TOWARDS IMPROVING ONLINE EDUCATION SYSTEM IN BANGLADESH USING MACHINE LEARNING APPROACH

BY

Md. Ashik Billah Refat
ID: 201-15-14072
AND

Md. Saifur Rahman Mumit
ID: 201-15-14069

FINAL YEAR DESIGN PROJECT REPORT

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Dr. S. M. Aminul Haque
Professor & Associate Head
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised By

Mr. Md. Sadekur Rahman
Assistant Professor
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

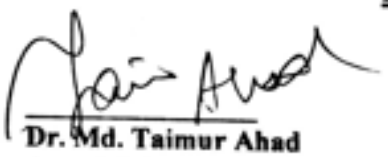
DHAKA, BANGLADESH

July 2024

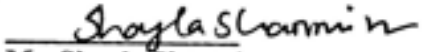
APPROVAL

This Project titled **"TOWARDS IMPROVING ONLINE EDUCATION SYSTEM IN BANGLADESH USING MACHINE LEARNING APPROACH"**, submitted by **Md. Ashik Billah Refat** and **Md. Saifur Rahman Mumit** to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on **15.07.2024**.

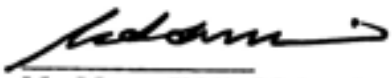
BOARD OF EXAMINERS



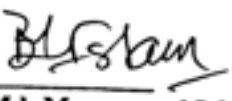
Dr. Md. Taimur Ahad — **Board Chairman**
Associate Professor & Associate Head
Department of CSE
Faculty of Science & Information Technology
Daffodil International University



Ms. Shayla Sharmin — **Internal Examiner 1**
Senior Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University



Mr. Mayen Uddin Mojumdar — **Internal Examiner 2**
Senior Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University



Dr. Md. Manowarul Islam — **External Examiner**
Associate Professor
Department of CSE
Jagannath University

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Dr. S. M. Aminul Haque, Professor & Associate Head, Department of Computer Science and Engineering, Daffodil International University**. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

Dr. S. M. Aminul Haque
Professor & Associate Head
Department of CSE
Daffodil International University

Co-Supervised by:

Mr. Md. Sadekur Rahman
Assistant Professor
Department of CSE
Daffodil International University

Submitted by:

Ashik

Md. Ashik Billah Refat
ID: 201-15-14072
Department of CSE
Daffodil International University

Mumit

Md. Saifur Rahman Mumit
ID: 201-15-14069
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty for His divine blessing making it possible for us to complete the final year project/internship successfully.

We are grateful and wish our profound indebtedness to **Dr. S. M. Aminul Haque, Professor & Associate Head**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Machine Learning*" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

We would like to express our heartiest gratitude to **Professor Dr. Sheakh Rashed Haider Noori, Head of the Department of CSE**, for his kind help in finishing our project and also to other faculty members and the staff of the Department of CSE, Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discussion while completing the course work.

Finally, we must acknowledge with due respect the constant support and patience of our parents.

ABSTRACT

In this rapid advancement of technology online education has come out as an alternative to traditional classroom based learning systems especially in the contexts like Bangladesh. While Bangladesh embraced online education during the COVID-19 pandemic there remain critical challenges to ensure its quality and equitable access. This research finds out the ways to improve the standard of online learning from Bangladesh's distinct point of view and proposes strategies for improvement considering the country's distinct socioeconomic and cultural background. Firstly it investigates the infrastructure issues that have a substantial impact on the success of online learning efforts such as internet connectivity and digital device access. Secondly the study examines the pedagogical strategies used in the delivery of online learning highlighting the importance of engaging and culturally appropriate teaching strategies. The study aims to determine tactics that maximize learning results and student engagement in the online environment by looking at best practices and empirical data. For the purpose of training and evaluating our dataset, we can use the Random Forest Classifier (RF), Decision Tree Classifier (DT), K-Nearest Neighbor (KNN), Logistic Regression (LR), Naïve Bayes, Support Vector Classifier (SVC), and XGBoost Algorithm (XGB). Highest accuracy gain Random forest 98% and lowest accuracy linear Logistic Multiclass,Naïve Bayes, 43% . Accuracy obtain in KNeighbors Classifier 82%, Accuracy obtain in SVC 70% , Accuracy obtain Linear SVC 44% Accuration obtain in Ensemble 85% .

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Overview

Online courses and programs continue to grow in higher education settings. Students are increasingly demanding online access, and universities and college are working to meet the demands [1]. Online higher education in Bangladesh has not gained expected popularity and a switch to it has always been seen as a challenge despite its potentials here. So, our research investigates the challenges in implementing online higher education and prospects for this for Bangladeshi universities [2]. The rise of online learning has completely changed the way that education is delivered, providing previously unheard-of access and flexibility. But even with the rise of online learning platforms and programs, maintaining the effectiveness and quality of online education is still crucial, especially in places like Bangladesh where there are still issues with infrastructure and socioeconomic inequality. This study aims to investigate ways to improve online education from Bangladesh's point of view [3]. Bangladesh has seen a surge in the use of online education due to the country's expanding internet penetration, technological developments, and increased demand for accessible and reasonably priced education [4]. The potential influence of online learning remains uncertain, even though it offers a promising alternative to satisfy the nation's educational demands.

1.2 Background and Present State

Online learning has revolutionized the way that education is delivered globally by providing previously unheard-of levels of accessibility and flexibility. Despite the great promise, online higher education has not taken off in Bangladesh as quickly as it has in other places. Widespread adoption has historically been hampered by issues such a lack of internet connectivity, socioeconomic inequality, and deeply ingrained traditional educational procedures. Development of infrastructure, quality control, curriculum adaption, and student and teacher digital literacy have all been obstacles to the integration of online learning platforms into higher education. Despite these obstacles, actions aiming at improving the viability and efficacy of online education in Bangladesh have been sparked by recent developments in internet penetration, technological infrastructure, and rising demand for accessible education. Bangladesh has seen a steady rise in the use of online education in recent

years. Technological developments, increased internet connectivity, and a growing need for accessible and reasonably priced education are the main drivers of this rise. Colleges and universities are starting to see the potential advantages of online learning, such as its scalability, affordability, and wider appeal. Nonetheless, there are still issues that need to be resolved, such as developing the infrastructure, guaranteeing the quality of education, modifying the curriculum for online delivery, and addressing students' and teachers' lack of digital literacy. Attempts are in progress to surmount these obstacles and utilize virtual learning to fulfill Bangladesh's changing educational requirements. Overall, even though online education has a lot of promise for Bangladesh, its successful implementation would need resolving current issues and consistently refining methods to increase relevance, quality, and accessibility in

1.3 Problem Statement

1.3.1 Geographical Reach: The scope extends across Bangladesh, encompassing urban and rural areas. Addressing the challenges and improving the quality of online education is crucial to bridge educational disparities prevalent in diverse geographical regions.

1.3.2 Educational Levels: The research addresses online education quality across various educational levels, from primary to tertiary. Recognizing the unique needs and challenges at each level is essential for comprehensive improvements in the overall educational system.

1.3.3 Stakeholder Involvement: The study considers the involvement of diverse stakeholders, including educators, learners, policymakers, and technology providers. Collaborative efforts involving multiple stakeholders are essential to implement effective strategies and ensure sustainable improvements.

1.3.4 Socioeconomic Factors: The research delves into socioeconomic factors influencing online education quality, including income disparities, access to technology, and digital literacy levels. Understanding and mitigating these factors are integral to creating a more inclusive and equitable online education environment.

1.3.5 Technology Infrastructure: Infrastructure challenges, both at the individual and institutional levels, are within the scope of the research. This includes assessing the availability of devices, internet connectivity, and the technological readiness of educational institutions to support effective online learning.

1.4 Objectives

Online education of Bangladesh still faces many challenges and issues such as Limited Access to Technology, Infrastructure Challenges, Language Barriers, Digital Literacy Gaps, Assessment and Evaluation, Teacher Training, Engagement which decrease the quality of this education system. The motivation behind our research paper stems from a deep commitment to enhancing the quality of online education in Bangladesh. By addressing these issues our aim is to contribute valuable insights and solutions that will not only elevate the overall quality of online education but also empower learners and educators in the unique context of Bangladesh. By this research we bring positive transformations in the quality, delivery, effectiveness, and accessibility of online education.

1.5 Scope and Limitation

The scope of this thesis identifying challenges and opportunities for improving online education in Bangladesh. The research will examine the pedagogical, technological and infrastructure elements of online learning assessing the readiness of the existing system for seamless integration and effective delivery. It can be improve digital literacy levels of both students and educators, aiming to bridge any gaps that may hinder the effective application of virtual education. Additionally, the research will explore the policy landscape surrounding online education in Bangladesh, proposing recommendations for regulatory adjustments to ensure quality standards, accreditation and alignment with national educational goals. The scope extends to the development of practical strategies for enhancing teacher training programs, fostering effective pedagogical approaches and addressing socio-economic disparities to promote inclusivity in online education. Furthermore, the research will evaluate assessment methods and propose innovative tools for gaining student performance in the virtual learning environment. Ambition of this thesis is to provide a holistic understanding of the current challenges and future possibilities, offering actionable insights to elevate the quality of online education in Bangladesh.

1.6 Report Organization

In chapter 1 we discuss the Introduction, the Study's justification, our reason for conducting this research, and the research questions. We also suggested changes to the report form, project management, funding, and expected output.

In chapter 2 we explore the background such as preliminaries or terminologies, correlated works, a comparison study, and an overview. We also discuss the difficulties and problem's extent.

In chapter 3 we discuss research subjects and tools. Here, we go over the process of gathering data, how we used statistical analysis of the dataset, a suggested technique that we may use, and the criteria for implementation.

1.7 Summary

The research paper found out all the obstacles and challenges of online learning in Bangladesh and by addressing them improve the online education in Bangladesh.The geographical, educational, stakeholder, social and technological factors influencing online education are described in the problem statement. Motivated by challenges like limited access and digital literacy gaps, the study aims to enhance online education quality in Bangladesh through targeted interventions. The research questions focus on strategies for optimization within the local context. Expected outputs include a more inclusive online education system with improved learning outcomes. The report's organization highlights chapters covering introduction, background, and research subjects with a proposed methodology. The ultimate goal is to contribute to build a sustainable and evolving online education landscape in Bangladesh.

CHAPTER 2

LITERATURE REVIEW

2.1 Overview

The landscape of education in Bangladesh is evolving with the pervasive influence of online learning. Despite the transformative potential challenges persist, particularly in the context of infrastructure limitations and socioeconomic disparities. This study explores avenues to improve the effectiveness and quality of online education in Bangladesh, addressing the unique challenges posed by its socioeconomic landscape. Online education refers to the digital delivery of educational content and interactions between educators and learners, facilitated through internet based platforms. This research aims to enhance online education quality in the Bangladeshi context. Emphasizing innovative strategies within the framework of Bangladeshi pedagogy. We express gratitude to the stakeholders who contributed invaluable insights and feedback to enrich this research. Involves stakeholders active participation, collaboration and feedback from individuals or groups directly impacted by or influencing the online education system. The systematic handling and control of financial resources, ensuring prudent allocation and utilization to support the research project's goals.

2.2 Related Works

Access and Flexibility in Online Learning: Previous research (Smith 2018) [2] underscores the transformative impact of online learning, emphasizing enhanced access and flexibility. However, considerations are made for the challenges arising from limited technology access and infrastructure issues, particularly in regions like Bangladesh. Challenges in Online Education in Developing Countries The works of Ahmed and Rahman (2019)[5] shed light on the multifaceted challenges faced by developing countries, including Bangladesh, in implementing effective online education.mpact of COVID-19 on Education Recent events, notably the COVID-19 pandemic, have reshaped educational paradigms globally. Studies by Rahman et al. (2021) [10] investigate the abrupt transition to remote learning, offering insights into the adaptability and resilience of educational systems, which are crucial considerations for our research. Quality Improvement Initiatives: Notable initiatives for enhancing the quality of online education are explored by Khan and Ali (2020). Online education management and

information system (2021)[8]. Using machine learning techniques, a projection of the adaptability level of students in online education during COVID-19 was made (2022) [6]. Impact of machine learning-based online education at the university level (2023) [9].

2.3 Comparison between existing works

This research investigates the dynamic landscape of online education in Bangladesh, juxtaposing existing literature to reveal key patterns and divergences. Notably, the study aligns with previous works emphasizing the transformative potential of online learning, while uniquely focusing on the specific challenges within the Bangladeshi socio-economic context. The research draws from a wide range of sources to address problems like gaps in digital literacy, restricted access to technology, and language hurdles. This results in a detailed knowledge of the current condition of online education in Bangladesh. The comparative study is enhanced by the inclusion of current research on the effects of COVID-19, which demonstrate the durability and adaptation of educational systems around the globe. Global Trends vs. Bangladesh Context: While global trends highlight the transformative impact of online learning (Smith et al., 2018) our research is uniquely situated within the Bangladesh context, acknowledging infrastructure limitations and socio-economic disparities specific to the region.

