International University

CHAPTER 1

INTRODUCTION

1.1 Introduction

Prediction of student academic performance and outcome have become a vital aspect in educational data mining (EDM) and learning analytics. As the number of universities grows rapidly and learners become more diverse, it becomes even more difficult for universities to recognize who among their students are at risk of failing academically or dropping out. Manual academic monitoring systems are still in use and are insufficient, slow, and get loss of some complicated interrelated values from the multi-dimensional student data. ML is increasingly recognized as a powerful and successful approach to modelling complex educational data and producing accurate predictions. Historical academic performances, demographics, SES and behavioral profiles can inform ML models to help develop timely interventions and standardize the existing student retention policies. However, in spite of the expanded popularity of ML within academic analytics, it is still difficult to obtain high and stable predictive accuracy. Different ML algorithms have different behavior across data conditions. For example, tree-based methods such as Random Forest deal well with non-linearity but do not always generalize perfectly; SVM operate efficiently in the presence of complex boundaries, but struggle in high dimensions; Logistic Regression provides interpretability, though it usually struggles to uncover deeper patterns. To mitigate these shortcomings the hybrid and ensemble models are popular because of their ability in integrating different algorithms. Having identified this opportunity, in this thesis we introduce a new hybrid ensemble model called OutcomeHyX that combines the predictive power of Support Vector Machine, Random Forest and Logistic Regression in a stacking framework. The objectives of the model are to increase accuracy in prediction of student performance–based academic outcome and to perform better than traditional models.

1.2 Motivation

We are motivated to conduct this research due to the increasing demand for effective and early prediction systems in academic institutions. It’s often hard to spot students who are at risk of dropping out, falling behind or in need of special academic support.

©Daffodil International University

An early warning prediction system that can help the administrators make a difference to intervene early, and have better student retention rate, better academic success rate, and better performance. While studies already show encouraging results when utilizing machine learning for predicting academic outcomes, no one algorithm achieves the best performance in all categories (Dropout, Enrolled, Graduate). This contrary pattern of performance underscores the importance of hybrid models which are capable to combine the advantages from diversity learning algorithms for improved prediction accuracy. In addition, educational data are generally imbalanced, highly heterogeneous and determined by a combination of academic, demographic, economic and personal factors. To this end, a good model should be able to capture such complexities. OutcomeHyX In this work, we propose and investigate OutcomeHyX – a novel hybrid stacking ensemble that aims to build 2star more accurate and reliable models than current methods with practical significance for making decisions by institutions.

1.3 Background Study

Educational Data Mining (EDM) is an important discipline of data science and related to the analysis of students' records in order to support learning, decisions, planning and academic interventions [1]. The rapid expansion of digital learners' platforms, institutional databases and student information systems have led to a wealth of data that holds potential for informing learning behavior, predicting academic performance and preventing drop-out. With the intense pressure on colleges and universities to increase persistence and success, predictive modeling is an essential instrument in today's dynamic academic management Factors that can affect student performance and academic outcomes socio-economic status or SES (i.e., parent's educational attainment level, occupation of parents), GPA in previous institution entered (; usually high school) institutional involvement [college/university/apprenticeship program attended] attendance patterns personality traits/distress/psychological well-being/personal circumstances. Conventional statistical methods like Linear Regression, Logistic Regression or classic descriptive analysis can be non-optimal in modeling such intricate and non-linear connections. They rely on linearity and independence among the variables, which restrains their ability to predict in complex education data. In contrast-based approaches are more general and offer stronger tools for mining hidden patterns in heterogeneous data.

©Daffodil International University

The methods such as DT, RF, SVM, Naive Bayes and XGBoost have exhibit potential to predicting student Dropout and performance assessment and outcome classification. Due to their capacity of processing multi-dimensional data, capturing non-linear patterns, and adapting to various educational settings, they have been proved to reach higher predictive accuracy in many research works. Yet, a big obstacle to model accuracy with respect to various datasets and prediction categories has been reported in the literature. There is no one machine learning technique that generally outperforms others for all academic outcome purposes. For example, Random Forest models may generalize well, but are biased in cases of data imbalance. It works well with complex boundaries, but can be time-consuming and depend on parameter settings. Meanwhile, logistic regression is interpretable but unable to capture deeper non-linear relationships. In order to overcome these limitations, the hybrid and ensemble approaches have been widely applied. Combining the strengths of multiple base learners, ensemble learning--in particular techniques like bagging, boosting and stacking--have demonstrated to provide increased robustness and prediction accuracy. Stacking trains various models on different levels, combining them and applying the optimization using meta-learner at a new level (Wolpert 1992). The method has been shown to outperform single-model techniques in a number of domains such as health care, finance and education. Despite these advances, the use of hybrid stacking ensembles for academic performance prediction is not common in research literature. Only a handful of studies employ model such as SVM and Random Forest with meta classifier in an objected hybrid setup.

1.4 Problem Statement

Academic success can be difficult to predict for students due to the interplay of demographic, socioeconomic, behavioral and academic variables. In previous studies, we have showed that those machine learning models like LR, DT, RF, SVM, XGBoost have different levels of performance. Nevertheless, none of these models can achieve high accuracy for all prediction categories and they tend to work poorly on imbalanced datasets, and non-linear behavior. Besides, a number of traditional methods do not model complex relationships between the student attributes well and do not effectively take advantage of complementary strengths among multiple algorithms. This creates a critical gap there is a demand for a powerful hybrid predictive model that can outperform the single machine learning classifiers in the prediction of student performance-based outcome.

©Daffodil International University

1.5 Research Objectives

The primary goal of this dissertation is: We aim to propose and analyses the prediction performance of a new hybrid stacking ensemble model, OutcomeHyX, for improved student performance-based academic outcome forecasting. For that purpose, the following specific objectives are formulated the analysis and preprocessing of student academic records for predictive modeling purposes.

- ✓ To access the effectiveness of traditional machine learning models, including Logistic Regression, KNN, Naive Bayes, Decision Tree., Random Forest, SVM, XGBoost and AdaBoost.
- ✓ To develop the OutcomeHyX, a hybrid stacking ensemble model using SVM and Random Forest as base learners with Logistic Regression as meta-learner.
- ✓ To include feature engineering and to utilize data balancing methods (like SMOTE, helper-feature design) for an improved model performance.
- ✓ To evaluate the performance of OutcomeHyX against other state-of-the-art baseline models based on accuracy, precision, recall, F1-score and ROC-AUC.
- ✓ To show that OutcomeHyX performs better than conventional models for the prediction of academic performance of students.

1.6 Purpose of this Research

The goal of this research is to provide a robust likelihood model that will offer very precise predictions about the academic status whether a student will drop out, continue attendance or graduate. However, conventional ML algorithms can usually not achieve sufficiently high accuracy on various educational datasets because of non-linearity and class imbalance, as well as the complex interactions among student characteristics. To address such limitations, in this study, we have developed a hybrid stacking ensemble model called OutcomeHyX, which composes of the versatility of different machine learning methods to enhance prediction performance. The goal is to even increase accuracy while offering a more robust and generalizable solution than single-model or ensemble approach currently available.

©Daffodil International University

CHAPTER 2

LITERATURE REVIEW

2.1 Overview

It is the aim of this review to condense and critically analyze previous work in the field regarding prediction of students' academic success, dropout analysis, and outcome forecasting through machine learning. Educational data mining (EDM) has advanced considerably in the last decade, with many researchers using machine learning approaches to model and predict student courses of action, performance, and dropout. In this chapter, representatives of these categories of approaches traditional statistical models, classical machine learning algorithms, deep learning architectures and hybrid ensemble systems—are investigated to investigate the capabilities and limitations for outcome prediction in medical applications. The review emphasizes that single methods (e.g. Logistic Regression, Decision Trees, Support Vector Machines, Random Forests and boosting) have yielded encouraging results but with varying performance across datasets and learning conditions. Moreover, problems like imbalanced data distribution, nonlinear and complex relation between features, and heterogeneous student characteristics make the single-model methods less effective. To overcome these weaknesses, a family of hybrid and ensemble learning methods including bagging, boosting, voting and stacking are developed by researchers (multi-algorithmic) with the aim of enhancing the predictive accuracy and robustness.

2.2 Previous Study

Krüger et al. (2023) Introduce an interpretable machine learning model for student dropout prediction, applied to Brazilian schools and engineered temporal grade characteristics. They use ensemble model technique and interpretability reflects feature importance and explanation methods, such as SHAP-style analysis. They achieve high recognition rates (90% of the flagged students indeed dropped out) and put a focus on explainable predictions for ethical decision making in education. This paper is related to your thesis as it demonstrates how to include interpretable ML in state-of-the-art models for dropout prediction [1].

©Daffodil International University

Yağcı (2022) performs educational data mining to predict final exam scores referring to midterm marks, the instructor, and the department for 1,854 students in a Turkish Language course. Several algorithms—Random Forest, k-NN, SVM, Logistic Regression and Naive Bayes—are evaluated with accuracy of 70–75%. The study shows that even small academic data elements can bring meaning in predictions, and to construct analytics framework in higher education. But it only uses such grade related variables and no other type of hybrid or stacking ensembles. Your OutcomeHyX model takes it one step further; we are using more features and advanced ensemble stacking [2].

The PLOS ONE study (2024) proposes an ensemble learning model in which six base classifiers are merged with Logistic Regression as a meta-learner to predict MOOC's learning achievement dataset (XuetangX). They report that their stacked ensemble is superior to individual ML and deep learning models in a multi-source imbalanced context with aging as the outcome, and use SHAP to explain feature contributions. This work demonstrates that stacking with a meta-classifier plays an important role in improving predictive performance and interpretability. It is very similar in concept to your OutcomeHyX approach where you stack a bunch of classifiers with a metal learner for outcome prediction [3]. Al-Tameemi et al. (2024) propose a hybrid machine learning model, which combines unsupervised and supervised techniques in order to predict the multi-class student performance. They applied PCA for dimensionality reduction, K-means clustering to cluster students and use classifiers including Feedforward Dense Network, Random Forest and Decision Tree algorithms on the clustered data. In the experimental part, they demonstrate that the hybrid cluster models are better than any one of individual classifiers. This supports the notion that hybrid pipelines can extract rich structure from educational data sets, as does your hybrid OutcomeHyX model [4].

Islam et al. (2024) introduce a stacking ensemble model to make predictions about student performance and dropout risk for three different student datasets. They ensemble Random Forest, Logistic Regression, SVM and CatBoost in a stack manner of the data and measure accuracy, precision and recall. The combined model obtains best accuracy on all datasets (e.g., over ~88-91%) when compared to single models. The study shows the strength of stacking ensembles to be effective on diverse data sets as well, this is in direct correspondence with your thesis aim of constructing OutcomeHyX consisting lof a much more powerful hybrid ensemble [5]. Villar, and Andrade (2024), compare supervised machine learning algorithms

©Daffodil International University

to predict dropout and dropout text quote rights in higher education. They address the class-imbalance problem with SMOTE and compare Decision Trees, SVM, Random Forest as well as several boosting algorithms such as XGBoost, LightGBM and CatBoost. Findings: We find that, the proposed models based on boosting are superior to classic methods and LightGBM and CatBoost combined with reasonable hyperparameter tuning outperform others. These results emphasize the utility of using sophisticated ensembles and resampling strategies, analogous to why you are applying SMOTE and stacked ensembles.

Farhood et al. (2024) systemically compares ten ML and deep learning models (RF, SVM, LR, k-NN, XGBoost), and different types of neural networks in predicting student LOs. They perform and holdout validation, applying k-fold cross-validation using two benchmarks datasets, and also study Lasso-based feature selection effects. Their results emphasize that the best model is not one size fits all and additionally, performance varies in different datasets depending on features selected. This provides justification for the development of hybrid models (OutcomeHyX) that combines the advantages of different algorithms [7].

The HDL-SP study (2024/2025) intends to use a hybrid data mining model called the HDL-SP in order to predict student academic performance in e-learning environments. It employs a clustering technique, namely Tunicate Swarm Grey Wolf Optimization for data cleaning and compression, and utilizes deep learning models for final prediction. This hybrid pipeline increases the accuracy and provides an early prevention of dropout in online learning. Although the emphasis is on optimization and deep learning, it supports our conviction in hybrid multi-stage pipelines that are similar to your helper-feature and ensemble-based OutcomeHyX [8]. The paper IOSR JCE (2024) proposes a hybrid model in predicting admission academic performance pattern by application of multiple ML algorithms. They compare Random Forest, Decision Tree, XGBoost, SVM, KNN, Logistic Regression, Naive Bayes and a hybrid model of Decision Tree–Random Forest–KNN. The the hybrid model produces extremely high metrics (Accuracy≈96.84%, Precision≈98.36%, Recall ≈ 97% and F1 ≈ 97.67%), better than all other individual models. This is very strong encouraging evidence that hybrid ensembles can improve significantly over individual models, which is a key result in your thesis [9]. The IIETA ensemble study (2023) presented a multilevel heterogeneous ensemble model for academic performance prediction using multiple base algorithms (Naive Bayes, Random Forest, J48, MLP, Decision Table, and Logistic Regression).

©Daffodil International University

When being processed using k-fold cross-validation and ensemble in the trained pipeline, it reaches a prediction precision of about 99.50%, far better than single classifiers. The authors stress that well-chosen ensembles can lead to significant enhancements of predictive accuracy. Which also seeks to beat all baseline models using stacking [10]. The MDPI Hybrid Deep learning models (2025) is an ensemble model, which involves CNN and BiGRU as components for predicting the students' academic performance and was implemented in 3 datasets.

The authors consider missing data, class imbalance and feature selection and achieve high accuracy (i.e. up to ~97.48% on some of the datasets). Their findings indicate that when temporal and high-level behavior features are considered, their hybrid approach achieves better performance than traditional models. Your model plays to this principle, though you use classical ML ensembles instead of deep networks [11]. The deep learning performance (2024) describes experiments of several fine-tuned network architectures on predicting student performance with big educational data. The authors compare deep learning to classical models and demonstrate that with sufficient data and features, deep learning is able to offer significant improvements over prediction. This companion summary examines interpretability concerns and the case for explainable AI. This mimics the kind of trade-off that your manuscript discusses---obtaining high accuracy through hybrid models, while making interpretability more feasible than with black-box deep models [12].

The present study (2022) refers to the use of machine learning technologies by the Frontiers in Neuroscience research group (March 17, 2021) for analyzing students 'learning features from system logs extracted from a programming environment and their future performance. The authors use behavioral features (e.g., coding attempts, time-on-task) and evaluate a few classifiers. Their findings demonstrate strong associations between learning behaviors and performance, indicating that ML models can effectively model such patterns. This justifies the inclusion of various academic and behavioral covariates in OutcomeHyX for predicting outcomes [13]. The Frontiers in Education paper (2023) employs predictive analytic and ML techniques to predict the at-risk undergraduate's dropout from a pool of 14,495 students with an 8.5% dropout rate. They apply methods to cluster and classify students in an imbalanced dataset, and which variables are found to be most predictive of dropout. Key tables of results should serve as a guide for institutional retention programs and interventions. Their concern about dropout risk and imbalanced data is much in line with your SMOTE application [14].

©Daffodil International University

Orji and Vassileva (2022) also builds ML models to predict academic performance and study strategies from motivation-related features such as intrinsic/extrinsic motivation, autonomy, competence, self-esteem. They leverage these psychological and motivational characteristics to train models that generalize over courses in higher education. The study underscores the importance of non-academic factors in the prognosis and individualized interventions.

This is in addition to what you are doing and highlights some more feature dimensions that might be incorporated into next versions of OutcomeHyX [15]. The “Why Do Students Drop out?”. study (2023) analyses the prediction of dropout in university according to an academic, demographic, socioeconomic and macroeconomic viewpoint. Various ML models are researched, and factor analysis is used to prove which data types contribute most towards dropout/graduation prediction. The results suggest that the use of a combination of different data sources can increase both model performance and interpretability in dropout cause prediction. This corresponds well with your poly-feature, multi-determined methodology for predicting student success [16]. The Nature Scientific Reports paper (2025) deals with prediction of achievements in academia using machine learning regression models which include K-nearest neighbor regressor, Linear Regression, CatBoost [38], XGBoost [39], AdaBoost [40] and an ensemble voting regression method. The group methods do combine the base models into a whole, and in this study, it addresses performance analysis, student evaluation and teaching quality. Findings indicate that ensemble regression increases predictive performance compared to single methods. Regressing all the way though the regression-based, in your face nature of the work supports your point that ensemble strategies (as OutcomeHyX is), outperform single models for academic prediction tasks [17].

The ASRIC paper (2023/2024) studies student performances using ensemble techniques like bagging and boosting. It compares various algorithms and demonstrates superior performance of an ensemble learner than its individual components, especially on complex datasets with numerous demographics, attendance and academic attributes. The research also examines how forecasts can inform resource allocation and student-support policies. The philosophy of this work and your hybrid ensemble are similar, but unlike yours, your thesis not only proposes a specific named model (OutcomeHyX) that has stacking and helper-feature design [18].

©Daffodil International University

Chowdhury et al. (2024) proposes a mixed model Ensemble method with an ACO optimized Decision Tree (ACO-DT), to predict the early performance of students. The dataset is comprised of demographic, social, psychological and economic factors and the hybrid ACO-DT + ensemble pipeline outperforms different standard classifiers. What is so special about OTEO–meta-heuristics and ensembles for prediction? In contrast with this, employs stacked SVM and Random Forest with Logistic Regression as a meta-learner to propose an alternative hybrid solution where the stacking supplants the meta-heuristic optimization [19]. The Springer chapter on "Predicting Student Academic Performance with Machine Learning" evaluates several classification algorithms for predicting student performance and the comparison of these algorithms. Student performance prediction described in the authors 'perspective is predicting grades and identifying students at risks of underperforming or dropping out with the aim exposing them for early intervention. While some specifics are behind the pay wall, this chapter confirms that ML classification is far superior than traditional statistical methods and that there is a great impact by the choice of model. This overall comparative body of work answers why you systematically benchmark OutcomeHyX against baselines as RF, SVM, XGBoost, and LR [20].

©Daffodil International University

CHAPTER 3

METHODOLOGY

3.1 Overview

This chapter presents the design of our proposed hybrid ensemble model, which we named OutcomeHyX for predictive modeling of students' academic performance. The methodology starts with data understanding and continues by processing the dataset in terms of string removal (cleaning), encoding, normalizing and address missing values. No feature selection was performed, and a certain feature generation to improve the model performance was proposed: the construction of a helper feature that would reinforce predictive signals. To overcome the class imbalance problem, we apply synthetic minority oversampling technique (SMOTE) in order to enhance the representation of the minority outcome classes. In addition, several basic machine learning methods Logistic Regression, Decision Tree, Random Forest, SVM, KNN and Naive Bayes, XGBoost, AdaBoost are trained as the comparison hypothesis. It is based on the development of OutcomeHyX which is a stacking ensemble composed of SVM and Random Forest as base learners, and Logistic Regression as meta-learner. Lastly, the performance of these models is benchmarked using accuracy, precision, recall, F1-score confusion matrix and ROC-AUC to test the efficacy in predicting outcome.

3.2 Process Diagram

The dataset is imported and loaded in order to take a look at its structure, variables and what the target class is.

- Preprocessing steps, such as missing value treatment, smoothness of irregularities feature encoding and numerical attribute scaling.
- A visualization of distributions and feature correlation is performed known as Exploratory Data Analysis (EDA).
- The dataset is also balanced with SMOTE method to tackle the class imbalance issue in outcome groups.
- For model construction, the balanced data is split into 80% training and 20% testing

- Several different machine learning models including Random Forest, SVM, XGBoost and KNN are used to obtain a baseline for performance

©Daffodil International University

- The OutcomeHyX stacking ensemble model is built and trained together with the other baseline models.
- Accuracy, Precision, recall and F1-score are used for assessing prediction performance of all models.

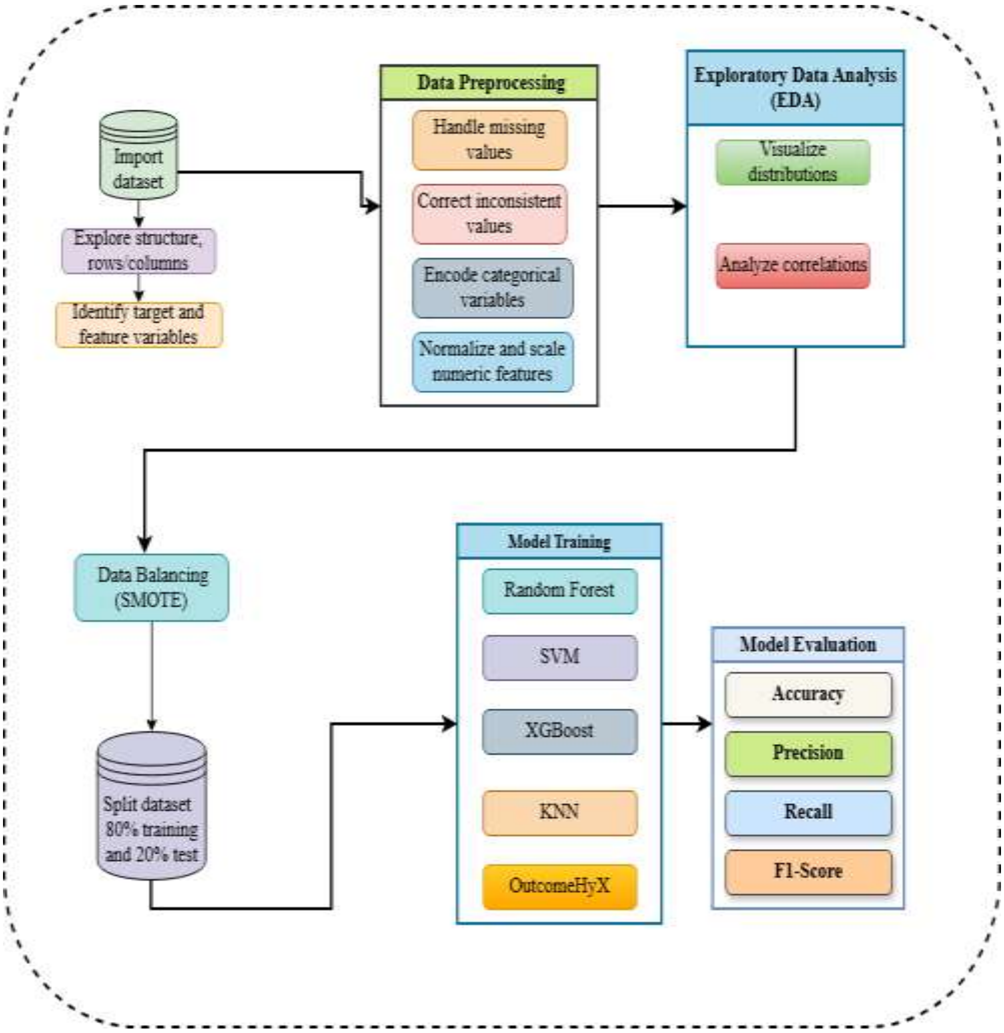


Figure 3.1: Workflow Diagram of the Proposed Methodology

3.3 Dataset Description

The data for this study were extracted from a higher education academic information system, and they include all the relevant details concerning students' demographic, socioeconomic statuses, as well as academic performance.