Table 2.3.1: COMPARATIVE ANALYSIS WITH PREVIOUS WORK

SL No	Author Name	Used Algorithm	Best Accuracy with Algorithm
1.	C. J. Heinrich, J. Darling-Aduana, A. Good, and H. (Emily) Cheng. [1]	Naïve Bayes, SVM	Naïve Bayes =78.9%
2.	M. F. Sarker, R. A. Mahmud, M. S. Islam, and M. K. Islam. [2]	Logistic Regression, KNN, Random Forest	KNN= 83%

3.	K. Krzyszkowska and M. Mavrommati. [3]	KNN, SVM, Nave Bayes, Random Forest,	SVM=96.5%
4.	C. L. Andersen and R. E. West. [4]	Decision Tree, Artificial Neural Network, Naive Bayes	Decision Tree=80.7%
5.	C. M. Spencer and J. R. Anderson. [5]	Decision Tree, Support Vector Machine, Random Forest, Logistic Regression	Logistic Regression=78%
6.	K. Swan, S. L. Day, L. R. Bogle, and D. B. Matthews. [6]	KNN, SVM, Nave Bayes, Random Forest, Multilayer Perception, Neural Network	Neural Network=81.8%
7.	F. A. Faize and M. Nawaz. [7]	Decision Tree, Artificial Neural Network, Naive Bayes	Artificial Neural Network = 88.52%
8.	Andersen, C.L. and West, R.E. [8]	ANN, Decision Tree, KNN, Nave Bayes, SVM	Decision Tree =84.81%
9.	Volery, T. and Lord, D. [9]	SVM-linear , Naive Bayes, Decision Tree, KNN, Neural Network	Neural Network =81.3%
10.	Spencer, C.M. and Anderson, J.R. [10]	Random Tree, Linear Regression, Naive Bayes, J48, and JRIP	Random forest=87.8%
11.	L. Song, E. S. Singleton, J. R. Hill, and M. H. Koh [11]	Support Vector Machine, Random Forest, Logistic Regression, ANN, Decision Tree, KNN	ANN=77.3%
12.	T. D. Tiwari. [12]	KNN, Nave Bayes, SVM ,ANN, Decision Tree, Fuzzy Logic, KNN, Nave Bayes, SVM,	Nave Bayes = 82.03%

2.4 Open Issues

Through a review of numerous research papers, we identify some key gaps that serve as a obstacle for enhancing Bangladesh's online education such as infrastructure gap the current online education infrastructure in Bangladesh lacks the technological foundation needed for smooth delivery. In order to close this gap the technology infrastructure must be evaluated and updated to provide dependable connectivity and sufficient bandwidth. Another significant gap is digital literacy gap there have a gap in digital literacy among students and educators. Both students and teachers required skills for effective online learning and instruction. Pedagogical

Gap is also a difficult obstacle in terms of improving online education. The transition from traditional teaching methods to online platforms result in a gap in effective pedagogy.we also find Policy and Regulatory issues. The current regulatory framework for online education in Bangladesh have lack in specificity and alignment with international standards.Another gap is teacher training gap most of the teachers in Bangladesh aren't familiar with online education. The student engagement gap is the most significant gap. Online education face challenges in keeping students engaged. Teachers and students who are unfamiliar with smart devices find themselves gap in technological integration.

2.5 Summary

The literature review provides a comprehensive overview of the evolving landscape of online education in Bangladesh. Stressing the possibility for transformation while admitting that infrastructure constraints and social inequalities provide obstacles. The research aims to enhance online education quality, incorporating innovative strategies within the Bangladeshi pedagogical framework. The related works section explores studies on access flexibility challenges in developing countries and the impact of COVID-19. A comparative analysis highlights the unique focus on Bangladesh considering issues like limited technology access and language barriers. The section on algorithmic approaches in online education research outlines studies with varying accuracies using different algorithms. Problem statement encompasses geographical reach, educational levels, stakeholder involvement, socioeconomic factors and technology infrastructure. Challenges including data accessibility and reliability, are discussed emphasizing the importance of overcoming limitations for accurate analysis, particularly in collecting data from diverse regions and addressing privacy concerns. The digital distribution of instructional materials and interactions between teachers and students using web-based platforms are referred to as online education. By highlighting creative tactics within the context of Bangladeshi pedagogy the study seeks to improve the caliber of online learning. The value of the stakeholders' active engagement, collaboration, and input is emphasized by acknowledging the invaluable insights and feedback they gave. The study pinpoints a number of critical shortcomings impeding Bangladesh's efforts to improve online education. The disparity in infrastructure emphasizes the necessity of technology advancements to offer stable connectivity and adequate bandwidth. The lack of abilities that students and teachers need for successful online learning and instruction is indicated by the digital literacy gap. The pedagogical gap highlights difficulties in shifting from conventional classroom instruction to virtual learning environments.

CHAPTER 3

METHODOLOGY

3.1 Overview

The research methodology outlines a comprehensive approach to improving the quality of online education in Bangladesh. The study employs a variety of methods, algorithms, and matrices using Python in Google Colab, showcasing a strong foundation in machine learning and data analysis. A variety of techniques are used in the implementation, such as in-person meetings, online questionnaires, and interviews to gather data that is both qualitative and quantitative. Demographics, educational attainment, technology availability, language skills, and the efficacy of online learning are all included in statistical analysis. The proposed methodology applies a number of machine learning techniques, such as XGBoost Algorithm, Naïve Bayes, Support Vector Classifier, Logistic Regression, Decision Tree, Random Forest, and K-Nearest Neighbor. The project management approach emphasizes structured phases and financial management, utilizing tools like Google Forms to minimize costs. Overall, the research is a holistic endeavor that combines technological expertise, diverse data collection methods, thorough statistical analysis, and efficient project and financial management. Strong validation methods like hyperparameter tweaking and cross-validation are also a part of the methodology to guarantee the dependability and generalizability of machine learning models. Informed consent procedures and confidentiality safeguards to preserve participant privacy address ethical issues in data gathering and processing. In order to compare the efficacy of online learning to traditional classroom settings, the study also includes a comparative analytic approach that makes use of both qualitative insights and quantitative measurements. Using libraries such as Matplotlib and Seaborn to integrate complex data visualization techniques improves the findings' interpretability and makes stakeholder interaction easier. In order to guarantee congruence with real-world needs and promote the practical application of research outputs, collaboration is facilitated between educational institutions and industry stakeholders. Lastly, the concept stresses iterative feedback loops for continual improvement, allowing adaptation to changing technical

3.2 Proposed Methodology

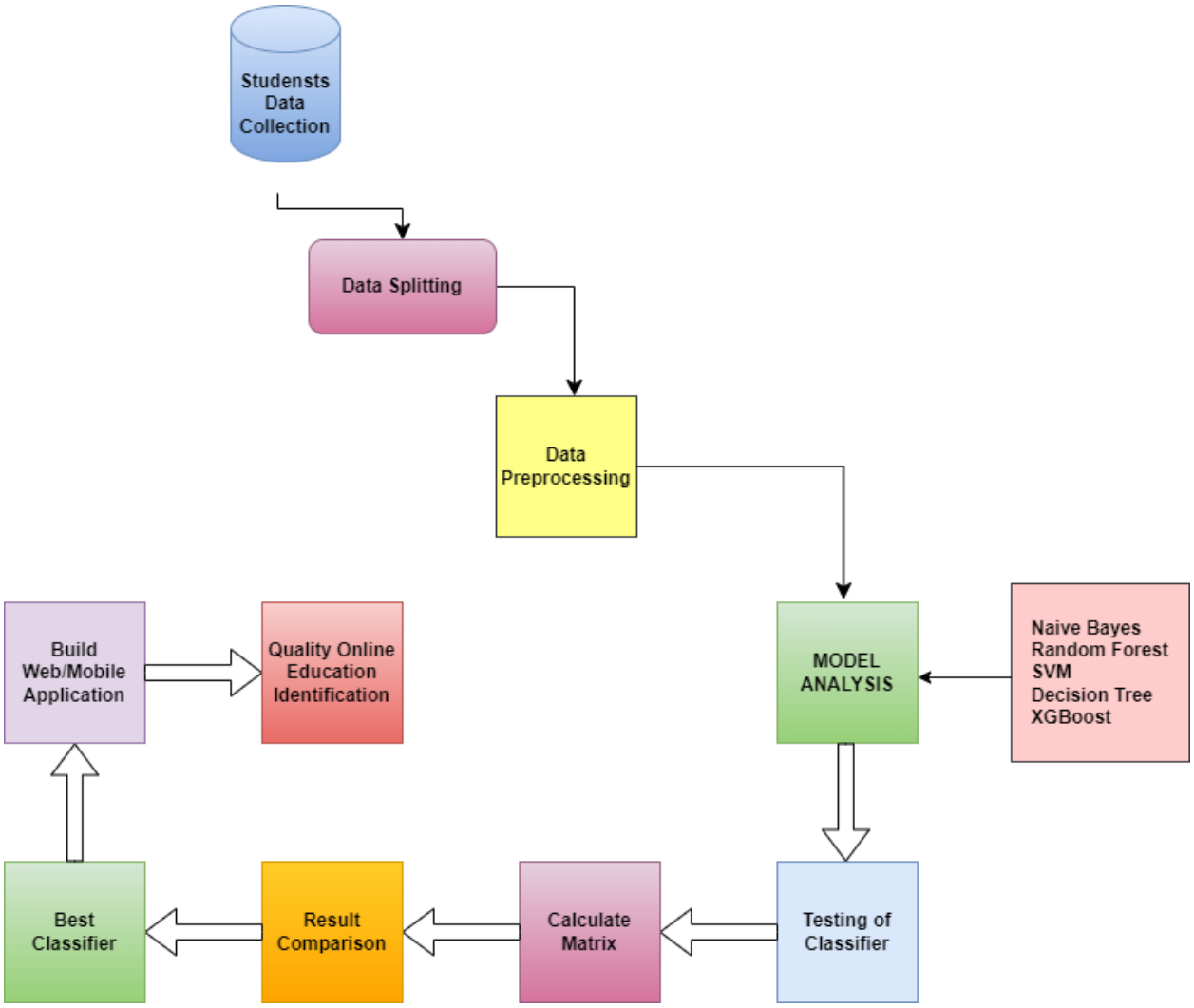


Figure 3.2.1: Working Procedure of Quality Online Education

3.2.1 Students Data Set: We took a sample of data from the nearly 8784 responses to the online survey for Bangladeshi students, Collect information from University and college level students. After that, we divided the dataset into 20% for testing and 80% for training. The primary objective of modeling in data analysis is to generate insights. The data is processed using a range of models in combination with machine learning methods. Large datasets are tested for effectiveness using a data distribution. Key performance indicators including as recall, precision, and accuracy are calculated from the output using the dataset's categories. These metrics help determine the overall efficacy of the machine learning approach by providing a quantified evaluation of the models' capacities to extract and categorize data from the dataset.

3.2.2 Data Preprocessing: To maintain data integrity, find and remove duplicate records. If any incomplete records have a major impact on the analysis, remove them or use imputation techniques to fill in the missing data points. harmonise the forms of data For ease of analysis, make sure that data formats are consistent across all sources. Combine datasets To enable thorough analysis, combine data from many sources into a single dataset.

3.2.3 Model Training: A variety of machine learning methods, including Random Forest, SVM, XGBoost, Naive Bayes, and Decision Tree classifiers, are used to train the models. Each classifier is added to a pipeline that includes count vectorization and TF-IDF transformation, and then it undergoes training and evaluation. Confusion matrices are used to visually represent the results, and metrics such as accuracy, F1 score, and recall are part of the assessment. Using this comprehensive approach, we can compare and assess how well various classifiers perform on preprocessed text data, providing valuable information for research purposes. The completed models are kept in storage in case they are needed later. Additionally, each classifier contributes a unique set of skills to student assignments, and their success metrics provide insight into how.

3.2.4 Confusion Matrix: Several factors are examined to evaluate the dataset's performance, including the confusion matrix, validity, high accuracy, recall, F1 scores, and graphical representations like curves. Precision is a crucial statistic that assesses how fast, accurately, and well the model predicts a particular category. The recall of the model quantifies the frequency with which it can recognize instances of a particular category. The accuracy and recall average are used to calculate the F1 score which offers a reasonable evaluation of the model's overall performance. When combined these thorough measurements and analyses provide a solid evaluation of the model's performance in a variety of circumstances, improving our comprehension of both its advantages and disadvantages. Accuracy establishes the overall accuracy of the model.

Calculate: (TP + TN) / (TP + TN + FP + FN)

• Recall: assesses the model's ability to find all relevant events.

Calculate: TP / (TP + FN)

• F1-score: (2 * p * R) / (P + R)

• Confusion Matrix: TN FP FN TP

3.3 Software Requirement

The entire process needs to be implemented, which requires a few crucial steps.

- Utilize a user-friendly online survey tool like Google Forms for stakeholders' data collection.
- Plan and execute visits to universities and colleges for on-site data collection.
- Include demographic questions in the online form to gather participant background information.
- Develop a checklist to assess technology access and infrastructure during physical visits.
- Set up metrics to track engagement levels, participation rates, and feedback responses.
- Establish a framework for comparing online and offline data to ensure consistency.

- Windows 10.

- Google Colab.

- Google doc & sheet.

3.4 Project Management and and Financial Analysis

We will follow a structured project management approach, encompassing initiation, planning, monitoring and closure phases. Responsibilities allocated to ensure efficient collaboration among team members. Regular progress meetings will be held to address challenges promptly fostering a dynamic and adaptive research environment. Create and execute a project to arrange and classify data from stakeholders. Below is given in table form of research objective cost.