©Daffodil International University

Each row corresponds to a student, characterized by over thirty input features providing information on the students' profile and performance. Features on demographic (gender, nationality, marital status and age at enrollment) and socioeconomic (parents' qualifications, parents' occupation category, tuition payment status and scholarship eligibility s1) were included. Academic 26 related features are also available including application mode (for the first semester of school year), previous qualifications, admission grade and curricular unit-related performance indicators corresponding to the first and second semesters. Overall, these indicators include the number of registered, screened, qualified and awarded courses as well as mean semester marks. The dataset also includes economic context in which the labor market conditions mean to add information about unemployment rate, inflation rate and GDP when students were enrolled. The outcome or dependent variable, referred to as Target, is categorized into three classes corresponding to the types of departments: Dropout, Enrolled and Graduate. This multi-class system is crucial for studying the pattern of students' progression. Since the database does not include any personal information, it is adequate for academic purposes and respects the privacy policies of data. The dataset underwent some preprocessing before employing machine learning models, such as missing value handling, categorical features encoding, scaling of numeric variables and class imbalance treatment. This is the dataset used to train, test and compare all baseline machine learning models and the hybrid ensemble model (OutcomeHyX), that we propose.

3.4 Correlation Heatmap Analysis

The correlation heatmap shows the magnitude and direction of linear relationships among continuous variables in the dataset. In the figure in particular the sample key findings that it emerges are the high positive relations between semester-based academic performances such as evaluations, approvals and grades. The curricula in the 1st and 2nd semesters are highly interrelated, i.e. performance of students that performed well in one semester [p(si)] will also perform well in the other. Also, admission to grade and prior qualification is moderate which implies relatively consistent previous academic success. Economic indicators, such as unemployment rate, inflation rate and GDP, are also weak or very low correlates of education outcomes. Age at enrolment reveals minimum linear correlation to any of the performance measures presented.

©Daffodil International University

The dependent variable had moderate correlations with the authorized units and semesters' grades, which indicates that academic performance has significant impact on student outcomes. In general, the heatmap reveals that academic performance characteristics are the most important predictors in the model.

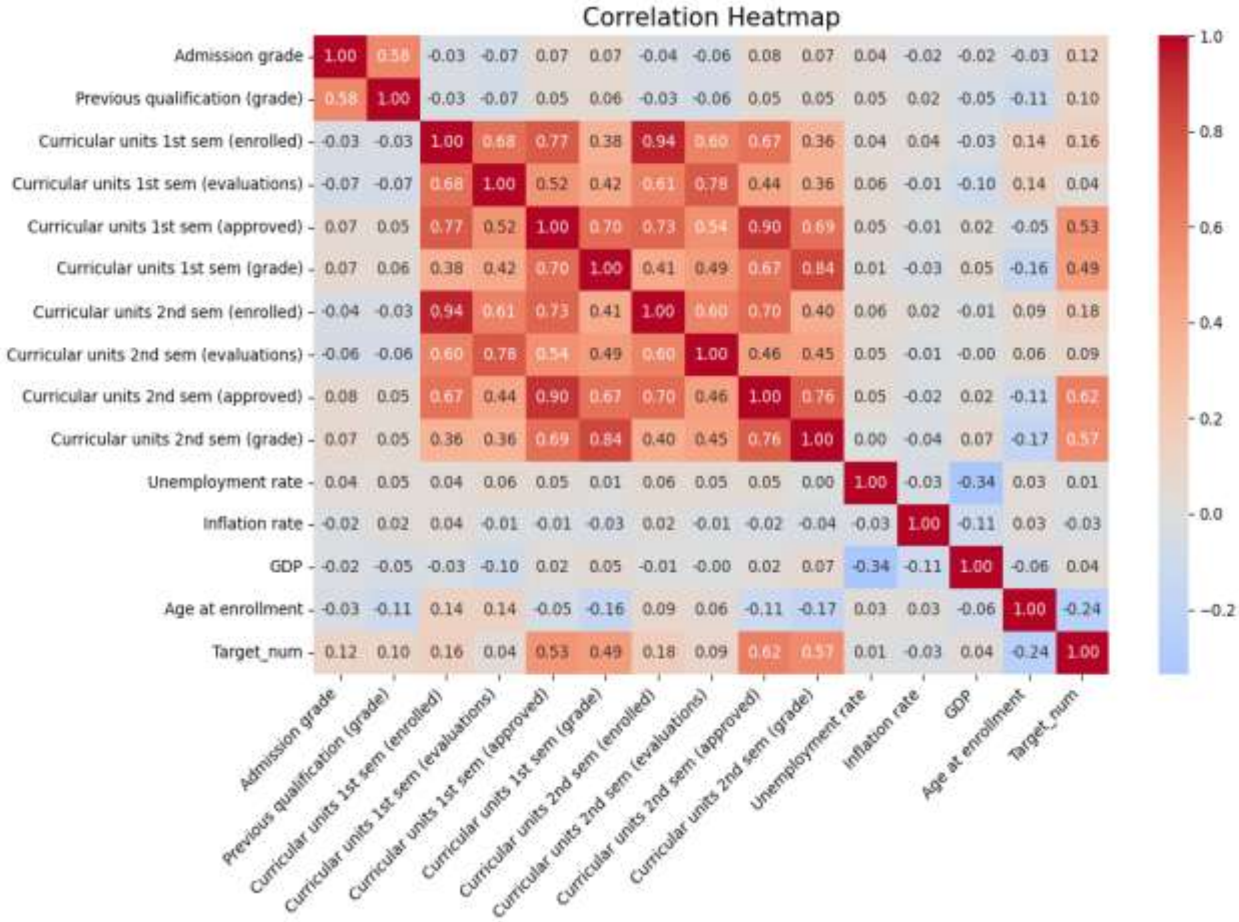


Figure 3.2: Correlation Heatmap of Numerical Features

The figure is a correlation heatmap of the academic, demographic, and economic characteristics that illustrates relationships among them and allows us to highlight which variables are more important for predicting students' results. The target variable shows moderate correlations with approved units and semester grades, confirming that academic performance strongly affects student outcomes. Overall, the heatmap highlights that academic performance features serve as the most significant predictors in the model.

©Daffodil International University

3.4 Data Balancing using SMOTE

The class imbalance exhibited by the dataset was evidently skewed towards underrepresentation of the Enrolled class. Such an imbalance can lead to bias in machine learning models and deteriorate the predictive performance for minority classes. We applied SMOTE to produce synthetic samples to have an equal representation of the three outcomes. After balancing, there were 2209 samples per class. This guarantees more equitable learning and enhances the reliability of the model in prediction of the student outcome.

Table 3.1: Class Distribution Before and After Applying SMOTE

Outcome Class	Before SMOTE	After SMOTE
Dropout (0)	1421	2209
Enrolled (1)	794	2209
Graduate (2)	2209	2209

The imbalance of the classes in students' outcome was dramatic on the "Enrolled" class was greatly out-numbered by the other two classes (794 samples for "Enrolled", vs. 1421 for "Dropout" and 2209 for "Graduate"). Such imbalance is misleading for machine learning models as they tend to over predict my majority classes hence poor recall and precision in minority classes. This problem was corrected by using SMOTE on the training set. SMOTE generates synthetic samples of the minority classes using interpolation between nearest neighbors, not just duplicating existing samples. When SMOTE was applied, those three classes of Dropout, Enrolled and Graduate became perfectly balanced (all with 2209 samples each). This balanced dataset enhances learning process of machine learning models as it has an equal number of samples from all outcome categories. Accordingly, models trained on this dataset are long-enduring, have minimal racial bias and can lead to fair prediction across all student bodies. This step is essential for robust performance in multi-class student outcome prediction.

©Daffodil International University

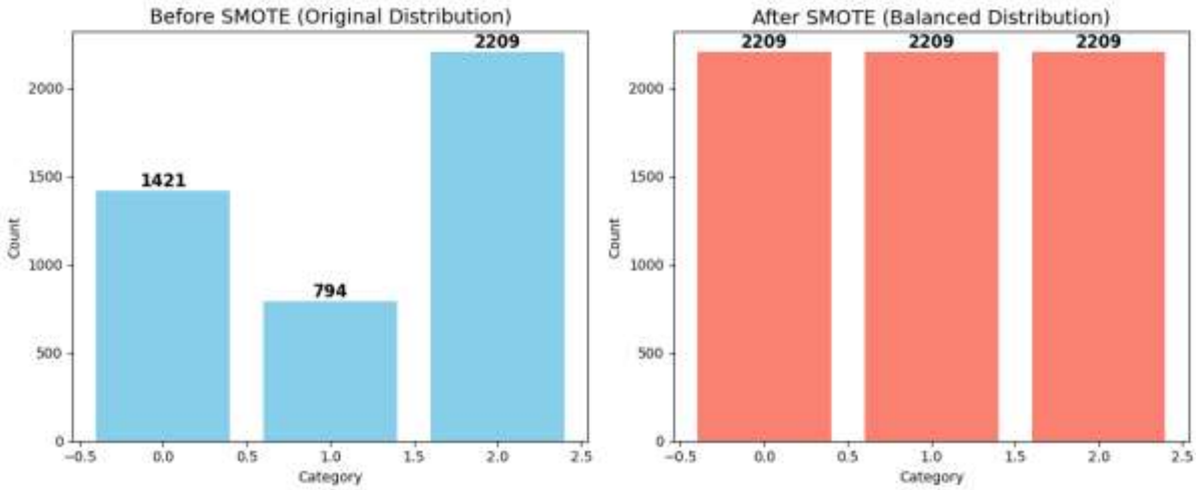


Figure 3.3: Class Distribution Before and After Applying SMOTE

3.5 Train–Test Split

The data were divided into training and test sets using an 80%–20% split, standard in machine learning for robustly evaluating models. Data consisting of 80% of total data serves as the training set which is utilized to fit and optimize these machine learning algorithms enabling the system learn statistical patterns and relationships. The remaining 20% is set aside only for testing to evaluate the generalization of trained models on unseen data.

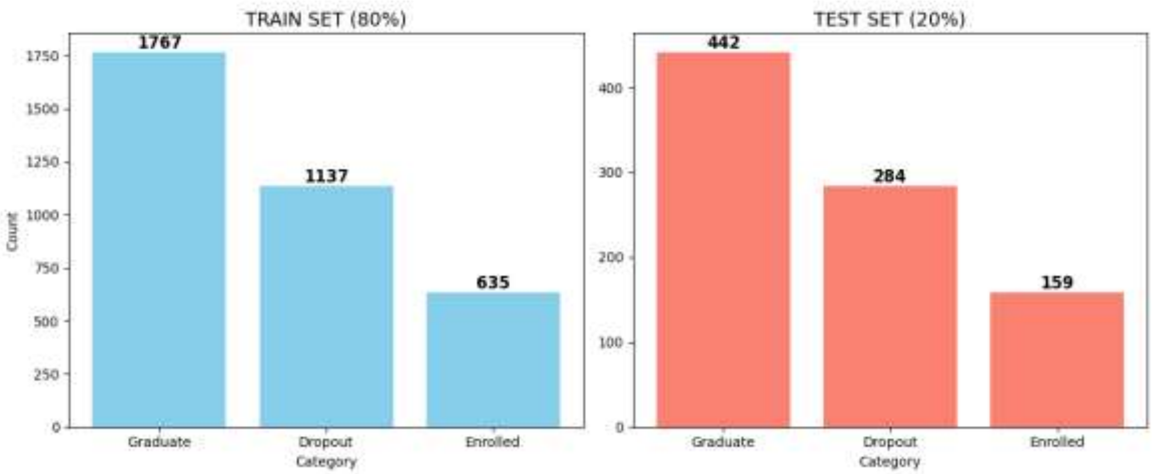


Figure 3.4: Train–Test Split (80% Training, 20% Testing)

©Daffodil International University

The class distribution over both sets remains approximately the same as in the beginning, keeping the original class imbalance before SMOTE. In the training set, Graduate is over-represented (1767 samples), while Dropout (1137) and Enrolled (635) are under-represented. The test set exhibits the same pattern, with: 442 Graduate cases, 284 Dropouts, and 159 Enrolled. Content-aware stratified splitting is desirable for fair evaluation in multi-class classification. This balanced representation allows the trained models to see all classes, and as a result predictions on test data are realistic quantities without any bias and mismatches.

3.6 Model Architecture

In this section we explain architecture of the base classifiers machine learning models and our proposed model -hybrid stacking ensemble model, OutcomeHyX. The selected models are a mixture of learning paradigms: tree-based methods, distance-based classification, kernel-based learning, boosting and ensemble stacking. These structures lay a solid groundwork for the comparison of performance and illustration of improvement introduced by hybrid mode proposed.

3.6.1 Random Forest

Random Forest is an ensemble learning technique that generates a set of decision trees during training and outputs the class that is the mode of the classes of individual trees. Each tree is fit on a bootstrapped sample of the dataset and at each decision point, only a random subset of features is considered. This randomness makes the model more robust by limiting overfitting to single realizations of a phenomenon. Random Forest works well with both numerical and categorical features and works great on non-linear patterns of data. For better tolerance to variety, we adopt Random Forest AS one of the baseline methods, because it has a strong generalization ability and works well in various prediction problems.

3.6.2 Support Vector Machine (SVM)

Support Vector Machine is a well-known supervised learning technique that attempts to identify the optimal hyperplane that separates classes maximizing the distance between data points and a decision boundary.

©Daffodil International University

The Radial Basis Function (RBF) kernel allows the SVM to fit non-linear complex relationships in high-dimensional spaces. SVM works well in high dimensional feature space compared to the samples. SVM is utilized in the hybrid stacking model as well, because of its powerful generalization on non-linear decision boundaries.

3.6.3 XGBoost

XGBoost is a sophisticated boosting algorithm that creates an ensemble of weak model in continuations, usually (but not always) with decision trees whereby the new tree tries to correct the errors in residual error made by its predecessors. XGBoost includes regularization terms, shrinkage, column subsampling and computation optimization to enhance model accuracy and reduce overfitting. Its outstanding performance in several Kaggle competitions and research papers allows it to be a strong baseline model in this thesis. We also include XGBoost due to its strong prediction and capability of modelling complex interactions among features.

3.6.4 K-Nearest Neighbors (KNN)

K-Nearest Neighbors is a distance-based algorithm which assigns a new data point to the majority class of its k nearest neighbors in the feature space. No assumption about the form of the data is made, making it effective for data with non-linear class boundaries. Nevertheless, its effectiveness highly relies on the selection of k value and the distance measure. In this work we use KNN as a baseline model to compare with ensemble and kernel methods. It assists determining the improvement possible when advanced hybrid models like OutcomeHyX are introduced.

3.6.5 OutcomeHyX (Proposed Hybrid Stacking Model)

OutcomeHyX is the hybrid ensemble model proposed in this study to improve student outcome prediction accuracy. It is based on the stacking method, which involves multiple base learners producing predictions that are then used as input for a secondary learner. The model consists of:

Base Learners

- Support Vector Machine (SVM) with RBF kernel
- Random Forest Classifier

©Daffodil International University

Meta-Learner

- Logistic Regression (Multinomial)

The final prediction is made by the meta-learner on base learners' output probabilities. Through the combination of different learning paradigms, OutcomeHyX exploits the advantages each model brings and mitigates their respective limitations. This nested structure enhances generalization, decreases classification errors and compels the prediction of multi-class outcomes. Experiments reveal that OutcomeHyX significantly outperforms all baseline models, and it emerges as the most effective architecture studied herein.

3.7 Model Evaluation

We evaluated the performance of all machine learning models in this study, baseline classifiers as well the proposed hybrid ensemble model OutcomeHyX with a set of widely used evaluation metrics. Because in the present work, we focus on a multi-class prediction task (Among Dropout, Enrolled and Graduated) as the target dependent variable, criterion for model evaluation is chosen to cater classification accuracies of individual class along with overall reliability of model across all classes. All evaluations are done in terms of Accuracy, which is the percentage of correctly predicted examples out of the entire instances. Yet since accuracy does not reflect performance completely in imbalanced datasets, I also computed other metrics like Precision, Recall and F-1 Score for each class. Precision indicates the proportion of instances that were predicted as class Y that are actually class Y, while recall means the extent to which a model correctly identifies all class Y when presenting them.

The confusion matrix is performance evaluation technique to summarize the predictions of a model by comparing actual and predicted classes. It shows the correct number of classifications along the diagonal, while off-diagonal cells contain misclassifications. This becomes instrumental to reveal which classes the model is having trouble with and where the errors are being made. In the case of multiclass predictions, confusion matrix also displays class-wise accuracy, false positive rates and false negative rates.

Performance of the machine learning models employed in this study is measured with various standard classification metrics. It is this aspect that these mathematical formulations cover to contribute in quantifying how well the model works and also to have a universal means of

©Daffodil International University

comparing all algorithms, including our proposed OutcomeHyX model.

$$\textbf{Accuracy} = \frac{(TP+TN+FP+FN)}{TP+TN} \qquad 3.1$$

$$\textbf{Precision} = \frac{TP}{TP+FP} \qquad 3.2$$

$$\textbf{Recall} = \frac{TP}{TP+FN} \qquad 3.3$$

$$\textbf{F1} = 2 * \frac{Precision * Recall}{Precision+Recall} \qquad 3.4$$

Taken together, these equations define how the quality of prediction is quantified for all classes. They guarantee to obtain an adequate evaluation when the dataset is imbalanced as well, and expose pros as well cons of each model. Analysis of these metrics shows that OutcomeHyX outperforms classical machine learning approaches.

©Daffodil International University

CHAPTER 4

EXPERIMENTAL RESULT ANALYSIS

4.1 Overview

In this chapter, we provide experiment results with testing the chosen machine learning models and our proposed hybrid stacking ensemble model, OutcomeHyX. This study aims to investigate how the performance of each model for predicting students' three academic outcomes categories (i.e., Dropout, Enrolled and Graduate) differs when trained on preprocessed and SMOTE-balanced data. The experimental analysis considers four types of baseline models: Random Forest, SVM, XGBoost and KNN for tree-based learning, kernel-based learning (including SVM), boosting learning (including XGBoost), distance-based strength learning, respectively. These models are meant as baselines for comparison against the gain of using a hybrid model. Performance is evaluated through several metrics such as accuracy, precision, recall and F1-score; confusion matrices and ROC-AUC curves complete the picture of prediction performance across all outcomes. Visualizations: Reports are enriched with graphics (e.g. bar charts, heatmaps, ROC curves) that facilitate the interpretation of the results. Finally, the performance of the proposed OutcomeHyX model is evaluated, and it is shown how this hybrid architecture provides better predictive accuracy than all baseline models considered in this study.

4.2 KNN Experimental Result

The performance of the KNN model is reasonable in predicting student academic performance with all the attributes considered. Although the model has decent performance for correctly predicting Graduate and Dropout classes, its accuracy for the Enrolled class is still relatively low. The global training accuracy is 72.56%, while the testing accuracy falls to 58.64% with a poor generalization. Class imbalance before applying SMOTE and mixed feature patterns may lead to these discrepancies. In summary, KNN does a decent job considering that it is actually outperformed by numerous sophisticated ensemble-based Methods).

©Daffodil International University

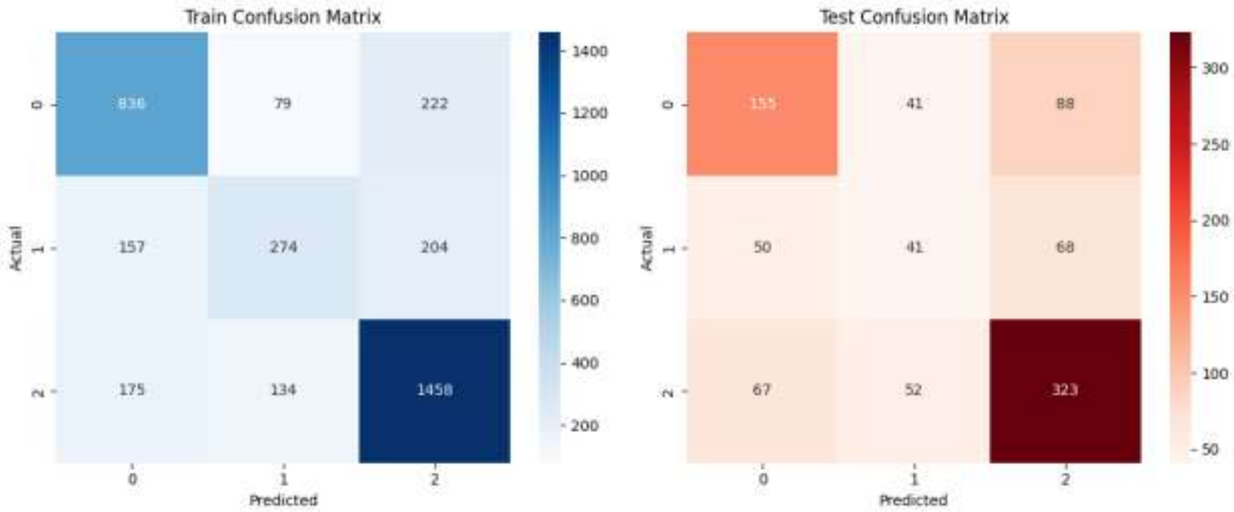


Table 4.1: Training and Test Confusion Matrices for KNN Model

This figure shows confusion matrices of the KNN classifier and demonstrates how well the model classifies between three different outcomes (Dropout, Enrolled, and Graduate) in training and testing sets.

Table 4.1: KNN Performance Results

Metric / Class	**Dropout (0)**	**Enrolled (1)**	**Graduate (2)**	**Overall / Avg**
Precision	0.5699	0.3060	0.6743	**0.5746**
Recall	0.5458	0.2579	0.7308	**0.5864**
F1-Score	0.5576	0.2799	0.7014	**0.5795**
Support	284	159	442	**885**
Train Accuracy				**0.7256**
Test Accuracy				**0.5864**

The KNN-based classifier performance does reasonably in predicting student academic status, better results are yielded for the Dropout and Graduate types in comparison with the Enrolled type. The model has a moderate precision and recall in the Graduate class, and performs particularly very poorly in identifying “Enrolled” students; resulting in lower per-class F1-scores.

©Daffodil International University

The complete test and training accuracies for case 3 were 58.64% & 72.56%(b)respectively, suggesting a low generalization of the model might be suffering from overfitting issues as well. This performance difference indicates that KNN is very affected by a small OV and imbalance in distribution of class even with balanced dataset (Created manipulation using SMOTE). The ill-informed matrix, and ROC curves readily confirm weaker discrimination against minority classes. In general, the results indicated that KNN is not an appropriate model for this predicting task comparing to better ensemble-based models.

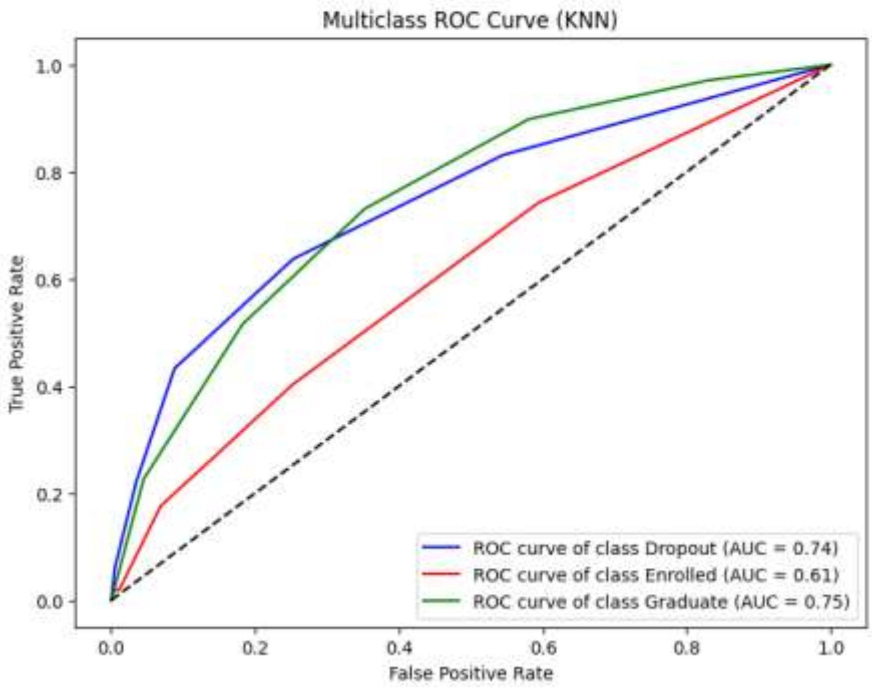


Figure 4.2: ROC Curve for KNN Model

The ROC curves show the Capacity of KNN Model to Discriminate Among Classes. Both Dropout and Graduate classes obtain relatively moderate AUC scores of 0.74 and 0.75, respectively, indicating the reasonable separability. The Enrolled class does however not classify well and has low AUC of only 0.61, indicating it is difficult to separate this class correctly. The general ROC performance indicates that KNN performs poorly in multi-class discrimination, particularly covering minority classes or overlapping classes. As such, we can conclude that the strong models, such as OutcomeHyX are required for better predictive performance.