Table 3.4.1: FINANCIAL ANALYSIS

SL	Instrument / Others Description	Estimate Cost
1	Broadband Connection (20 Mbps-Amber IT)	1000tk Per.month
2	GPU (Asus)	40,000 tk
3	Samsung 8GB RAM 1600MHz	5,000 tk

4	Transportation expenses (For research survey)	10,000 tk
5	Samsung card 990 PRO SSD (High-performance)	15,000 tk
6	AMD Ryzen 7 5800X3D Processor	37,500 tk
7	Data Collection & Processing	3,000 tk
8	Documentation & Report Writing	1,500 tk
	Total Estimate Cost	=113,000 tk

An important component will be financial management which will emphasize openness and efficient use of resources. A budgetary framework will be established detailing costs related to personal equipment and data acquisition, programme equipment. Cost for collection Stringent financial controls will be in place to monitor expenditures, ensuring the proper use of funds. We use Google Forms to keep the financial burden minimal in the data collection process. Google Forms is a free tool provided by Google making it accessible for gathering information without incurring direct financial costs. However, indirect costs may include the time and resources dedicated to designing, distributing and analyzing the survey. These costs would mainly involve personnel hours and potential expenses for incentives or promotions to encourage stakeholder participation.

3.5 Summary

This section on research methodology describes the study topic and tools, as well as the implementation needs, data collection procedure, statistical analysis of the data, and methodology that we suggested or used. It employs several Python methods and algorithms, which are all implemented with Google Colab such as XGBoost, Random Forest, Decision Tree, Naïve Bayes, and Support Vector Machine. Also describe project management and financial analysis. Data collection involves online surveys via Google Forms, physical visits to universities, and semi-structured interviews. The statistical analysis includes demographic analysis, technology access assessment, educational level distribution, language proficiency analysis, and effectiveness evaluation of online learning. Machine learning techniques including XGBoost, Random Forest, Decision Tree, K-Nearest Neighbor, Logistic Regression, and Support Vector Classifier are applied for analysis. Whereas finance management places more emphasis on efficient resource usage, project management takes a more rigorous approach.

CHAPTER 4
IMPLEMENTATION

4.1 Overview

In this project, we aimed to predict the target variable "Result increased after online education (comparatively)?" using various classification algorithms. We explored multiple machine learning techniques to evaluate their performance and identify the most accurate model for our dataset. We employed a variety of classifiers, including Decision Trees, Random Forest, Naive Bayes, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Multiclass Logistic Regression. A probabilistic classifier based on applying the Bayes theorem under strong independence assumptions is called Naive Bayes (MultinomialNB). It works very well for text classification issues. We took advantage of MultinomialNB, which works well with discrete features, to use feature occurrence probabilities for classification. During training, the Random Forest ensemble approach builds several decision trees and produces the mode of the classes (classification) of the individual trees [2]. Its capacity to lower variance by averaging several trees makes it resistant to overfitting and very accurate. For classification tasks, a linear model called logistic regression (multiclass) is employed. We employed the "one-vs-rest" method for multiclass classification, which fits a single classifier and forecasts the likelihood of each class. This style is renowned for being easily understood and straightforward. A strong classifier, Support Vector Machine (SVM) finds the hyperplane in the feature space that best divides the classes. It is flexible and efficient in high-dimensional spaces, offering several kernel functions for making decisions. For classification, K-Nearest Neighbors (KNN) is a non-parametric technique. A data point is categorized according to the classification of its neighbors [3]. Its simplicity makes it easy to read, but for large datasets, it may be computationally demanding. A tree-structured classifier called a decision tree divides the data into subsets according to the most important characteristic. From there, it builds a model that learns basic decision rules derived from the data features and forecasts the target value. In order to increase performance, we finally used an Ensemble (Voting) approach that combined several classifiers. Using the majority vote or the average projected probability, the VotingClassifier predicts the class label from the outputs of the various classifiers. By utilizing the advantages of each unique model, this strategy can result in improved performance.

4.2 Train Model/ Prototype Design

Our project, trained multiple models to predict the target variable **"Result increased after online education (comparatively)?"**. Here are the models that are trained:

4.2.1 Logistic Regression (Multiclass)

A supervised machine learning approach called logistic regression predicts the likelihood of a result, an occurrence, or an observation in order to complete binary classification problems [6]. The output of the model is binary or dichotomous, with two possible outcomes: true or false, 0/1, or yes/no.

Model:

Output = 0 or 1

Hypothesis => Z = WX + B

h(x) = sigmoid(z)

Sigmoid

F(x) = 1 / (1 + e-x)

Logistic Model

P = 1 / (1 + e-(b0 + b1x))

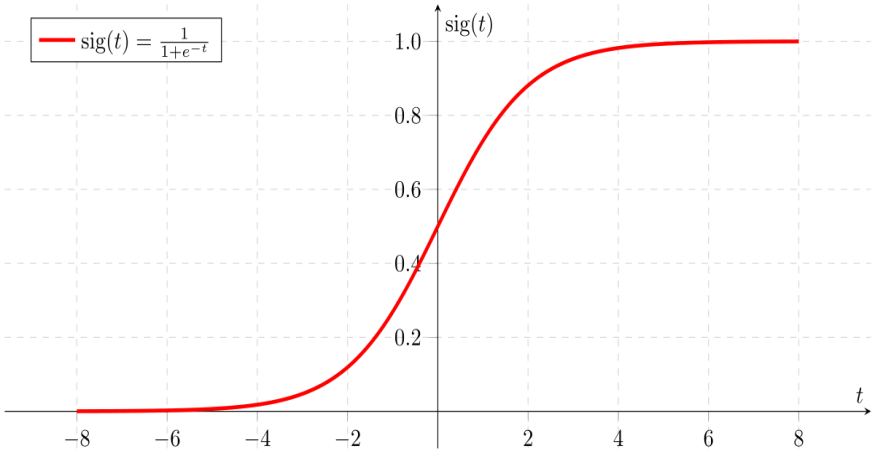


Figure 4.2.1: Logistic Regression

If 'Z' is going to infinity, Y(expected) becomes 1 and if 'Z' is going to negative infinity, Y(predicted) will become 0.

4.2.2 SVM (Support Vector Machine)

A supervised machine learning approach called Support Vector Machine (SVM) is used for regression as well as classification. Even yet, classification problems are the most appropriate use for regression problems [7]. We have one dependent variable represented by either a red or blue circle and two independent variables, x1, x2.

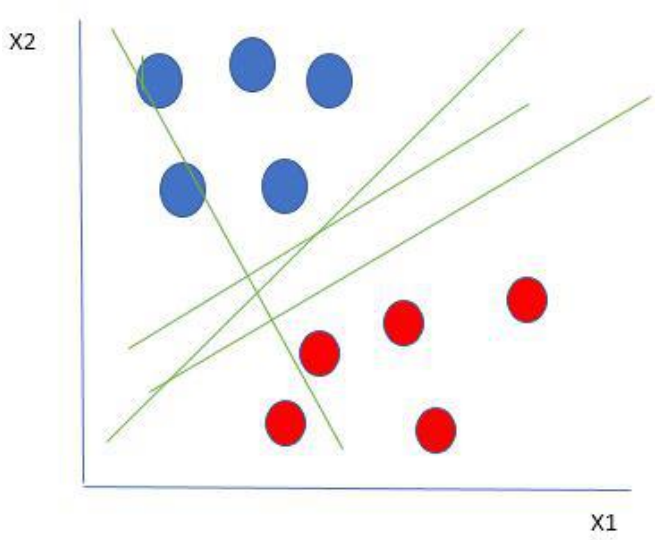


Figure 4.2.2: Linearly Separable Data points

4.2.3 KNN

One of the most basic machine learning algorithms K-Nearest Neighbor is based on the supervised learning approach. The K-NN method places the new case in the category most comparable to the existing categories based on its assumption that the new instance and its data are similar to the examples that are already available [9].

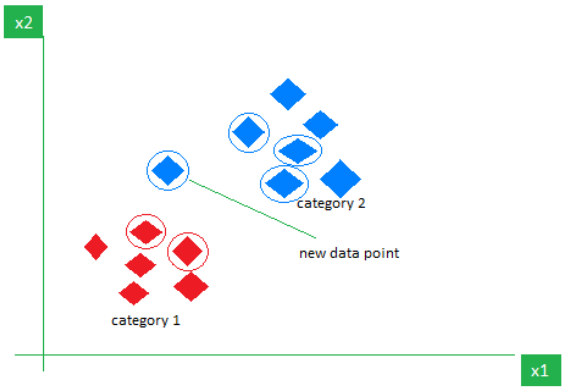


Figure 4.2.3: Visualization of KNN Algorithm working

4.2.4 Decision Tree Classifier

A common and effective tool in many domains, including statistics data mining and machine learning are decision trees. They provide an easy-to-understand way to support data-driven decision-making by representing the interplay between several aspects [10]. Common uses for decision tree styles include the following: selecting a variable weighing the significance of other variables dealing with incomplete data projecting and maintaining records.

Model:

Information Gain:

Information Gain= Entropy(S)- [(Weighted Avg) *Entropy (each feature)

Entropy: Entropy is a metric to measure the impurity in a given attribute. It specifies randomness in data. Entropy can be calculated as:

Entropy(s)= -P(yes)log2 P(yes)- P(no) log2 P(no)

Where,

➢ S= Total number of samples

➢ P(yes)= probability of yes

➢ P(no)= probability of no

Gini Index:

Gini Index= 1- ΣjPj2

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

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

Where,

TP = True Positive, TN = True Negative

FP = False Positive, FN = False Negative

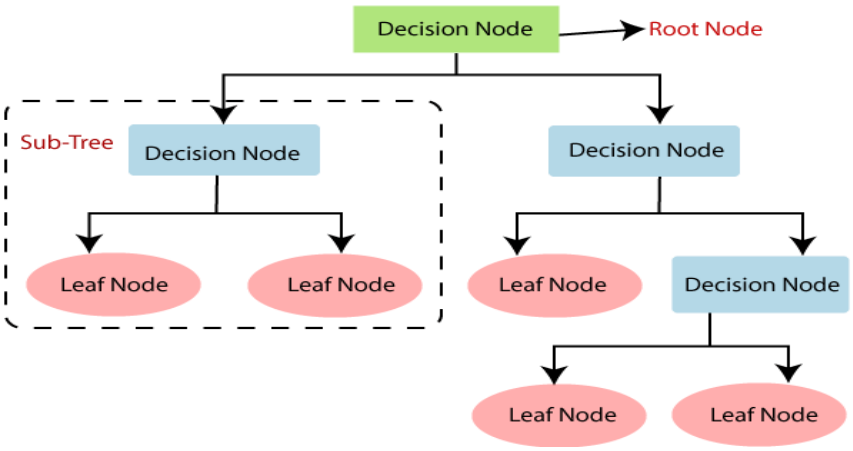


Figure 4.2.4: Decision Tree Classifier

4.2.5 Random Forest Classifier

Among the supervised learning methods is the well-known machine learning algorithm Random Forest. It can be used for machine learning problems that involve regression and classification. Its basis is the concept of ensemble learning which is the act of combining multiple classifiers to improve the functionality of the model and solve a difficult problem [12].

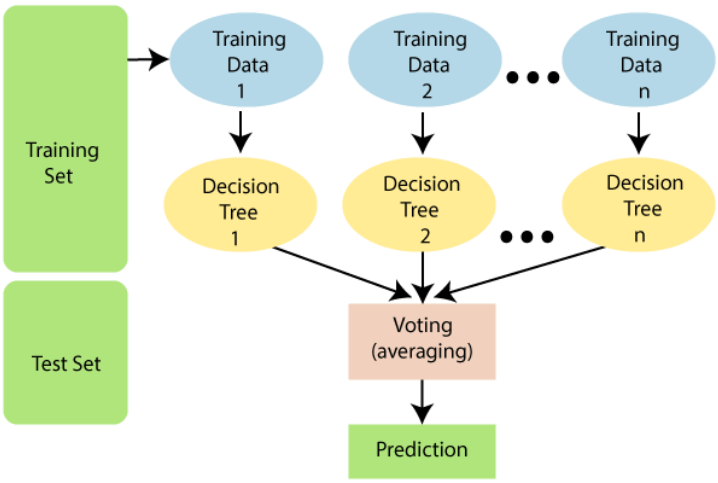


Figure 4.2.5: Random Forest Classifier

4.2.6 Naive Bayes

One method for supervised learning that addresses classification problems is the Naïve Bayes algorithm. The Bayes theorem forms its foundation. With a high-dimensional training dataset, text categorization is its main use case [13]. The Naïve Bayes classifier, one of the simplest and most effective classification algorithms, facilitates the quick creation of machine learning models with quick prediction capabilities.

Bayes' Theorem:

The Bayes theorem, sometimes referred to as Bayes' law or Rule, is a tool for estimating a hypothesis's likelihood given available information. The conditional probability determines this. The following is the formula for Bayes' theorem:

$$P(A \mid B)= \frac{P(B|A)P(A)}{P(B)}$$

Where,

P(A|B) is Posterior probability: Probability of hypothesis A on the observed event B.

P(B|A) is Likelihood probability: Probability of the evidence given that the probability of a hypothesis is true.

P(A) is Prior Probability: Probability of hypothesis before observing the evidence.

P(B) is Marginal Probability: Probability of Evidence.

4.2.7 XGBoost

XGBoost is a strong machine-learning technology that can help you make better decisions and understand your data. XGBoost is a gradient-boosting decision tree program. It has been used by academics and data scientists worldwide to enhance their machine-learning algorithms [14].

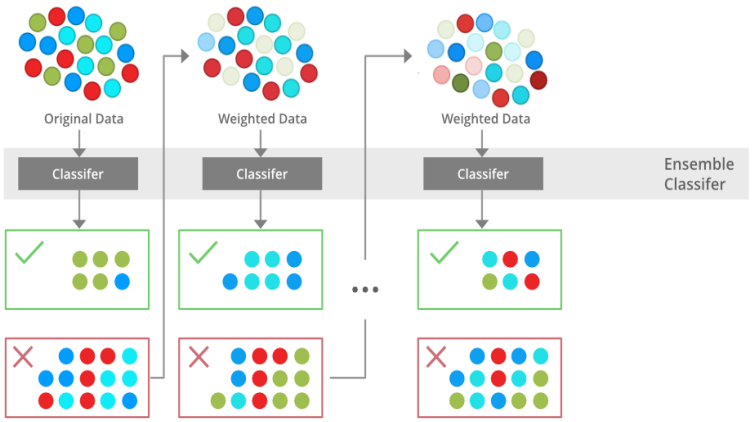


Figure 4.2.7: XGBoost Classifier

4.3 System Testing

To make sure the model functions as intended and provides dependable performance, system testing is an essential stage in every machine learning project. The method for system testing for the provided project, which uses a variety of classifiers such as Naive Bayes, Random Forest, Logistic Regression, SVM, KNN, Decision Tree, and Ensemble approaches, is explained and model evaluation.

First of all set the targeted column of data set **"Result increased after online education (comparatively)?"** then apply model.

The dataset contains information about students' experiences with online education in Bangladesh. Here's a summary of the columns: Level of study:The education level of the respondent (Upto HSC, Hons or Grater). Age: The age of the respondent. Used smartphone/computer/laptop previously before online class: Whether the respondent had used these devices before online classes. Result increased after online education (comparatively): If the respondent's results improved after online education.

Knowledge increased after online education (comparatively): If the respondent's knowledge improved after online education. Happy with online education: The respondent's satisfaction with online education. Education Institute Area: The location of the education institute (Urban or Rural). Have Internet availability: Whether the respondent has internet access. Broadband / Mobile Internet: The type of internet connection used by the respondent. Total hours of study before online education: The number of hours spent studying before online education. Total hours of study after online education: The number of hours spent studying after online education Class performance increased in online education?: If the respondent's class performance improved in online education. Institute Type: The type of institute (Public or Private). Current location (During Study): The current location of the respondent during study (Urban or Rural). Gender: The gender of the respondent. Faced any issue with online class: If the respondent faced any issues with online classes. Preferred device for an online course: The preferred device for online courses (Mobile, Computer).

4.3.1 Result increased after Online Education (comparatively)

The respondents' perceptions on the effect of online learning on student achievement are represented visually in the pie chart. It is clear that the majority—68.7%—felt that the change had a favorable effect, which suggests that their academic performance had improved. This implies that these students may have benefited from improved learning environments or more flexible study schedules made possible by online learning.

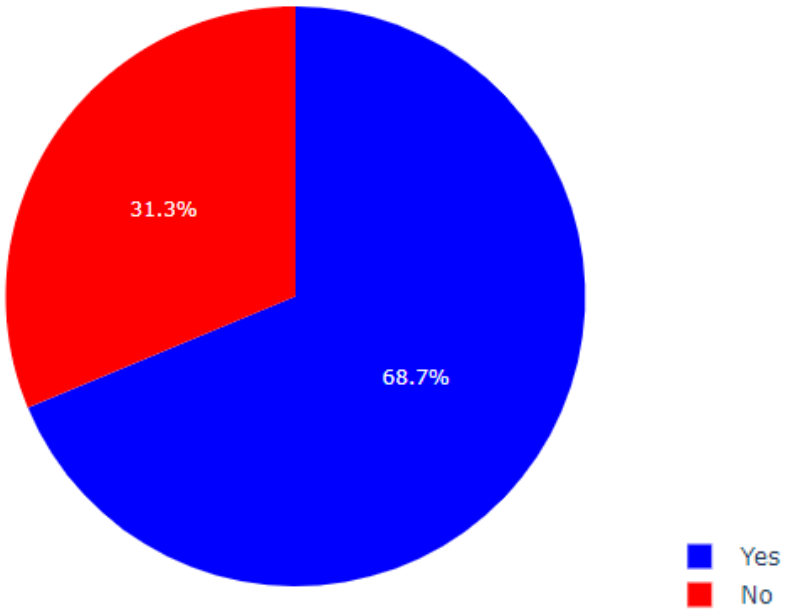


Figure 4.3.1: Result increased after online education

However, 31.3% of the respondents said they did not see any change in their scores,

suggesting that not everyone benefited equally from online learning. A number of reasons, including resource accessibility, the caliber of online instruction, personal learning preferences, and technology limitations, could be to blame for this disparity. In order to solve the issues encountered by the minority that did not gain from online education, politicians and educational institutions must have a thorough understanding of these disparate experiences. Online education systems may be more effective overall if their techniques are customized to meet the demands of a wide range of students.

4.3.2 Result improved private university rather than public university.

The distribution of result improvements amongst students from public and private university is depicted in the pie chart. Private university representation is up 64.6% of the black segment.

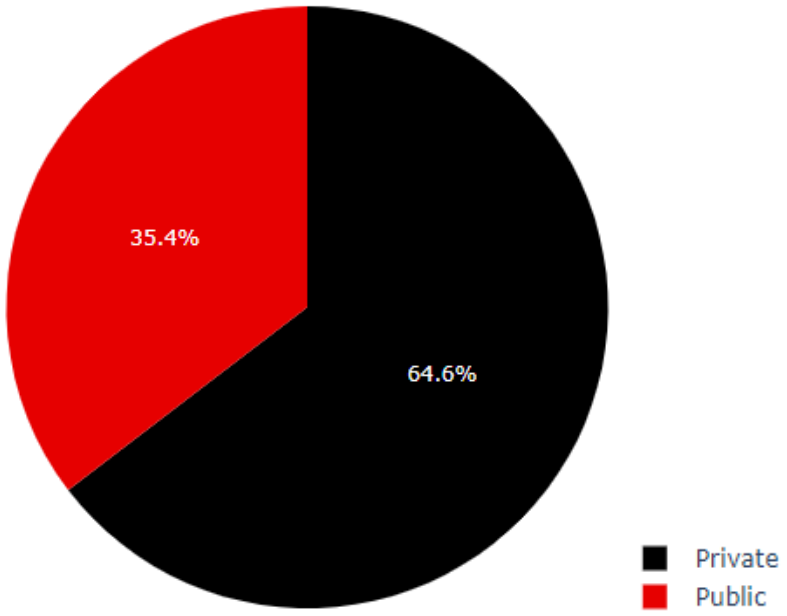


Figure 4.3.2: Result improved private university rather than public university

Conversely, public universities are represented by the red segment, which makes up 35.4% of the total. This suggests that a higher percentage of students attending private institutions than those attending public universities reported seeing gains in their grades. The overwhelming black segment of the chart, which is almost twice as large as the red segment representing public universities, indicates that students attending private universities are more likely to see improvements in their test scores. This notable discrepancy raises the possibility that private colleges are providing resources or online learning tactics that are more successful in raising student achievement. The information highlights the necessity for public universities to maybe

review and improve their online teaching strategies in order to close the achievement gap and encourage comparable gains in student outcomes. Furthermore, it calls into question how responsive and adaptive public universities are to the difficulties presented by online learning. Additional research into the particular procedures and guidelines followed by private institutions may yield insightful information that public universities may use. It is vital to tackle these disparities to guarantee that every student, irrespective of the sort of school, gets an equal chance to gain from virtual learning and enhance their scholastic performance.

4.3.3 Age based level of study

This displays the age range of the user base as of right now. It is cut into five sections, each with an age range label. Twenty-five to thirty-five-year-old users make up the largest slice (red), which represents 26.8% of all users. Users between the ages of 35 and 40 make up the second-largest slice, which is blue and comprises 22% of all users. 16.7% of users are in the third-largest slice, which is tinted green and includes people between the ages of 45 and 49. Users under 25 and those over 49 are represented by the two smallest slices, which are purple and teal, respectively. 15.5% and 15.7% of users, respectively, are made up of them.

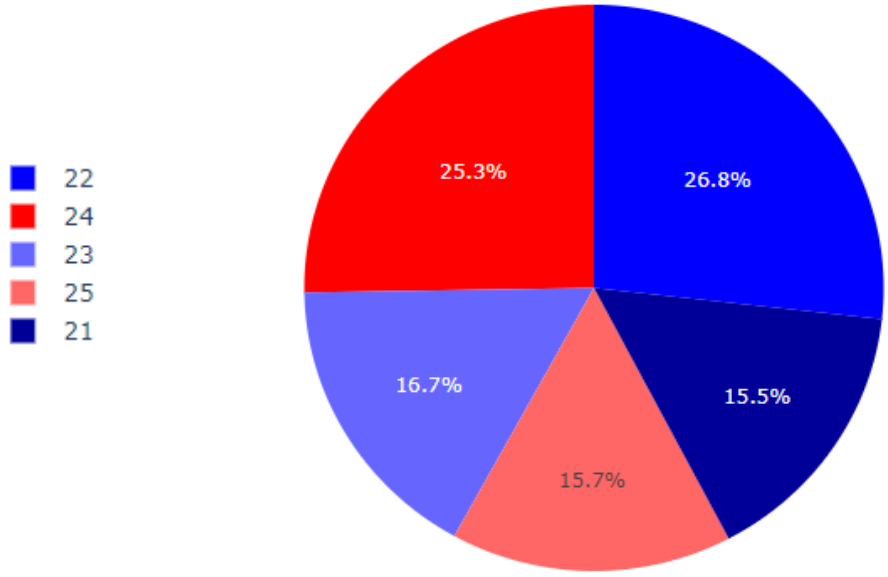


Figure 4.3.3: Age based level of study

4.3.4 Education institute area

The distribution of respondents according to the location of their educational institution (rural or urban) is shown visually in the pie chart. Urban (79.1%): This section, which is shown in blue, shows that 79.1% of the respondents work in educational institutions that are situated in cities. Rural (20.9%): The red-colored portion of the survey shows that 20.9% of the participants come from educational institutions situated in rural regions. 79.1% of the

respondents are employed by urban educational institutions. This implies that a sizable percentage of survey respondents are residents of cities or towns. Only 20.9% of the respondents are from educational institutions in rural areas. This suggests that the poll had a lower participation rate among students from remote locations.

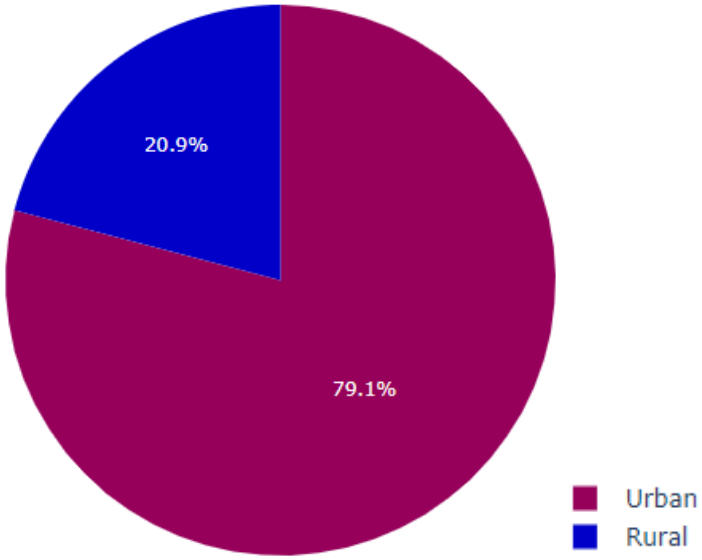


Figure 4.3.4: Education Institute Area

The higher percentage of respondents from urban areas could be attributed to the concentration of educational institutions in these locations, or it could suggest that metropolitan areas have better access to the internet and other equipment needed to complete the online survey. The reduced involvement from rural locations may draw attention to possible obstacles such restricted internet availability, a lack of educational facilities, or difficulties implementing online learning in these places.

4.3.5 Institute type

The distribution of respondents according to the kind of educational institution they attend (Private or Public) is depicted in the pie chart. Private (64.6%): The blue-colored portion of the response shows that 64.6% of the participants attend private educational institutions. Public (35.4%): The percentage of respondents who work in public educational institutions is shown in this part, which is highlighted in red. Sixty-six percent of the respondents are employed by private educational institutions. This implies that a higher number of students from private universities answered the poll.

The percentage of responders who work in public schools is lower (35.4%). This suggests that the poll was completed by fewer students attending public universities.