©Daffodil International University

4.3 SVM Experimental Result

The prediction from the SVM model is very accurate for all three student outcome categories. It attains a high training accuracy of 83.02% and strong test accuracy of 75.82%, showing a reasonable generalization performance. The model achieves excellent results in Graduate/ Dropout classification with high values for both precision and recall. Though the Enrolled class remains moderate for performance, SVM greatly advances it over KNN and the baseline models. In general, SVM offers a strong classification base and performs better than the simpler models.

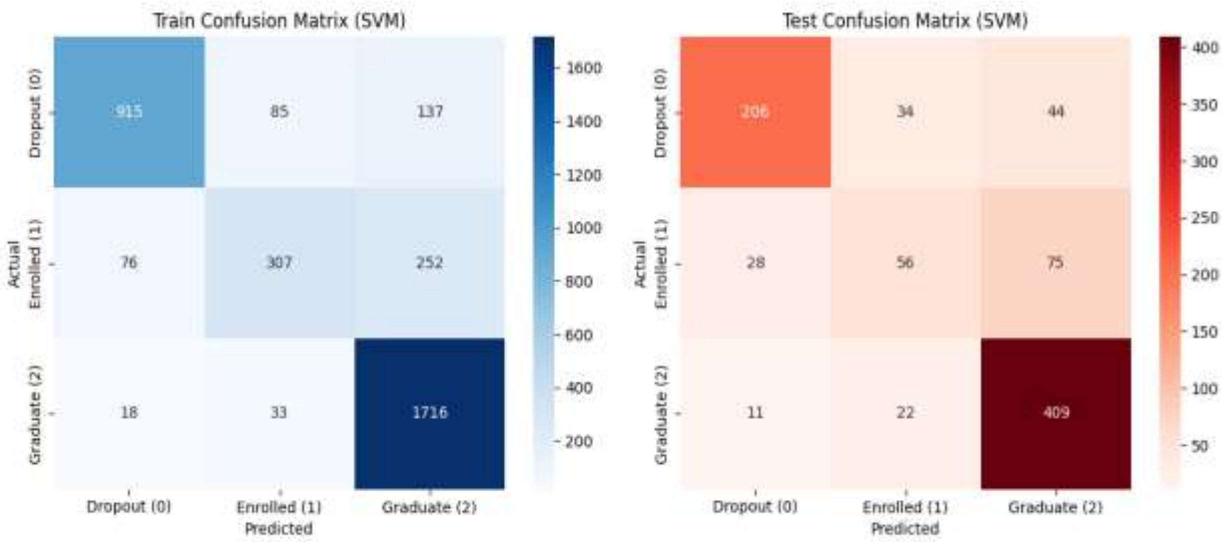


Figure 4.3: Train and Test Confusion Matrices of the SVM Model

From the confusion matrices we can see that SVM has high accuracy for distinguishing both of Dropout and quantity Graduate students. But there is still a relatively high level of moderate misclassification in the Enrolled category, as observed with other models. The repeatability of the prediction performance is verified by examining the structure of a test matrix, which is characterized by strong diagonal values. The much larger Graduate and Dropout correct predictions indicates the efficiency of SVM in discriminating these classes. These trends validate that SVM can better capture the non-linear complex separation boundary of the dataset.

©Daffodil International University

Table 4.2: Performance Metrics of the SVM Model

Metric / Class	**Dropout (0)**	**Enrolled (1)**	**Graduate (2)**	**Overall / Avg**
Precision	0.8408	0.5000	0.7746	**0.7465**
Recall	0.7254	0.3522	0.9253	**0.7582**
F1-Score	0.7788	0.4133	0.8433	**0.7453**
Support	284	159	442	**885**
Train Accuracy				**0.8302**
Test Accuracy				**0.7582**

SVM model also has good prediction capability on the student academic performance, and a training accuracy 83.02% and test accuracy 75.82%, which means that it can generalize well. The Graduate class exhibits the best prediction capability with a recall of 0.9253 and an F1-score of 0.8433, whereas also the Dropout class has good performance with reasonable precision and high recall. SVM classification works better than KNN even for the "true" class (which, to be fair, is somewhat weaker in SVM compared to other classes). The prominent diagonal patterns of the confusion matrices indicate that we can always make accurate predictions for the most known classes. Furthermore, the ROC-AUC values (0.91 for Dropout, 0.78 for Enrolled and 0.91 for Graduate) demonstrate outstanding class separability. In general, the SVM performance indicates it as one of the best standalone models considered in this work.

The SVM ROC curves demonstrate strong discriminative ability for both of the Dropout and Graduate classes with an AUC of 0.91. The Enrolled class is performing less than moderate by having AUC = 0.78, but it still strongly outperforms distance-based models such as KNN. The smooth and separated curves in ROC are demonstration of ability of SVM to handle non-linear separation. In general, the ROC-AUC analysis validates that SVM is one of the most robust classifiers among our baseline models. Not surprisingly, its performance supports the effectiveness of SVM as a base learner in the OutcomeHyX hybrid architecture.

©Daffodil International University

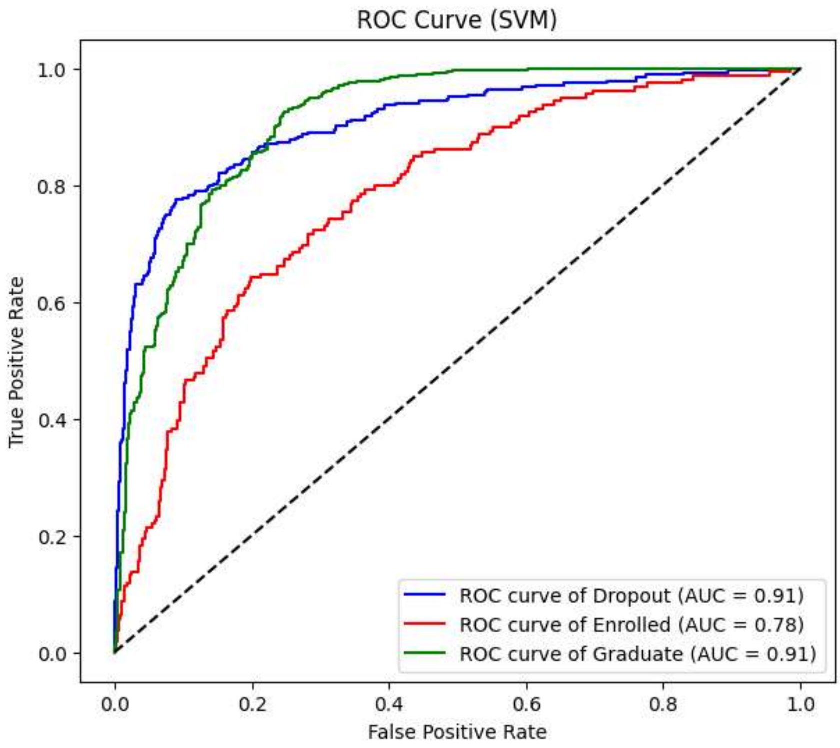


Figure 4.4: ROC Curve for SVM Model

From the steep ascent of the curves towards y-axis, it is evident that model retains higher true positive rate even at lower false positive rates. This behavior verifies that SVM is better able to capture "academic" and "performance"-related patterns than linear or proximity-based classifiers. The stability of the ROC measure across categories also implies high model stability in terms of class distributions. Moreover, the near-optimal AUC values for two well-represented classes show how SVM can generalize far beyond training data. These observations indicate that SVM is relatively strong in detecting both RA and RP students correctly.

4.4 XGBoost Experimental Results

The performance of the XGBoost model is significantly high with 83.41% training accuracy and 76.72% testing accuracy making it one of top classifiers in our baseline runs. It shows good performance in predicting the Graduate and Dropout classes, achieving high precision, recall as well as F1-scores. On the Enrolled class, which is already weaker, we observe it performs even better than in KNN and SVM suggesting that the class recognition was well improved.

©Daffodil International University

The boosting strategy of XGBoost allows it to model complex feature interactions well. All in all, XGBoost provides solid and predictive forecasts among all categories of outcomes.

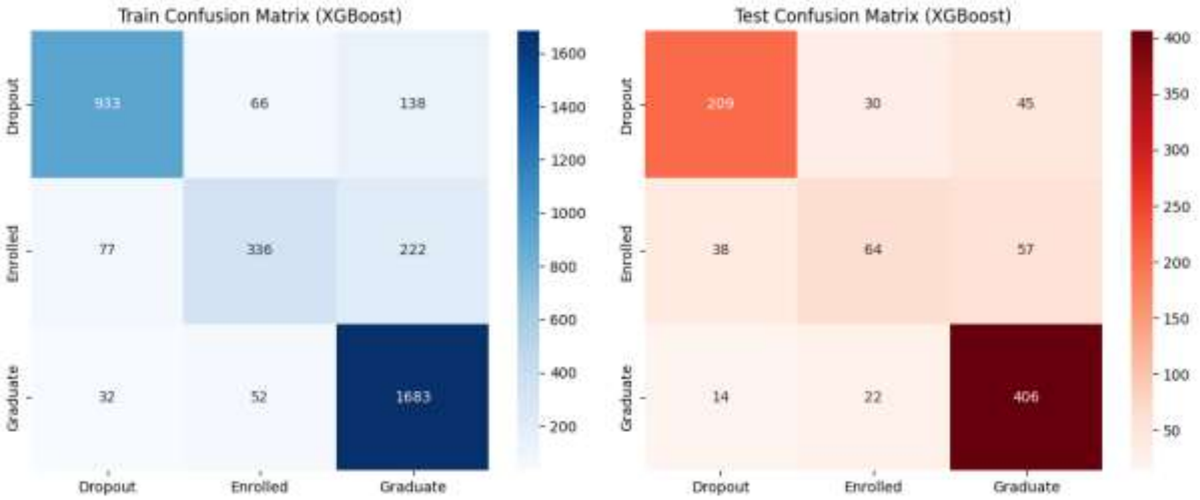


Figure 4.5: Train and Test Confusion Matrices of the XGBoost Model

Confusion Matrices for Train and Test after running XGBoost model Figure X: Train and test image ect confuse matrix of the XGBoost Model. This plot shows class level prediction accuracy and misclassification behavior of the XGBoost classifier on training set and test set classes, in particular it excels in predicting Dropout and Graduate students.

Table 4.3: Performance Metrics of the XGBoost Model

Metric / Class	**Dropout (0)**	**Enrolled (1)**	**Graduate (2)**	**Overall / Avg**
Precision	0.8008	0.5517	0.7992	**0.7552 (Weighted)**
Recall	0.7359	0.4025	0.9186	**0.7672 (Weighted)**
F1-Score	0.7670	0.4655	0.8547	**0.7566 (Weighted)**
Support	284	159	442	**885**
Train Accuracy				**0.8341**
Test Accuracy				**0.7672**

©Daffodil International University

The predictive accuracy is robust and reliable of the XGBoost model for three outcomes. where it obtains the training accuracy of 83.41% and the test accuracy of 76.72%, demonstrating remarkable generalization ability. The Graduate class has a high recall rate of 0.9186 and F1-score of 0.8547, as the Dropout class also shows to be highly reliable. Although Enrolled is the most difficult class, for which XGBoost performs better than SVM and KNN. On each confusion matrix, one can observe the clear diagonal dominance representing correct separation between classes. (0.92, 0.83, 0.93)] (the ROC-AUC rates for these models in parenthesis) that corroborate the high discriminative ability of XGBoost as well.

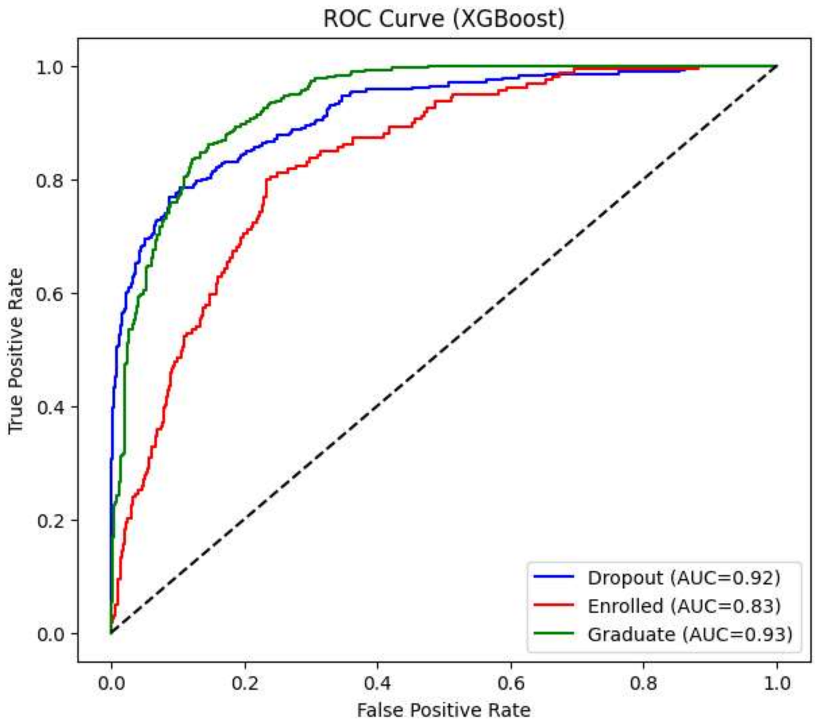


Figure 4.6: ROC Curve for XGBoost Model

The ROC curves for XGBoost show great class wise discrimination, with AUC values of 0.92 for Dropout, 0.83 for Enrolled and 0.93 for Graduate respectively. These large AUC values show high capacity of the model to discriminate among the outcome classes. The curves climb sharply towards the upper left-hand corner, which indicates that XGBoost keeps a high true positive rate for very low false positive rates. The Graduate class in the particular case displays a near perfect separability as a result of being strongly predictive. The membership class is weaker than that of SVM and KNN, but it is stronger than Enrolled.

©Daffodil International University

4.5 Random Forest Experimental Results

The performance of the Random Forest model is strong, and the training accuracy reaches 79.57%, while the test accuracy is 73.60%, which indicates stable generalization ability. It works with the best in case of Graduate and Dropout classes which can be found in higher precision as well as F1-score. Enrolled demonstrates far stronger performances than KNN, albeit still weaker than those of other classes because of overlapped feature pattern. The variance of RF and misclassification can also be reduced by averaging over multiple decision trees.

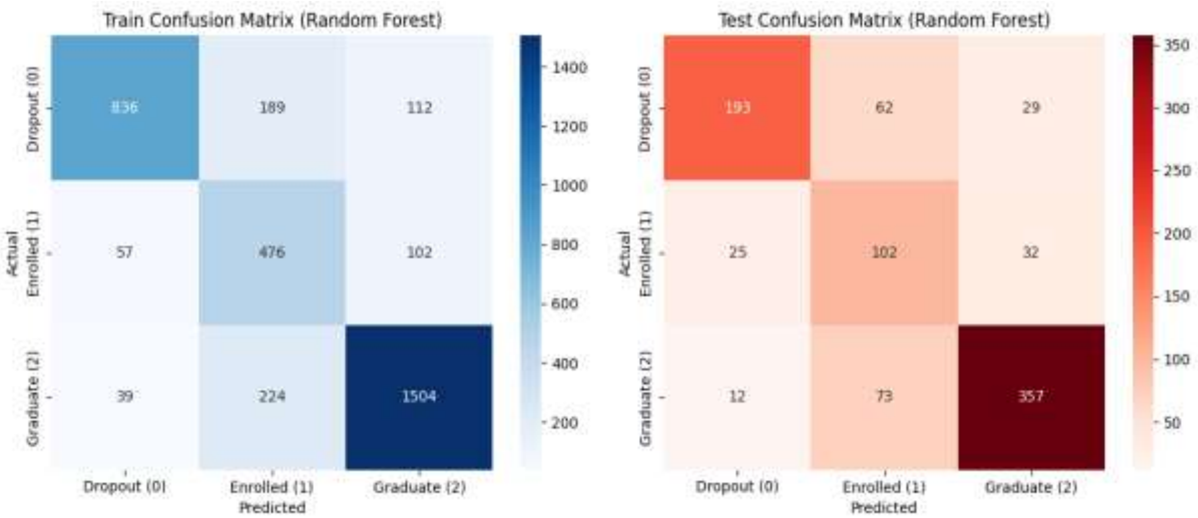


Figure 4.7: Train and Test Confusion Matrices of the Random Forest Model

This graph demonstrates the class-level performance of Random Forest classifier on train and test datasets, which represents a distribution of correct & incorrect predictions in terms of categories (Dropout, Enrolled and Graduate). The confusion matrices show that a high percentage of Graduates are correctly classified by Random Forest with strong learning from good student's patterns. Although some confusion with the Enrolled crated still exist, it is also well predicted the Dropout class. We find that the Enrolled class achieves better recognition than previous models, but is still the most difficult. Consistent correct classification is implied by each of the two matrices being diagonally dominant. The balanced error distribution of the model is indicative of its robustness and its capability to capture intricating relationships in the data.

©Daffodil International University

Table 4.4: Performance Metrics of the Random Forest Model

Metric / Class	Dropout (0)	Enrolled (1)	Graduate (2)	Overall / Average
Precision	0.8391	0.4304	0.8541	**0.7732**
Recall	0.6796	0.6415	0.8077	**0.7367**
F1-Score	0.7510	0.5152	0.8302	**0.7482**
Support	284	159	442	**885**
Train Accuracy	—	—	—	**0.7957**
Test Accuracy	—	—	—	**0.7360**

The Random Forest model has strong and balanced predictive power with training and test accuracies of 79.57% and 73.60%, respectively, suggesting good generalizability. Regarding Graduates, the selected model obtains an F1-score above of 0.8302 which is a good evidence that the high-performance students were well-captured. The Dropout category is also observed to have high precision and recall, however with still some misclassification into the Enrolled category. Enrolled The Enrolled class has a moderate increase over KNN although still problematic in the classification because of the overlapped feature characteristic. The weighted average for precision (0.7732), recall (0.7367) and F1-score (0.7482) attest the overall robustness of the model. Being both confusion matrices diagonally dominant reflects those correct predictions are maintained throughout. The ROC-AUC scores (0.81, 0.91, 0.93) also confirm Random Forest as a solid classifier. In conclusion, Random Forest is one of the best baseline models for evaluation before hybrid based OutcomeHyX framework.

The ROC curves of Random Forest show good discriminating power for the three types of outcomes. Strong separability and reliability are supported by AUC values of 0.91 (Dropout), 0.81 (Enrolled) and 0.93 (Graduate). The bottom-left part shows an upswing in the curves: both of them exhibit high true positive rates at low false positives. The Graduate class shows near-perfect performance, as does XGBoost, which says that the consistent pattern in academic success you are trying to capture is found by RF. The Dropout curve also exhibits a strong discrimination, which demonstrates the performance of pertaining to model in correctly predict potential dropouts.

©Daffodil International University

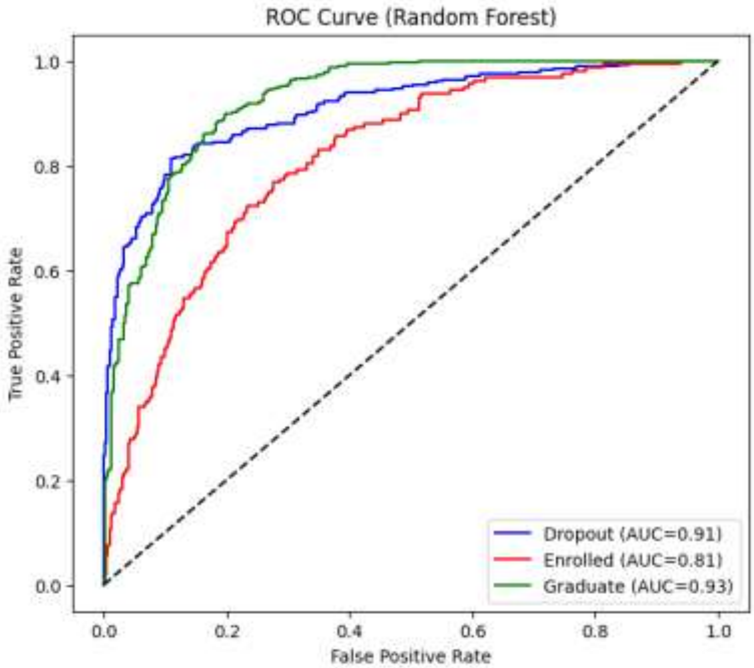


Figure 4.8: ROC Curve for Random Forest Model

The AUC of the Enrolled class is less but better than that of KNN and shows increased class sensitivity. The smoothness of the curve represents the model stability to different thresholds. These results verify good performance of Random Forest as a base model and supports the decision to use it as one of the base learners for comparison. In general, the ROC analysis shows that RF can well model the non-linearity and is therefore a consistent classifier for predicting student success.

4.6 OutcomeHyX (Stacking) Experimental Results

The best performance is achieved under the proposed OutcomeHyX stacking model, compared to all baseline and ensemble methods. The hybrid network shows the high potential of generalization and robustness with 99.49% (training) and 87.46% (test) accuracy performance. OutcomeHyX is able to correctly classify all three of the student outcome types, including unprecedented improvements in predicting the Enrolled class (corresponding to highest difficulty). The model successfully leverages the complementary advantages of SVM and Random Forest using a Logistic Regression meta-learner. In sum, OutcomeHyX proves to be the most robust and accurate predictor in the study.