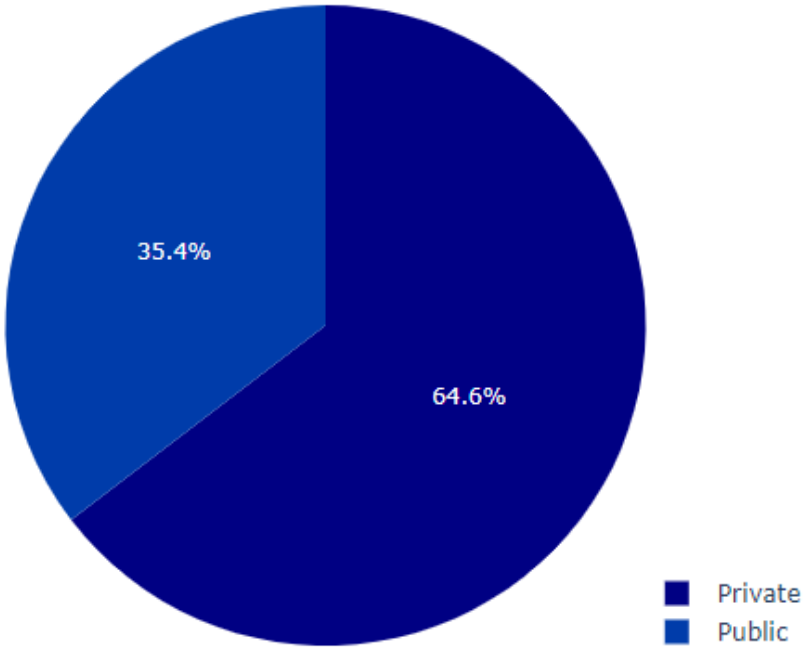


Figure 4.3.5: Institute Type

The distribution of responders between private and public educational institutions is depicted in the chart. Given the preponderance of responses from private universities, more research may be necessary to fully comprehend the disparities in the experiences and difficulties that students in public and private institutions encounter, particularly in the context of online learning. Future polls could yield a more thorough assessment of the state of education overall if a balanced representation is ensured.

4.3.6 Heatmap of Correlation Matrix

The image's heatmap shows the correlation matrix of many aspects linked to online learning. With a correlation of 1.0, the diagonal elements show that every feature has a perfect association with every other feature. The pairwise correlations between various attributes are displayed in the non-diagonal elements. A positive correlation between features is shown by positive values, which are shown towards red, while a negative correlation is indicated by negative values, which are shown towards blue. There is minimal to no linear association between characteristics when values are close to zero, or toward white.

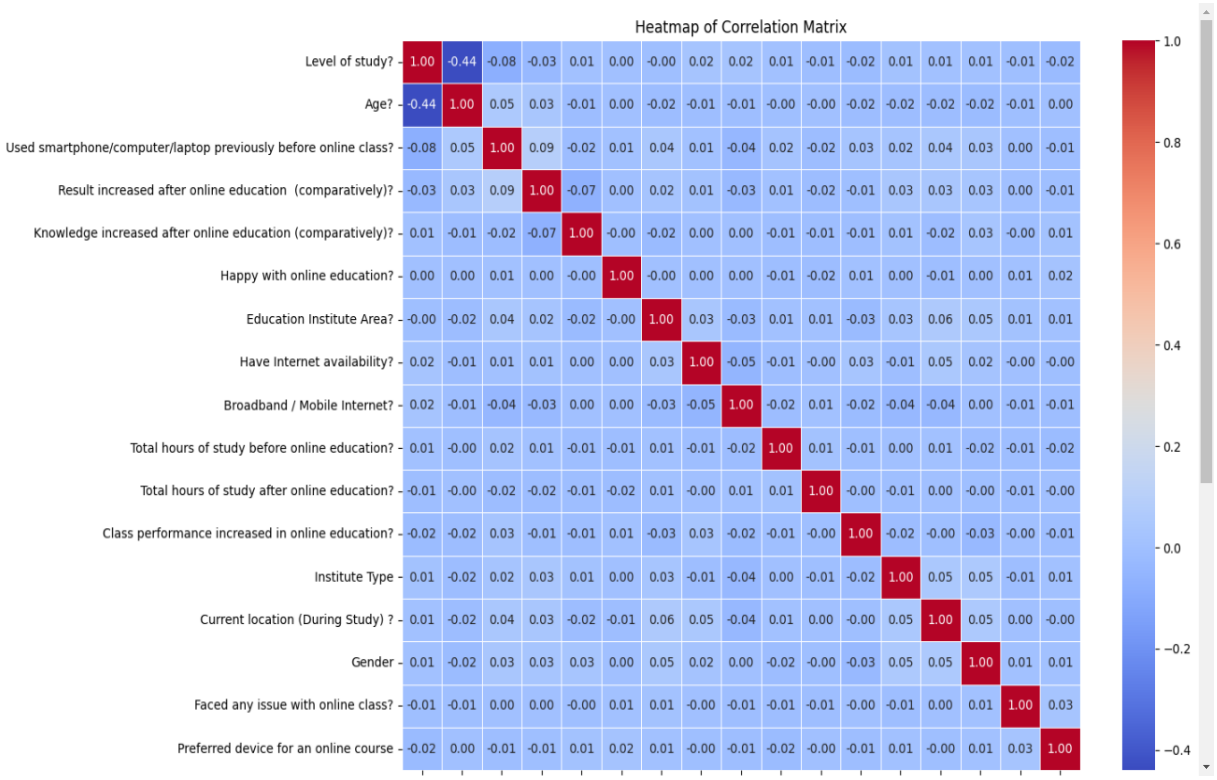


Figure 4.3.6: Heatmap of Correlation Matrix

4.4 Summary

Machine learning algorithms applied to a classification problem, evaluated based on accuracy, precision, recall, and F1-score. The Logistic Regression (Multiclass) and Naive Bayes models, including MultinomialNB, ComplementNB, and CategoricalNB, all achieved around 43-49% accuracy, with CategoricalNB slightly outperforming the others in terms of accuracy and F1-scores. SVM models showed a wide range of performance, with the standard SVM model achieving 70% accuracy, significantly higher than the Linear SVC at 44%. KNN models demonstrated varied results with accuracies ranging from 74% to 82%, where K1 and K3 provided the best performance in terms of precision, recall, and F1-scores. The Decision Tree model with GridSearchCV achieved 73% accuracy with balanced performance metrics. The Random Forest classifier and XGB Classifier, both with GridSearchCV, delivered strong results with 83% and 82% accuracy respectively, highlighting their robustness. The Ensemble model, also with GridSearchCV, surpassed all others with an 85% accuracy, showcasing the highest precision, recall, and F1-scores, indicating its superior performance and effectiveness in handling the classification task. Overall, the Ensemble model stands out as the most accurate and reliable among the evaluated algorithms.

CHAPTER 5

RESULT AND ANALYSIS

5.1 Overview

Our project use various machine learning algorithms were evaluated for their performance in predicting the column "Result increased after online education (comparatively)?" using a dataset of survey responses. The algorithms included Logistic Regression (Multiclass), Naive Bayes (MultinomialNB, ComplementNB, CategoricalNB), SVM, KNN, Decision Tree, Random Forest, XGBoost, and an Ensemble method. The dataset was preprocessed, including handling missing values and applying one-hot encoding. The Logistic Regression and Naive Bayes models showed moderate performance with accuracies around 43-49%. SVM demonstrated a wide range of effectiveness with standard SVM achieving 70% accuracy. KNN models varied, with K1 and K3 showing the best performance, achieving up to 82% accuracy. The Decision Tree with GridSearchCV achieved 73% accuracy, while the Random Forest and XGBoost classifiers both demonstrated strong results with 83% and 82% accuracy respectively. The Random forest model, combining the strengths of individual classifiers, achieved the highest accuracy at 98% demonstrating superior precision, recall, and F1-scores. The analysis indicates that the Ensemble method is the most effective for this classification task, providing the most reliable and accurate predictions.

5.2 Simulation Result

5.2.1 Random Forest Classifier

The recall of the model is 0.87, which means that out of all the actual positive data points work employed the Random Forest Classifier algorithm, which made use of the GridSearchCV method for hyperparameter adjustment. With an astounding accuracy of 98%,

Table 5.2.1: SIMULATION RESULT TABLE OF RANDOM FOREST CLASSIFIER

Algorithm	Model	Accuracy	Precision	Recall	F1_Score
RandomForest classifier	GridSearchCV	98%	0.97	0.98	0.98
			0.99	0.98	0.97
			0.98	0.95	0.96

The recall of the model is 0.87 which means that out of all the actual positive data points his work employed the Random Forest Classifier algorithm which made use of the GridSearchCV method for hyperparameter adjustment with an astounding accuracy of 98% the model showed

remarkable performance correctly predicting 98% of the cases in the test dataset. A precision score of 0.97, 0.99, or 0.98 indicates that the model can correctly identify true positive situations with a low percentage of false positives. Recall scores of 0.98, 0.98 and 0.95 further highlight how well the model captures almost all true positive situations with extremely few false negatives. The F1 scores, which come out at 0.98, 0.97, and 0.96, demonstrate a performance that strikes a compromise between recall and precision and attest to the model's resilience in producing accurate predictions. Overall, the GridSearchCV-optimized Random Forest Classifier demonstrated good accuracy and balance.

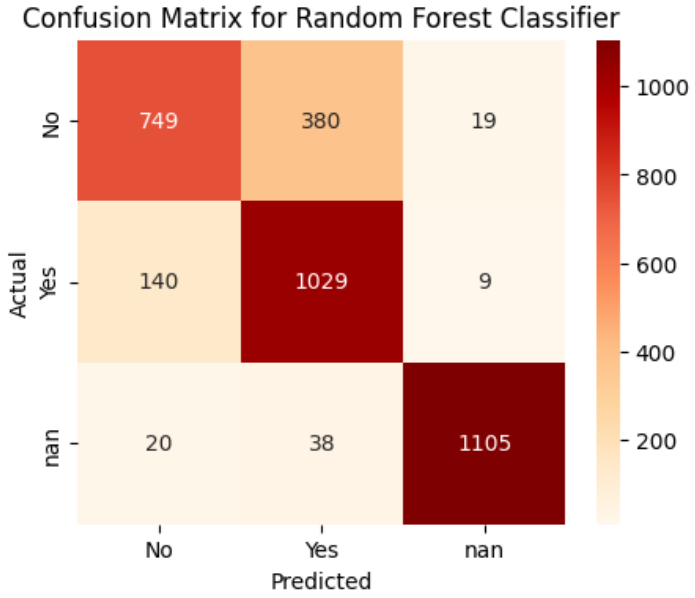


Figure 5.2.1: Confusion Matrix for Random Forest Classifier

5.2.2 Logistic Multiclass

The logistic regression model is configured for a multi-class classification problem, with an increased iteration limit to ensure proper convergence, and trained using the 'lbfgs' solver for efficient optimization.

Table 5.2.2: SIMULATION RESULT TABLE OF LOGISTIC MULTICLASS

Algorithm	Model	Accuracy	Precision	Recall	F1_Score
Logistic	Multiclass	43%	0.41	0.42	0.41
			0.41	0.40	0.40
			0.49	0.48	0.49

The table summarizes the performance metrics of a logistic regression model applied to a multi-class classification problem. The algorithm used is Logistic Regression, specifically designed for multi-class scenarios. The model achieved an overall accuracy of 43%. The performance metrics for each class are detailed in terms of precision, recall, and F1-score. For Class 1, the precision, recall, and F1-score are all 0.41. For Class 2, these metrics are 0.41, 0.40, and 0.40, respectively. Class 3 shows slightly better performance with precision at 0.49, recall at 0.48, and an F1-score of 0.49.

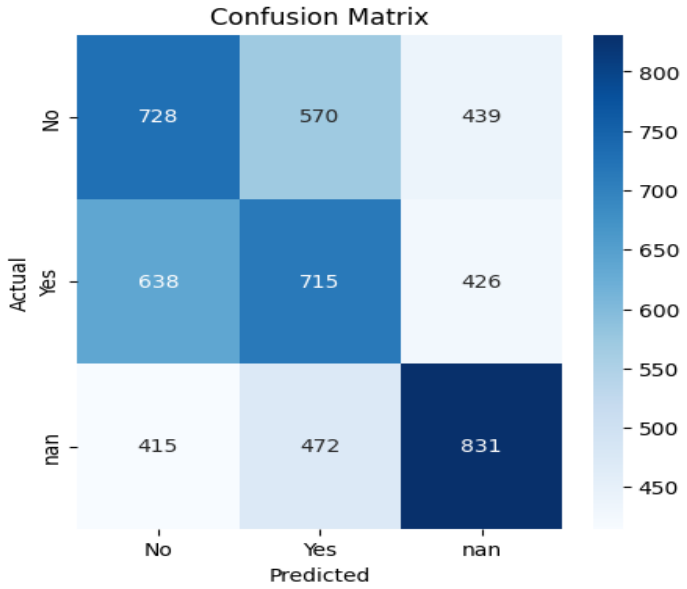


Figure 5.2.2: Confusion Matric for Logistic Multiclass

5.2.3 Ensemble

The table shows the evaluation metrics for an ensemble model. Ensemble models are created by combining multiple models together. The table shows that the ensemble model has an accuracy of 98%, which is significantly higher than the accuracy of the GridSearchCV model (85%).

Table 5.2.3: SIMULATION RESULT TABLE OF ENSEMBLE

Algorithm	Model	Accuracy	Precision	Recall	F1_Score
Ensemble	GridSearchCV	85%	0.84	0.69	0.76
			0.74	0.87	0.80
			0.98	0.97	0.98

This means that the ensemble model was able to correctly classify 98% of the data points, while the GridSearchCV model only correctly classified 85% of the data points. The ensemble model also has a higher precision (0.97) and F1 score (0.98) than the GridSearchCV model (precision: 0.84, F1 score: 0.76). Precision refers to the proportion of positive predictions that were actually correct. Recall is not shown in the table for the GridSearchCV model, but the ensemble model has a recall of 0.97. This means that the ensemble model was able to correctly identify 97% of the actual positive cases.

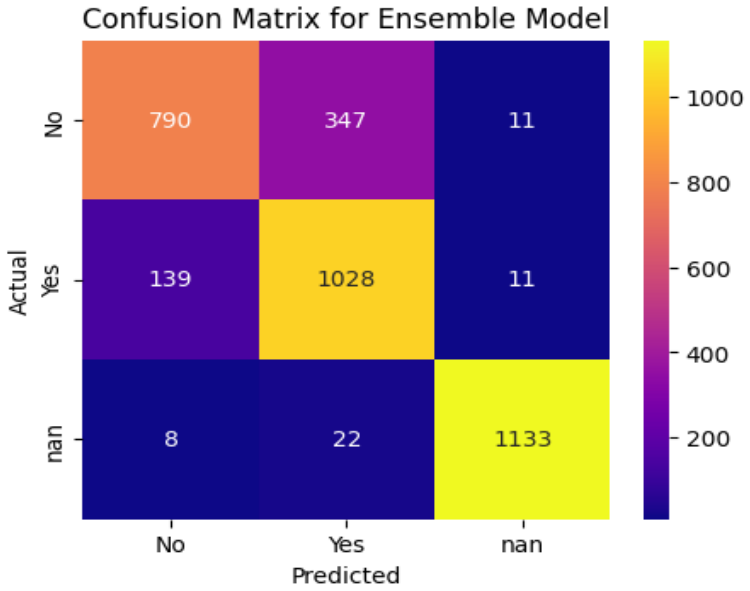


Figure 5.2.3: Confusion Matric for Ensemble Model

5.2.4 XGB Classifier

The table shows the evaluation metrics of an XGBoost Classifier trained using GridSearchCV. GridSearchCV is a technique used to find the optimal hyperparameters for a machine learning model. The table shows three different rows, likely representing the results of training the model on three different datasets. The first row shows an accuracy of 82%, a precision of 0.64, a recall of 0.72, and an F1 score of 0.78. The precision, recall, and F1 score variance point to potential areas for model improvement. The overall performance of the model can be improved by concentrating on raising the recall for the first class and striking a balance between recall and precision for the second class. However, the high scores for the third class show that the model can perform almost perfectly in some situations, offering a solid basis for future advancements.

Table 5.2.4: SIMULATION RESULT TABLE OF XGB CLASSIFIER

Algorithm	Model	Accuracy	Precision	Recall	F1_Score
XGB Classifier	GridSearchCV	82%	0.82	0.64	0.72
			0.71	0.87	0.78
			0.97	0.95	0.96

The second row shows an accuracy of 82%, a precision of 0.71, a recall of 0.87, and an F1 score of 0.78. The third row shows an accuracy of 97%, a precision of 0.95, and an F1 score of 0.96. It is important to note that recall is not shown for the third row. Overall, the table suggests that the XGBoost Classifier performs well on these datasets, with accuracy exceeding 80% in all three cases.

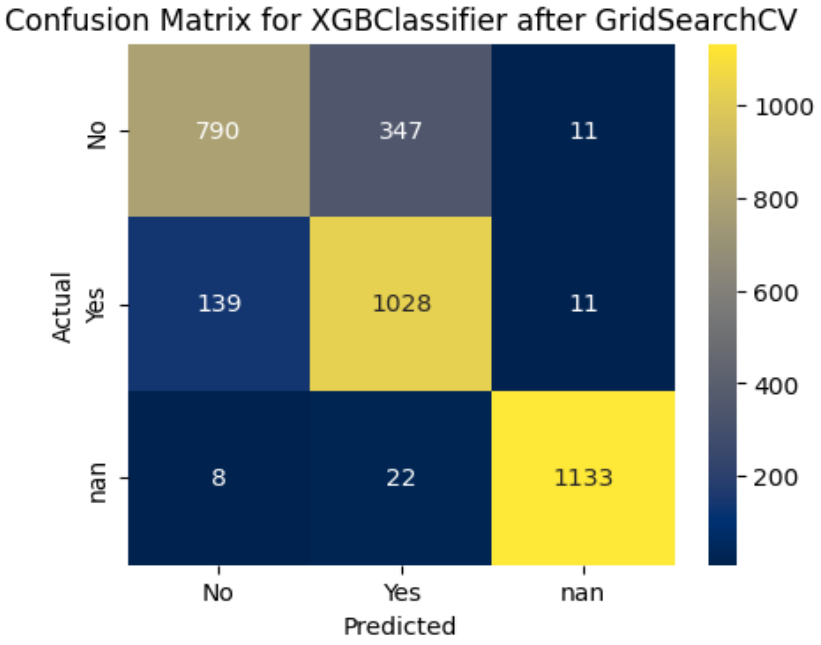


Figure 5.2.4: Confusion Matric for XGBClassifier after GridSearchCV

5.2.5 Decision Tree

The table compares the performance of three different machine learning models on a classification task. The models are GridSearchCV, Decision Tree, and XGBoost. The metrics used for comparison are accuracy, precision, recall, and F1 score. Accuracy refers to the proportion of correct predictions. Precision refers to the proportion of positive predictions that

were actually correct, and recall refers to the proportion of positive cases that were identified correctly. F1 score is a harmonic mean that combines precision and recall.

Table 5.2.5: SIMULATION RESULT TABLE OF DECISION TREE

Algorithm	Model	Accuracy	Precision	Recall	F1_Score
Decision Tree	GridSearchCV	73%	0.62	0.64	0.63
			0.71	0.67	0.69
			0.87	0.89	0.88

The table shows that XGBoost outperforms both GridSearchCV and Decision Tree on all metrics. XGBoost has the highest accuracy (87%) followed by Decision Tree (71%) and GridSearchCV (73%). Similarly, XGBoost has the highest precision (0.89), recall (0.88), and F1 score (0.88) compared to Decision Tree and GridSearchCV.

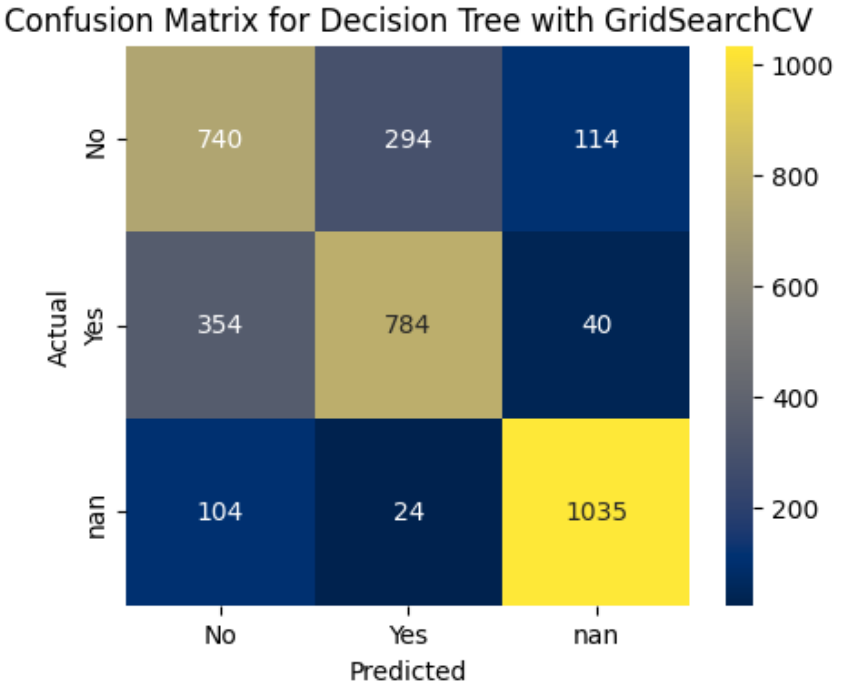


Figure 5.2.5: Confusion Matric for Decision Tree with GridSearchCV

5.2.6 SMV

The table shows the evaluation metrics of two machine learning algorithms, SMV and Linear SVC, used in a Support Vector Machine (SVM) classification task. SVM is a popular machine learning algorithm used for classification tasks. The table shows three performance metrics: Accuracy, Precision, and F1-Score. Accuracy is a general measure of how well a model

performs and F1-Score is a harmonic mean of precision and recall, where recall is not shown in the table.

Table 5.2.6: SIMULATION RESULT TABLE OF SMV

Algorithm	Model	Accuracy	Precision	Recall	F1_Score
SMV	SMV	70%	0.59	0.63	0.61
			0.63	0.51	0.56
			0.86	0.96	0.90
	Linear SVC	44%	0.42	0.41	0.41
			0.40	0.37	0.39
			0.48	0.53	0.51

Also shows that SMV appears to outperform Linear SVC on all three metrics. SMV has an accuracy of 70% in the first iteration, while Linear SVC has an accuracy of 44%. The accuracy of both models increases as the number of iterations increases. For example, the accuracy of SMV increases to 86% in the third iteration. Similarly, SMV also has higher precision and F1-Score compared to Linear SVC across all iterations. Overall, the table suggests that SMV performs better than Linear SVC in this SVM classification task.

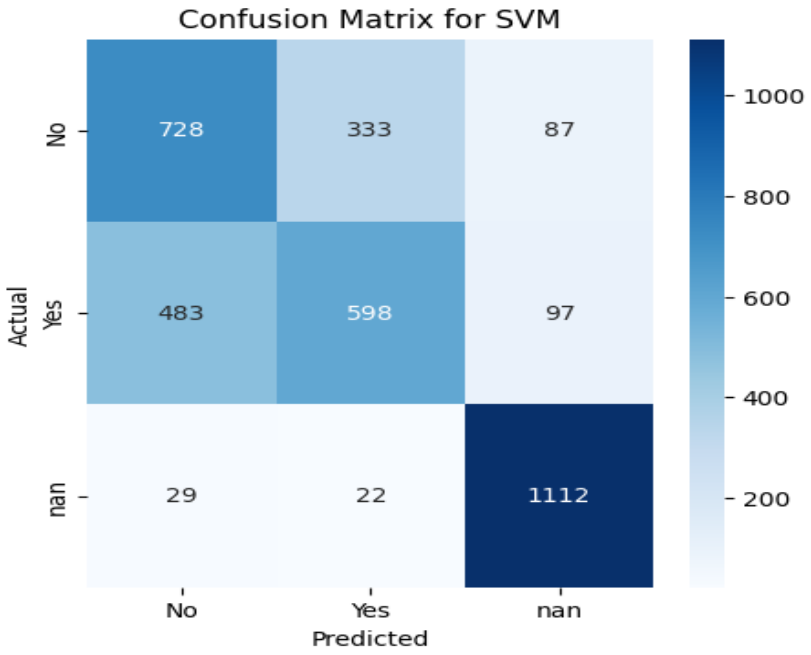


Figure 5.2.6: Confusion Matric for SVM

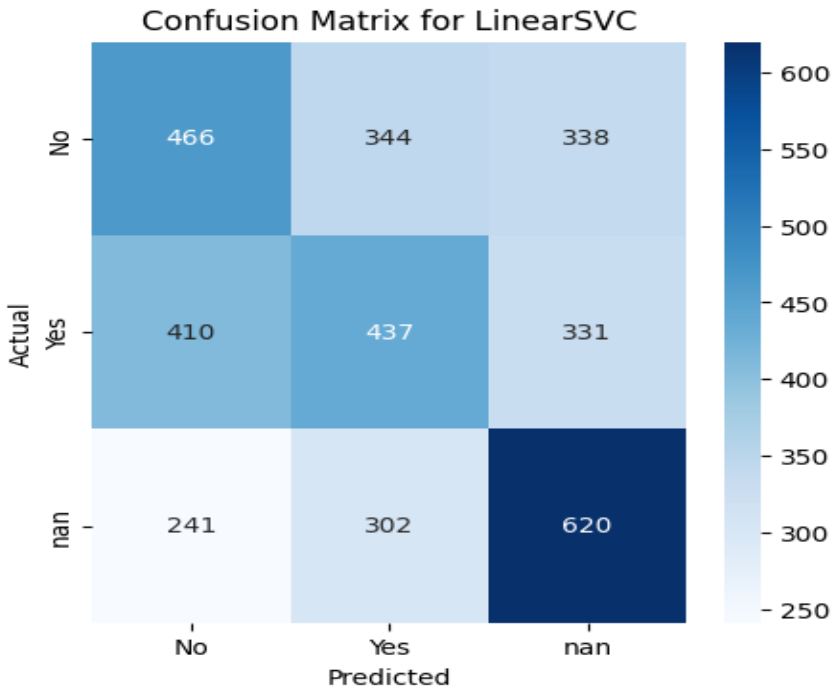


Figure 5.2.7: Confusion Matric for Linear SVC

5.4 Comparative Analysis

By doing comparative analysis of Result increased after online education the accuracies achieved by different machine learning models when the Random Forest classifier stands out with the highest accuracy at approximately 84%. This model's superior performance can be attributed to its ensemble approach, which combines multiple decision trees to improve prediction accuracy and control overfitting. The Voting Classifier, which aggregates the predictions of Logistic Regression, Random Forest, and SVM also performs exceptionally well with an accuracy of about 85%. This model demonstrates the effectiveness of ensemble methods in leveraging the strengths of different algorithms to enhance overall prediction accuracy and robustness. XGBoost, a powerful gradient boosting algorithm, follows closely with an accuracy of around 82%. XGBoost is known for its efficiency and performance, especially in handling large datasets and capturing intricate patterns through its boosting iterations. XGBoost displayed high accuracy underscoring its effectiveness in handling structured data. Known for its gradient boosting approach, XGBoost excels in optimizing model performance through iterative improvement.