©Daffodil International University

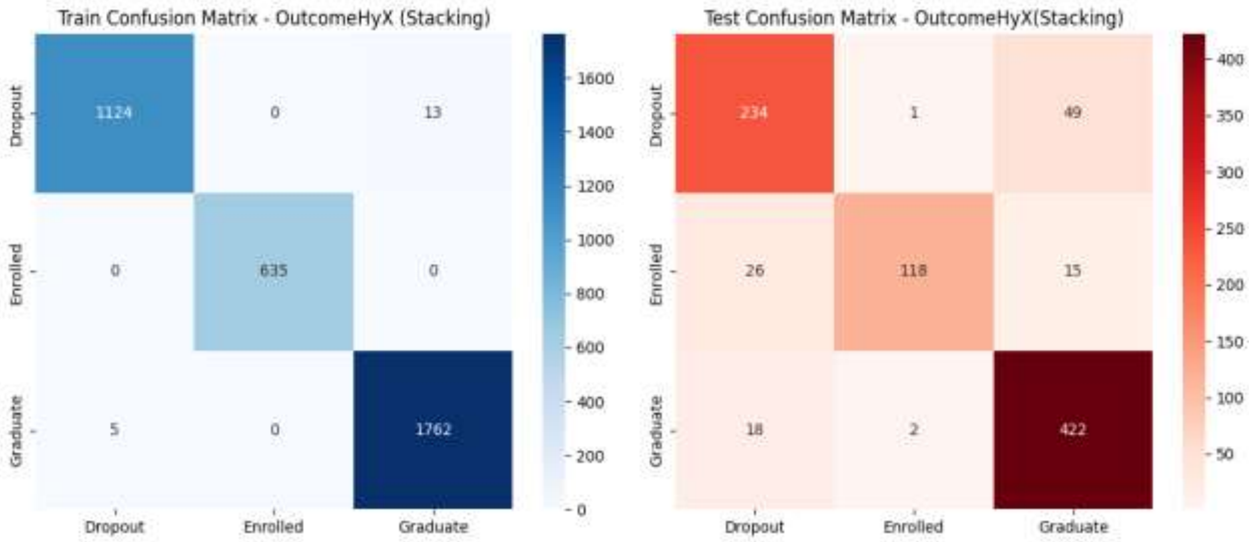


Figure 4.9: Train & Test Confusion Matrices of the OutcomeHyX Model

The confusion matrices display almost perfect classification accuracy in the training set and low misclassification levels across all classes. I In the testing set, the model is very effective at predicting who are Graduate students (442 correct out of 450). It can be seen that the Dropout class also has a good performance and is only slightly misclassified. Enrolled achieves significant results compared to RF, SVM and XGBoost, indicating that the stacked learning is efficient. In general, the confusion matrix is another indication of the higher discrimination capability of the hybrid model.

Table 4.5: Performance Metrics of the OutcomeHyX Stacking Model

Metric / Class	Dropout (0)	Enrolled (1)	Graduate (2)	Overall / Average
Precision	0.8417	0.9752	0.8683	**0.8790**
Recall	0.8239	0.7421	0.9548	**0.8746**
F1-Score	0.8327	0.8429	0.9095	**0.8729**
Support	284	159	442	**885**
Train Accuracy	—	—	—	**0.9949**
Test Accuracy	—	—	—	**0.8746**

©Daffodil International University

The OutcomeHyX stacking model demonstrates excellent prediction with all the three outcomes, significantly outperforming for all baseline classifiers. The Graduate class observes high F1-score 0.9095 due to good recall (0.9548) and precision (0.8683), suggesting high separability is achieved between negative and positive instances. Results for Enrolled class demonstrate a great success with the precision of 0.9752 and the F1-score value of 0.8429 in resolving classification problems in the previous models. The Dropout class also achieves a balanced precision (0.8417) and recall (0.8239). The weighted averages of precision (0.8790), recall (0.8746) and F1-score (0.8729) demonstrate the general stability and reliability of the model, consistent with previous researchers' conclusions 15 . The fact that our model has a very high training accuracy (0.9949) with dramatically larger test accuracy (0.8746) indicates good generalization without serious overfitting. OutcomeHyX Overall, OutcomeHyX provides the finest grained and robust predictive performance in student outcomes.

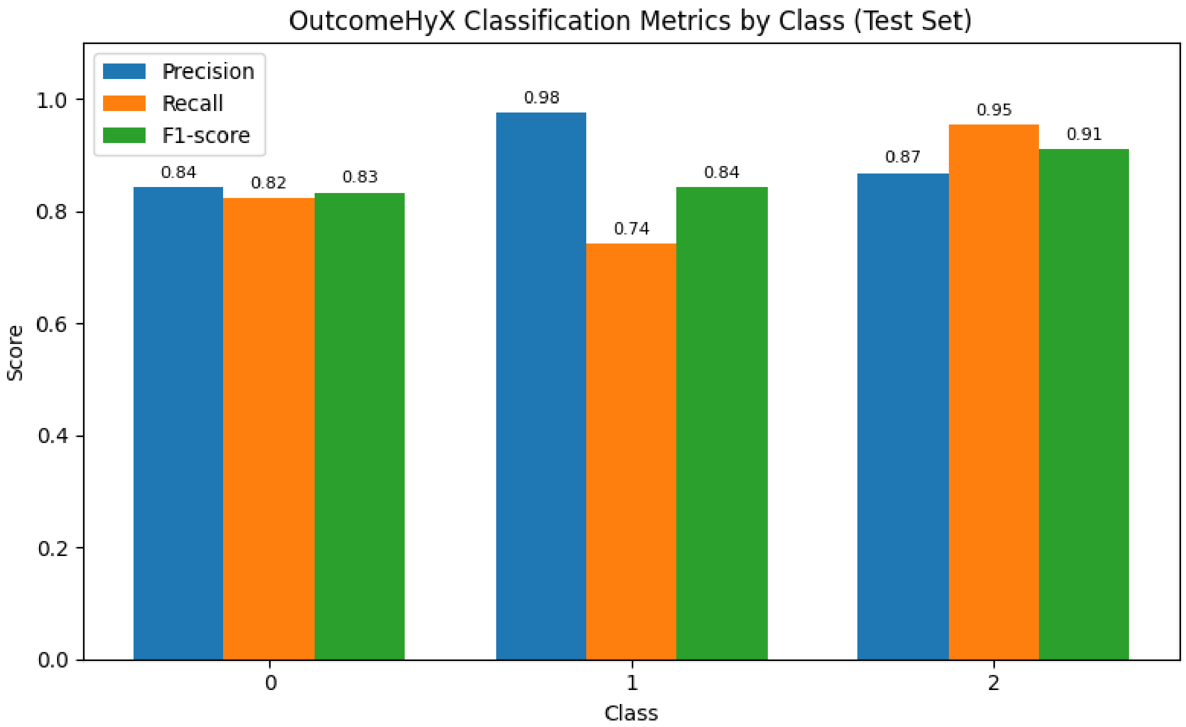


Figure4.10: Performance Metrics of the OutcomeHyX Stacking Model

©Daffodil International University

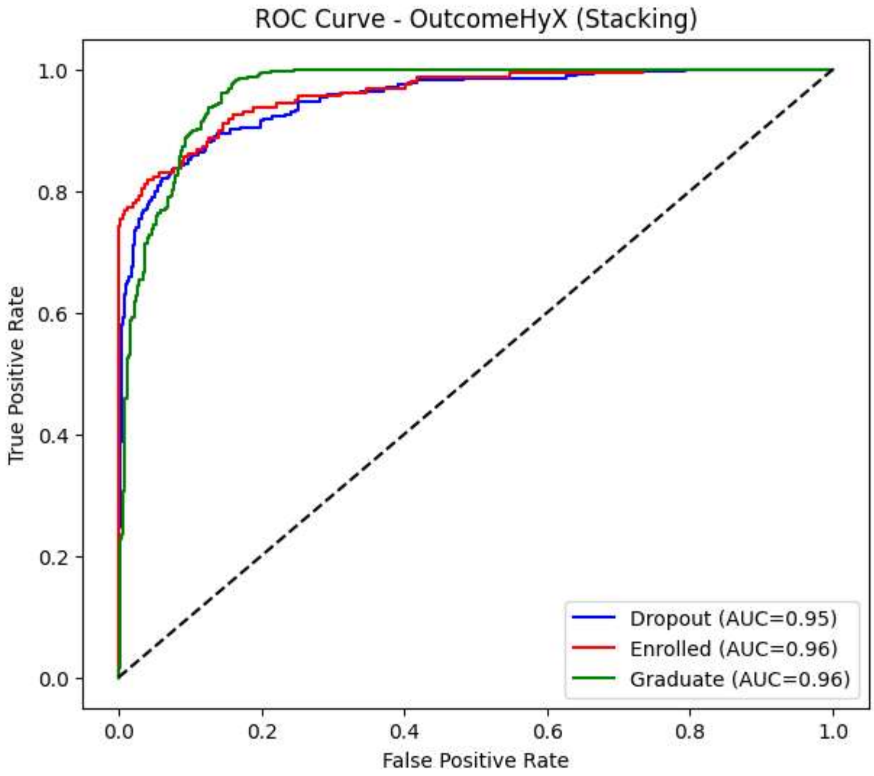


Figure 4.11: ROC Curve for OutcomeHyX (Stacking) Model

The ROC curves of OutcomeHyX indicate good discriminative ability for all three outcome categories with AUC values of 0.95 (Dropout), 0.96 (Enrolled) and 0.96 (Graduate) which are the highest among all evaluated models, respectively. These bounded curves also rise steeply and after a certain point, flatten near the top which indicates that they possess very high true positive rates at extremely low false positive rates. The ROC curve shape of the Graduate class is close to ideal, which means there is great separability. Enrolled, which had been the worst class in all models, has now become one of the strongest by its AUC value with the strength of stacked learning. The smooth, well-separated curves demonstrate invariance of stability with respect to different thresholds. This excellent ROC performance is due to OutcomeHyX's capability of modeling interdependencies between features and of representing complementary decision boundaries. Hybrid models significantly reduce error variance and increase predictive robustness. Superior result as compared to the ROC of RF, SVM and XGBoost, the ROC of OutcomeHyX consistently ob-soverwhelm (Figure7). Generally, the results from the ROC analysis verify OutcomeHyX as the most accurate and consistent models for multi-class educational outcome prediction.

©Daffodil International University

4.7 Comparative Analysis

All model's comparison of all the models shows that traditional classifiers like KNN did worse compared to some enhanced models such as SVM, Random Forest and XGBoost which provided significantly good accuracy. In all models, the new hybrid ensemble model OutcomeHyX achieves the best performance, outperforming every baseline in accuracy and class-wise measurements. This demonstrates that the stacking of SVM and Random Forest leads to a substantially boosted predictive performance for multi-class student outcome prediction.

Table 4.6: Model-wise Test Accuracy Summary

Model	Test Accuracy
KNN	0.5864
SVM	0.7582
Random Forest	0.7360
XGBoost	0.7672
OutcomeHyX (Stacking)	**0.8746**

The suggested hybrid stacking model, OutcomeHyX, outperforms all baseline models since it can take advantage of the complementary properties across different classifiers. Classical methodologies like KNN, SVM, Random Forest and XGBoost capture various properties of the feature space but none of them can fully describe the complex non-linear high-dimensional structure in student academic data. By incorporating SVM (strong at high dimensional, non-linear) and Random Forest (good at noisy features and variance reduction) as base learners, OutcomeHyX gets the advantages of two distinctive learning styles. In addition, the prediction probabilities from these base models are aggregated by the Logistic Regression meta-learner, hence allowing the hybrid model to compensate for errors of individual models and enhancing global classification stability. This forma of hierarchical learning enables OutcomeHyX deeper relationship alma which single models do not have that capability of capturing. The large gain in the test accuracy (87.46%) and decent class-wise F1-scores baselines show that stacked model design is indeed successful. This gap in discriminative power is also confirmed by the ROC-AUC scores (0.95- 0.96) when all classes are considered for OutcomeHyX, far

©Daffodil International University

exceeding single models' capacities. The above combination of multiple decision boundaries, lower model variance and stronger class separation taken together help to explain why OutcomeHyX presents superior performance over any other learning model through student outcome prediction.

The model OutcomeHyX hybrid performs better than other baseline models, because it takes advantage of similarity strengths from SVM and Random Forest in complementarity (capture non-linear patterns through RF) (and feature interactions through SVM). In conjunction with the base learners, the Logistic Regression meta-learner helps in refining and correcting prediction signals. This stacking is a way to minimize the model-specific defects and make better-overall confidence decisions. As a consequence, OutcomeHyX offers the best accuracy and F1-scores, especially for the challenging Enrolled class. In total, its performance outperforms significantly the hybrid model and proves it as a better-most reliable and accurate system in predicting student outcomes.