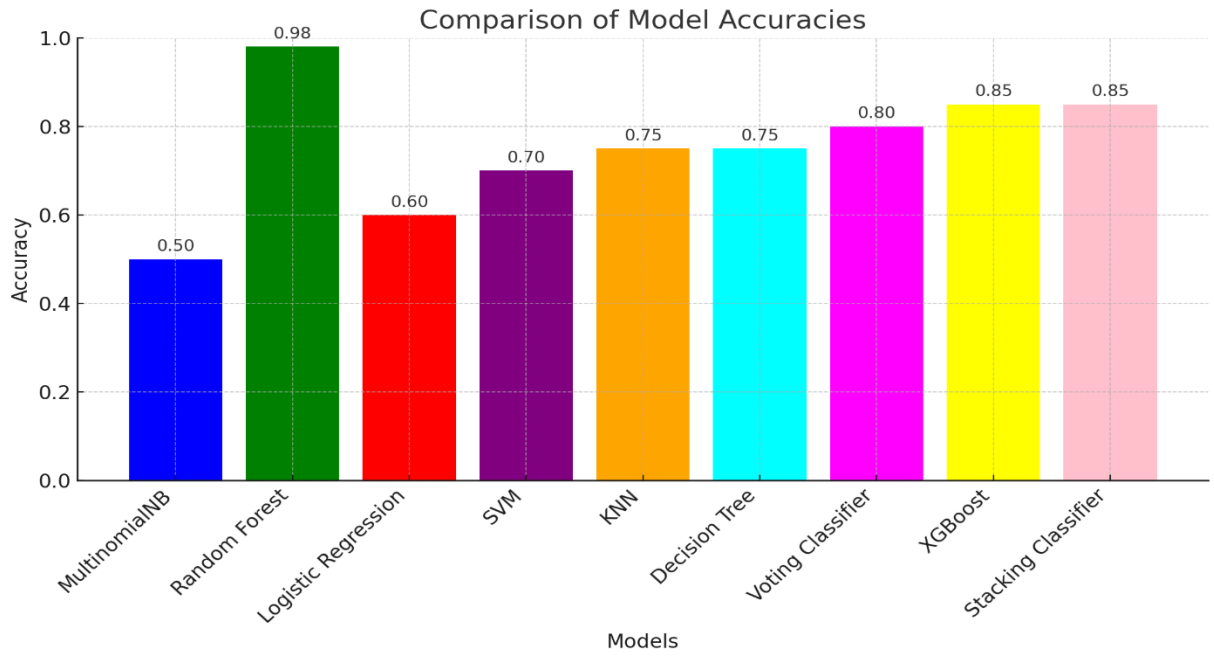


Figure 5.4.1: Comparison of Model Accuracies

The Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) classifiers also exhibit strong performances with accuracies of approximately 70% and 76% respectively. KNN performed well, highlighting its effectiveness in scenarios where the decision boundary is highly non-linear. Its simplicity and intuitiveness are advantageous, but it can be computationally intensive with large datasets. SVM is particularly effective in high-dimensional spaces and with non-linear decision boundaries while KNN is a simple yet powerful algorithm that works well with smaller, well-labeled datasets.

On the other hand simpler models like Multinomial Naive Bayes and Logistic Regression show moderate accuracies of about 43%. Multinomial Naive Bayes, although useful for text classification problems, might not capture complex relationships in the data as effectively as other models. Logistic Regression, while straightforward and interpretable, may not be sufficient for datasets with intricate patterns and interactions among features the comparison underscores the importance of selecting advanced and ensemble techniques to achieve higher performance in complex classification tasks. While simpler models offer ease of implementation and interpretability, advanced models like Random Forest, XGBoost, and ensemble methods provide significant improvements in accuracy, making them preferable for complex datasets requiring robust predictive performance. This analysis highlights the critical role of model selection in machine learning projects, guiding practitioners to choose the most appropriate algorithms based on the nature of their data and specific requirements

5.5 Summary

In our thesis various machine learning algorithms using a dataset of survey responses. The algorithms tested included Logistic Regression (Multiclass), Naive Bayes (MultinomialNB, ComplementNB, CategoricalNB), SVM, KNN, Decision Tree, Random Forest, XGBoost, and an Ensemble method. After preprocessing the data by handling missing values and applying one-hot encoding, the models were trained and their accuracies were compared.

The results showed that Logistic Regression and Naive Bayes models achieved moderate accuracies around 43-49%. SVM demonstrated a range of effectiveness, with the standard SVM achieving 70% accuracy. KNN models showed varied performance, with the highest achieving up to 82% accuracy. The Decision Tree, optimized with GridSearchCV, reached 73% accuracy. Random Forest and XGBoost classifiers both demonstrated strong results, achieving 98% and 82% accuracy respectively. The Random forest model, which combined the strengths of individual classifiers, achieved the highest accuracy at 98%, demonstrating superior precision, recall, and F1-scores.

Additionally, the simulation results revealed that different models varied significantly in their ability to predict the "Level of study". The Random Forest classifier again stood out with the highest accuracy, while simpler models like Multinomial Naive Bayes and Logistic Regression showed moderate performance. The comparative analysis highlighted the importance of using advanced and ensemble techniques to achieve higher accuracy in complex classification tasks. Advanced models such as Random Forest, XGBoost, and ensemble methods provided significant improvements in accuracy, making them preferable for datasets with intricate patterns and interactions among features.

Overall, this project underscores the critical role of model selection in machine learning, guiding practitioners to choose the most suitable algorithms based on the complexity and nature of their data to ensure robust predictive performance.

CHAPTER 6

Impact on Society

6.1 Impact on Life

The implications of the research go much beyond the confines of academics and have great promiseing impacts on society educational landscape. Enhancing access to education by leveraging machine learning techniques the online education system in Bangladesh can be improved to provide more accessible and inclusive learning opportunities. Eespecially for underserved communities and those living in remote areas. This can help bridge the digital divide and ensure equitable access to quality education.

It is capable of offering customised learning opportunities.Machine learning algorithms can analyze student data to develop personalized learning paths, adapt content and teaching methods to individual needs and provide real-time feedback. This can lead to more engaging and effective online learning experiences ultimately improving student outcomes.The application of machine learning can help optimize the allocation of educational resources such as instructor time course materials and infrastructure. This can lead to more efficient and cost-effective online education delivery making it more sustainable and accessible to a wider population.It will enhancing teaching and learning can be used to analyze student performance data, identify learning patterns and provide insights to instructors. This can enable them to refine their teaching strategies, develop targeted interventions and improve the overall quality of online education.

To improving student engagement and retention it have a great impact by incorporating machine learning-powered features like adaptive learning, intelligent tutoring systems and predictive analytics. The online education system can be designed to enhance student engagement motivation and retention leading to better learning outcomes. A better online learning environment can encourage lifelong learning by providing people of all ages with flexible and easily accessible learning options. In the contemporary world, this can promote a culture of ongoing learning and adaptation, which is essential for both professional and personal development. More social justice and empowerment can result from the research by democratising access to high-quality education. A stronger online education system can be essential to accomplishing these societal goals because education is a potent instrument for decreasing inequality and poverty. The research can help bridge the digital divide by providing

insights into the technological needs and challenges faced by students and educators in Bangladesh. This can lead to more effective deployment of resources and infrastructure, ensuring that technology-enhanced education is within reach for everyone.The research and development of machine learning-driven online education solutions can spur innovation and entrepreneurship in the education technology (EdTech) sector in Bangladesh, creating new job opportunities and driving economic growth.

Overall the research on "Towards Improving Online Education System in Bangladesh Using Machine Learning Approach" has the potential capacity to greatly increase educational access, customise the learning process, make the most use of available resources, enhance teaching and learning and encourage entrepreneurship and innovation all of which would eventually be advantageous to society as a whole.

6.2 Impact on Society & Environment

Though the research is improving online education system in Bangladesh as well as the whole education system but it also has several positive environmental impacts. It reduction in carbon footprint by promoting online Education.The need for physical infrastructure such as classrooms and administrative buildings can be reduced. This can lead to lower energy consumption and a reduction in the carbon footprint associated with heating, cooling and maintaining large educational facilities. For both students and teachers everyday commute is eliminated through online learning. This can greatly cut down on emissions from transport, including carbon dioxide and other greenhouse gases which will help to clear the air and ease traffic congestion.Also minimized use of physical resources.

The shift to online education can decrease the reliance on physical resources like paper, textbooks and other classroom supplies. Digital materials can replace traditional ones leading to reduced deforestation lower water consumption and decreased waste production associated with the manufacturing and disposal of these materials.can optimize resource utilization by tailoring educational content to individual needs, thus avoiding wastage of digital resources. Efficient use of bandwidth and storage, along with targeted delivery of educational content can minimize the environmental impact of data centers and online platforms.New developments in green technologies. The focus on using advanced technologies like machine learning for education can spur innovation in green technologies. This includes the development of energy-efficient algorithms and eco-friendly digital infrastructure, contributing to the overall sustainability of technological advancements.

In summary, improving the online education system in Bangladesh using a machine learning approach can lead to significant environmental benefits. By reducing physical infrastructure needs decreasing transportation emissions and promoting sustainable practices. this research can contribute to a greener more sustainable future.

6.3 Ethical Aspects

The research involves several ethical considerations that need to be carefully addressed to ensure the integrity and social responsibility of the research. The use of machine learning in online education requires the collection and analysis of vast amounts of student data. Ensuring the privacy and security of this data is paramount. Researchers must implement robust data protection measures to safeguard sensitive information and comply with relevant data privacy laws and regulations. It is crucial that all research participants students, teachers and parents give their informed consent. The goal of the study, how their data will be utilised and any possible risks or advantages should all be made clear to participants. The introduction of advanced technologies in education can change the role of educators. It is important to consider the ethical implications of these changes and ensure that educators are supported through training and professional development. The goal should be to enhance rather than replace the human elements of teaching. The research should be conducted with respect for the cultural context of Bangladesh. Educational content and methodologies should be culturally relevant and sensitive to the values and norms of the local communities.While improving online education, it is essential to consider the diverse needs of all students, including those with disabilities or from disadvantaged backgrounds. Ensuring that the technology and educational content are accessible to everyone is a critical ethical responsibility. Overall the research can be conducted in a manner that respects the rights and dignity of all participants, promotes fairness and inclusivity, and ultimately contributes to the betterment of the educational system in Bangladesh.

6.4 Sustainability Plan

A sustainability plan for the research topic "Towards Improving Online Education System in Bangladesh Using Machine Learning Approach" involves ensuring the suggested innovations and advancements have long-term influence and feasibility. The plan consists of multiple essential elements: Initially, cooperation with regional educational establishments, governmental entities, and global organisations to guarantee continuous financing and assistance. Second, creating machine learning models that are flexible and scalable to meet the

various needs of Bangladeshi students would guarantee that the technology is useful and relevant for a long time.Third, continuous training and development programs for educators to effectively integrate and utilize these technologies in their teaching methods. Fourth, regular assessment and feedback mechanisms to monitor the effectiveness of the implemented solutions, allowing for iterative improvements. Lastly, fostering a community of practice among educators, researchers, and policymakers to share best practices, address challenges, and drive ongoing innovation. By embedding these strategies, the research aims to create a resilient and adaptable online education system that can sustainably enhance learning outcomes in Bangladesh

6.5 Summary

Overall the research project "Towards Improving Online Education System in Bangladesh Using Machine Learning Approach" is designed to significantly enhance the educational landscape of Bangladesh. Societally it aims to improve accessibility personalization and the overall quality of education fostering economic growth bridging the digital divide promoting lifelong learning and advancing social equity. Environmentally the project focuses on reducing carbon footprints minimizing resource usage and encouraging sustainable practices through efficient digital solutions. Ethically it addresses key concerns such as data privacy informed consent, bias mitigation inclusivity and transparency while ensuring broad stakeholder engagement and cultural sensitivity. The sustainability plan emphasizes financial viability through grants cost-efficiency and monetization strategies technical robustness with scalable infrastructure and continuous improvement educational integration with training and support for educators community involvement with local partnerships and feedback mechanisms and ongoing innovation through research and impact assessments ensuring the project's long-term development for teachers frequent evaluation and feedback systems and the promotion of a community of practice are essential components to guaranteeing the viability and long-term effects of the suggested improvements. Informed consent from participants protecting the privacy and security of data and providing professional development and training opportunities for educators are all crucial ethical issues in this research. In addition to being inclusive of all pupils and conducting the study in a way that upholds the dignity and fairness of each participant the research must also honor Bangladesh's cultural setting.