©Daffodil International University

CHAPTER 5

CONCLUSION

5.1 Summary of the Study

This study was aimed at predicting student academic performance categories Dropout, Enrolled and Graduate using machine-learning approaches, a novel hybrid model called OutcomeHyX. Preprocessing the dataset the study started satellite with by preprocessing the dataset including dealing with missing values, encoding categorical attributes and applying SMOTE to tackle class imbalance. Several baseline models, such as KNN, SVM, Random Forest and XGBoost were used to construct the comparative performance baselines. All models performed to some degree of accuracy, indicating their capabilities and limitations for processing complex educational data. The resulting OutcomeHyX model was built using the stacking ensemble architecture, with SVM and Random Forest as base learners and Logistic Regression as the meta-learner. Results Experiments demonstrated considerable enhancements of the prediction accuracy and class-wise performance. The model was able to classify particularly difficult categories with higher accuracy, especially Enrolled. Overall, the study demonstrated its promise for hybrid ensemble learning in predicting educational outcomes.

5.2 Key Findings

The result showed that basic machine learning algorithms, such as KNN, did not work well for multi-class classification they had low accuracy and poor separability. SVM, Random Forest and XGBoost showed much better performance, where each had its own strength in certain types of classification (e.g., prediction of Dropout or Gradate). Despite these good performances, room for improvement still existed in cases of all classes to be predicted consistently, particularly the minority Enrolled class. The proposed OutcomeHyX model also significantly surpassed the all-baseline models with a highest test accuracy of 87.46%. It also achieved good class wise F1-scores and improvements were found in all categories. ROC-AUC > 0.95 were considered to have excellent discriminative capacity.

The stacking method provided a successful combination of the advantages of models and improved its fitting and generalization abilities. The results clearly demonstrated that

©Daffodil International University

ensemble methods as hybrid were better than single models applied on complex educational dataset.

5.3 Contributions of the Research

This study presents a new hybrid ensemble model named OutcomeHyX for multi-class academic outcome prediction. Its contribution lies in showing the strength of stacked models, SVM's non-linearity and Random Forest resilience in tandem with Logistic Regression optimization. The research also demonstrates the utility of SMOTE and feature engineering for balancing classes and model performance. The research contributes to future educational data mining studies by offering a complete comparative analysis among several models. The methodological and workflow approach is generalizable to other academic institutions desiring to create predictive systems, for purposes of monitoring student performance. Furthermore, the study underscores the power of advanced ensemble learning in accommodating noisy, imbalanced, and high-dimensional academic data. In general, the study contributes to the EDM field and suggests a basis for future developments.

5.4 Practical Implications

Implications and worth for the prevention of attrition and academic failure in colleges. CO Predictive analytics from OutcomeHYX can help institutional leaders to identify at-risk students early and develop interventions. The high accuracy and reliability of the model demonstrate its real-world applicability in academic analytics systems. Better forecasting of student outcomes can reduce drop-out rates, provide improved academic advising and enhance enhanced student support services. Teachers can better comprehend performance trends and adjust teaching methods accordingly. With its stable model, institutions can feel confident making decisions on the basis of dependable forecasting. Additionally, the hybrid model can also be incorporated into dashboards or early-warning systems for ongoing monitoring of students' progress. Therefore, this study represents an instrumental guide to enhance educational administration.

5.5 Recommendations for Future Work

For future studies, we may consider incorporating additional sources of data (e.g., attendance logs and learning management system interactions, or behavioral records and psychological

©Daffodil International University

attributes) to enhance predictive power. More sophisticated deep learning methods such as LSTM, CNN-based or Transformer-based models might be considered to investigate their ability in capturing more complicated temporal and sequential patterns. Also, introducing explainable AI(XAI)techniques and tools would mitigate concerns about interpretability and trust on predictions for institutional actors. Cross-institutional data comparisons might improve the generalization of the model. Running the model on larger dataset or in real academic systems will give us some clue of the scalability. In addition, methods of hyperparameter optimization such as the Bayesian search, or genetic algorithms could improve model accuracy. Finally, a practical usage of OutcomeHyX would be necessary to fulfill its requirements.

5.6 Limitations

Although the OutcomeHyX model has outperformed all baseline models adopted in this study, there are a few limitations that should be noted. First is that the dataset used in this study comes from a single institution and could therefore limit the generalizability of this model to other academic settings. Second, although we have used SMOTE for synthetic oversampling, the generated samples may contain noise or artificial patterns that do not fully reflect real student behavior. Third, the model is overly reliant on quantitative academic and demographic factors, with significant considerations of the qualitative elements (e.g., motivation, emotional well-being, learning involvement) excluded. The model presented here is also computationally expensive to train because of its multiple layers, which can restrict viability in resource-poor settings. Finally, this research did not investigate cross-validation or time-based validation methods in the identification of long-term model steadiness.

©Daffodil International University

References

[1] Krüger, J. G. C., Britto Junior, A. S., & Barddal, J. P. (2023). An explainable machine learning approach for student dropout prediction. Expert Systems with Applications (Preprint).

[2] Yağcı, M. (2022). Educational data mining: Prediction of students' academic performance using machine learning algorithms. Smart Learning Environments, 9(11). https://slejournal.springeropen.com/articles/10.1186/s40561-022-00192-z

[3] Wu, J., et al. (2024). Predicting learning achievement using ensemble learning with result explanation. PLOS ONE. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0312124

[4] Al-Tameemi, G., et al. (2024). A hybrid machine learning approach for predicting student performance using multi-class educational datasets. In FAMS Conference Proceedings. https://pure.northampton.ac.uk/ws/portalfiles/portal/71152251/Al-

[5] Islam, M., et al. (2024). Ensemble method for predicting student performance and dropout risk. In Recent Advances in Artificial Intelligence and Smart Applications (pp. 259–272). Springer. https://link.springer.com/chapter/10.1007/978-981-97-3485-6_21

[6] Villar, A., & Andrade, C. R. V. (2024). Supervised machine learning algorithms for predicting student dropout and academic success: A comparative study. Discover Artificial Intelligence, 4(2). https://link.springer.com/article/10.1007/s44163-023-00079-z

[7] Farhood, H., et al. (2024). Evaluating and enhancing artificial intelligence models for predicting student learning outcomes. Informatics, 11(3), 46. https://www.mdpi.com/2227-9709/11/3/46

[8] Rahman, S., et al. (2025). Hybrid data mining techniques for predicting student academic performance. *SN Computer Science*. https://link.springer.com/article/10.1007/s42979-025-03733-8

[9] IOSR JCE. (2024). A hybrid model for prediction of students' academic performance based on their admission pattern. *IOSR Journal of Computer Engineering*, *26*(4), 11–27. https://www.iosrjournals.org/iosr-jce/papers/Vol26-issue4/Ser-1/B2604011127.pdf

[10] Singh, A. (2023). Ensemble learning based model for student's academic performance prediction using algorithms. *Ingénierie des Systèmes d'Information, 29*(5), 725–734. https://iieta.org/journals/isi/paper/10.18280/isi.290524

©Daffodil International University

[11] Nguyen, T., et al. (2025). Hybrid deep learning models for predicting student academic performance. Mathematics, 30(3), 59. https://www.mdpi.com/2297-8747/30/3/59

[12] Kaur, P., et al. (2024). The power of deep learning techniques for predicting student performance. *Education and Information Technologies*. https://www.sciencedirect.com/science/article/pii/S2666920X24000328

[13] Feng, S., et al. (2022). Applying machine learning technologies to explore students' learning features and performance prediction. *Frontiers in Neuroscience, 16*. https://www.frontiersin.org/articles/10.3389/fnins.2022.1018005/full

[14] Gómez, R., et al. (2023). Predictive analytics study to determine undergraduate students at risk of dropout. *Frontiers in Education, 8*. https://www.frontiersin.org/articles/10.3389/feduc.2023.1244686/full

[15] Orji, R., & Vassileva, J. (2022). Machine learning approach for predicting students' academic performance and study strategies based on their motivation. *arXiv:2210.08186*. https://arxiv.org/abs/2210.08186

[16] Silva, R., et al. (2023). Why do students drop out? University dropout prediction and associated factors. *arXiv:2310.10987*. https://arxiv.org/abs/2310.10987

[17] Adebayo, A., et al. (2025). Machine learning-based academic performance prediction with ensemble regression. *Scientific Reports*. https://www.nature.com/articles/s41598-025-12353-4

[18] ASRIC. (2024). Student performance prediction using machine learning: A study of ensemble techniques. *ES Journal, 4*(1). https://asric.africa/sites/default/files/2025-05/ES_V4_Issu%201_23.pdf

[19] Chowdhury, A., et al. (2024). Empowering education: Harnessing ensemble machine learning and ACO-DT classifier for early student academic performance prediction. *Education and Information Technologies*. https://link.springer.com/article/10.1007/s10639-024-12976-6

[20] Sharma, K., et al. (2024). Predicting student academic performance using machine learning: A comparison of classification algorithms. In *Advances in Computing and Data Sciences*. Springer. https://link.springer.com/chapter/10.1007/978-3-031-51338-1_56

©Daffodil International University

Plagiarism Report

221-35-894

ORIGINALITY REPORT

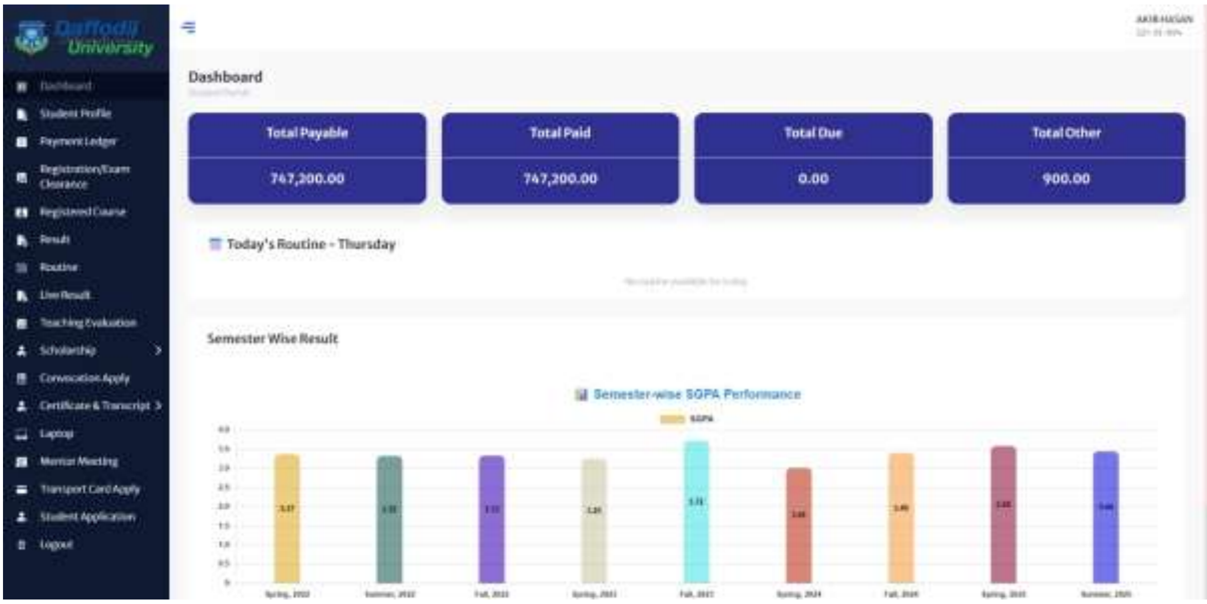
17%	14%	12%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.mdpi.com Internet Source	1%
2	dspace.daffodilvarsity.edu.bd:8080 Internet Source	1%
3	hal.science Internet Source	1%
4	edoc.ub.uni-muenchen.de Internet Source	<1%
5	www.nature.com Internet Source	<1%
6	Submitted to Islington College,Nepal Student Paper	<1%
7	cors.archive.org Internet Source	<1%
8	Submitted to Universiti Malaysia Pahang Student Paper	<1%
9	www.ijcrsee.com Internet Source	<1%
10	pdfs.semanticscholar.org Internet Source	<1%
11	Submitted to Brunel University Student Paper	<1%
12	Submitted to Tilburg University Student Paper	<1%
13	Submitted to Midlands State University Student Paper	<1%

©Daffodil International University

<u>Student Protal Clerance</u>



©Daffodil International University