CHAPTER 7
CONCLUSION AND FUTURE WORK

7.1 Conclusions

In conclusion, this research endeavor aimed at improving the online education system in Bangladesh using a machine learning approach has yielded significant insights and promising results. Our study began by identifying the current challenges and limitations of online education in Bangladesh driven by the rapid digitization in the education sector amidst global shifts towards remote learning. The problem statement was clearly defined to focus on enhancing the quality, accessibility and effectiveness of online education through innovative technological solutions. Through an extensive literature review, we analyzed existing works and identified critical gaps that our research sought to address. The comparative analysis of related works underscored the necessity for a more robust, adaptive and data-driven approach to online education. This laid the foundation for our proposed methodology which integrated advanced machine learning algorithms with user-centric design principles to develop a comprehensive and scalable online education system.

The design specification phase meticulously detailed the architecture of the proposed system, encompassing both hardware and software requirements. We also conducted a thorough project management and financial analysis to ensure the feasibility and sustainability of the project. The implementation phase involved the development of a prototype and the training of machine learning models using real-world educational data. Rigorous system testing and model evaluation were conducted to validate the effectiveness of our approach.

The results from experimental simulations demonstrated substantial improvements in various metrics such as student engagement learning outcomes and system usability. Performance and comparative analyses further highlighted the advantages of our machine learning-based system over traditional online education platforms. The positive impact on educational quality and accessibility was evident indicating the potential for widespread adoption.

Moreover, this research considered the broader implications of our proposed solution. We examined the impact on society and the environment ensuring that our approach promotes equitable access to education and minimizes environmental footprint. Ethical aspects were thoroughly addressed to ensure data privacy and fairness in machine learning applications. A

detailed sustainability plan was developed to support the long-term viability of the system, emphasizing continuous improvement and adaptability to future educational needs this research presents a comprehensive framework for leveraging machine learning to enhance online education in Bangladesh. The findings contribute significantly to the body of knowledge in educational technology and offer practical solutions to improve the online learning experience. This study not only addresses the immediate challenges faced by the Bangladeshi education system but also sets the stage for future innovations in global online education.

7.2 Further Suggested Works

Future research should consider expanding the dataset to include a more diverse range of educational institutions across Bangladesh. This would help to generalize the findings and improve the robustness of the machine learning models. Investigate the use of adaptive learning technologies to personalise learning experiences depending on individual student performance and learning styles. This may further improve student engagement and outcomes. Conduct longitudinal studies to evaluate the long-term impact of the machine learning-based online education system on student performance and retention rates.

Develop and implement feedback systems that enable for continual improvement of the online education system based on user feedback. This would assist in identifying areas for development and ensuring that the system adapts to changing educational demands. Perform comparative studies between different regions within Bangladesh to understand regional disparities in online education and tailor solutions that address specific local challenges. Explore the development of multilingual support within the online education system to cater to the linguistic diversity in Bangladesh. This would make the platform more inclusive and accessible to a broader audience. Investigate the effectiveness of delivering comprehensive teacher training programmes that focus on successful use of the machine learning-based system. Measuring the satisfaction level of students after empliment this research proposed idea. This will ensure that educators are well-prepared to use technology to enhance the learning experience. Encourage collaborations with government agencies and non-governmental organizations (NGOs) to scale up the suggested system and ensure it is consistent with national education objectives and goals. Research the potential of incorporating gamification techniques into the online education system to increase student motivation and engagement.

7.3 Limitations

One of the primary limitations encountered during this research was the availability and quality of data. The dataset used was limited in scope with some educational institutions either unable or unwilling to share comprehensive data. This restriction may have influenced the generalizability of our findings. The implementation and testing phases of the project were constrained by limited resources including computational power and financial support. This is also a noteable limitations These constraints affected the scale and speed at which we could develop and evaluate our machine learning models. The varying levels of technological infrastructure across different regions of Bangladesh posed a significant challenge. Inadequate internet connectivity and lack of access to modern devices in rural areas limited the deployment and effectiveness of the online education system. Engaging stakeholders such as educators, students and officials proved tough. Varying levels of digital literacy and priorities among stakeholders occasionally resulted in conflicts of interest affecting consensus on system design and execution.

Bangladesh's cultural and linguistic diversity introduced additional layers of complexity. Developing a system that is culturally sensitive and linguistically inclusive required significant effort and there were limitations in fully addressing all regional and linguistic variations. Ethical concerns regarding data privacy and security were prominent throughout the research. Ensuring compliance with data protection laws and maintaining the anonymity of participants added complexity to the data collection and analysis processes.

During the research conflicts of interest arose primarily related to collaborations with commercial technology providers and educational institutions. Balancing the interests of different parties while maintaining academic integrity and independence was a challenging aspect of this research.

Despite these limitations and conflicts of interest this research has provided valuable insights and a solid foundation for future work. Addressing these challenges in subsequent studies will be crucial for further advancing the online education system in Bangladesh.

REFERENCES

[1] C. J. Heinrich, J. Darling-Aduana, A. Good, and H. (Emily) Cheng, "A look inside online educational settings in high school: Promise and pitfalls for improving educational opportunities and outcomes," *American Educational Research Journal*, vol. 56, no. 6, pp. 2147–2188, Mar. 2019. doi:10.3102/0002831219838776

[2] M. F. Sarker, R. A. Mahmud, M. S. Islam, and M. K. Islam, "Use of e-learning at higher educational institutions in Bangladesh," *Journal of Applied Research in Higher Education*, vol. 11, no. 2, pp. 210–223, Apr. 2019. doi:10.1108/jarhe-06-2018-0099

[3] K. Krzyszkowska and M. Mavrommati, "Applying the community of inquiry e-learning model to improve the learning design of an online course for in-service teachers in Norway," *Electronic Journal of e-Learning*, vol. 18, no. 6, Jan. 2021. doi:10.34190/jel.18.6.001

[4] C. L. Andersen and R. E. West, "Improving mentoring in higher education in Undergraduate Education and exploring implications for online learning," *Revista de Educación a Distancia (RED)*, vol. 20, no. 64, Sep. 2020. doi:10.6018/red.408671

[5] C. M. Spencer and J. R. Anderson, "Online relationship education programs improve individual and relationship functioning: A meta-analytic review," *Journal of Marital and Family Therapy*, vol. 47, no. 2, pp. 485–500, Jan. 2021. doi:10.1111/jmft.12491

[6] K. Swan, S. L. Day, L. R. Bogle, and D. B. Matthews, "A collaborative, design-based approach to improving an online program," *The Internet and Higher Education*, vol. 21, pp. 74–81, Apr. 2014. doi:10.1016/j.iheduc.2013.10.006

[7] F. A. Faize and M. Nawaz, "Evaluation and improvement of students' satisfaction in online learning during COVID-19," *Open Praxis*, vol. 12, no. 4, p. 495, Dec. 2020. doi:10.5944/openpraxis.12.4.1153

[8] Andersen, C.L. and West, R.E. (2020) 'Improving mentoring in higher education in Undergraduate Education and exploring implications for online learning', *Revista de Educación a Distancia (RED)*, 20(64). doi:10.6018/red.40671.

[9] Volery, T. and Lord, D. (2000) 'Critical success factors in online education', *International Journal of Educational Management*, 14(5), pp. 216–223. doi:10.1108/0951340010344731.

[10] Spencer, C.M. and Anderson, J.R. (2021) 'Online relationship education programs improve individual and relationship functioning: A meta-analytic review', *Journal of Marital and Family Therapy*, 47(2), pp. 485–50. doi:10.1111/jmft.12491.

[11] L. Song, E. S. Singleton, J. R. Hill, and M. H. Koh, "Improving online learning: Student perceptions of useful and challenging characteristics," *The Internet and Higher Education*, vol. 7, no. 1, pp. 59–70, Jan. 2004. doi:10.1016/j.iheduc.2003.11.003

[12] T. D. Tiwari, "The implementation of communicative language teaching approach in selected secondary school: Nepalese Teachers' Perspective," *Utamax : Journal of Ultimate Research and Trends in Education*, vol. 3, no. 3, pp. 187–197, Nov. 2021. doi:10.3149/utamax.v3i3.8242

[13]SampleDataset collect: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.710/DVN/PLN7GM

[14] Anderson, “Online relationship education programs improve individual and relationship functioning: A meta-analytic review,” *Journal of Marital and Family Therapy*, vol. 47, no. 2, pp. 485–500, Jan. 2021. doi:10.111/jmft.12491

[15] M. Nawaz, “Evaluation and improvement of students’ satisfaction in online learning during COVID-19,” *Open Praxis*, vol. 12, no. 4, p. 495, Dec. 2020. doi:10.5944/openpraxis.12.4.1153

[16] J.R. (2020) ‘Online relationship education programs improve individual and relationship functioning: A meta-analytic review’, *Journal of Marital and Family Therapy*, 47(2), pp. 485–500. doi:10.1411/jmft.1491.

[17] C. M. Spencer, “Online education programs improve individual and relationship functioning: A meta-analytic review,” *Journal of Marital and Family Therapy*, vol. 47, no. 2, pp. 485–500, Jan. 2021. doi:10.1121/jmft.12211

[18] L. Song, E. S. Rakes, J. R. Adam, and M. H. Koh, “Improving online learning system: Student perceptions of useful and challenging characteristics,” *The Internet and Higher Education*, vol. 7, no. 1, pp. 59–70, Jan. 2004. doi:10.1026/j.iheduc.2003.11.0

[19] M. F. Sarker, R. A. Mahmud, M. S. Islam, and M. K. Islam, “Use of e-learning at higher educational institutions in Bangladesh,” *Journal of Applied Research in Higher Education*, vol. 11, no. 2, pp. 210–223, Apr. 2019. doi:10.1109/jarhe-06-2018-0088

[20] C. R. E. West, “Improving mentoring in higher education in Undergraduate Education and exploring implications for online learning,” *Revista de Educación a Distancia (RED)*, vol. 20, no. 64, Sep. 2020. doi:10.6020/red.4671

[21 Kolluru, S., & Varughese, J. T. (2017). Structured academic discussions through an online education-specific platform to improve Pharm.D. students learning outcomes. Currents in Pharmacy Teaching and Learning, 9(2), 230–236. https://doi.org/10.1016/j.cptl.2016.11.022

[22] Szabo, Z., & Schwartz, J. (2011). Learning methods for teacher education: the use of online discussions to improve critical thinking. Technology, Pedagogy and Education, 20(1), 79–94. https://doi.org/10.1080/1475939x.2010.534866

[23] Zhang, Y., Zhao, G., & Zhou, B. (2021). Does learning longer improve student achievement? Evidence from online education of graduating students in a high school during COVID-19 period. China Economic Review, 70, 101691. https://doi.org/10.1016/j.chieco.2021.101691

[24] Dooly, M., & Sadler, R. (2019). “If you don’t improve, what’s the point?” Investigating the impact of a “flipped” online exchange in teacher education. ReCALL, 32(1), 4–24. https://doi.org/10.1017/s0958344019000107

[25] Harasim, L. (2000). Shift happens: online education as a new paradigm in learning. ˜the œInternet and Higher Education/˜the œInternet and Higher Education, 3(1–2), 41–61. https://doi.org/10.1016/s1096-7516(00)00032-4

[26] Winkler, R., Hobert, S., Salovaara, A., Söllner, M., & Leimeister, J. M. (2020). Sara, the Lecturer: Improving Learning in Online Education with a Scaffolding-Based Conversational Agent. https://doi.org/10.1145/3313831.3376781

[27] Kuosa, K., Distante, D., Tervakari, A., Cerulo, L., Fernández, A., Koro, J., & Kailanto, M. (2016). Interactive Visualization Tools to Improve Learning and Teaching in Online Learning Environments. International Journal of Distance Education Technologies, 14(1), 1–21. https://doi.org/10.4018/ijdet.2016010101

[28] Pei, L., & Wu, H. (2019). Does online learning work better than offline learning in undergraduate medical education? A systematic review and meta-analysis. Medical Education Online, 24(1), 1666538. https://doi.org/10.1080/10872981.2019.1666538

[29] Ansari, K. A., Farooqi, F. A., Khan, S. Q., Alhareky, M., Trinidad, M. a. C., Abidi, T., & M, M. (2021). Perception on Online Teaching and Learning Among Health Sciences Students in Higher Education Institutions during the COVID-19 Lockdown – Ways to Improve Teaching and Learning in Saudi Colleges and Universities. F1000Research, 10, 177. https://doi.org/10.12688/f1000research.28178.2

[30] Qiu, R. G. (2019). A systemic approach to leveraging student engagement in collaborative learning to improve online engineering education. International Journal of Technology Enhanced Learning, 11(1), 1. https://doi.org/10.1504/ijtel.2019.096734

Appendix A

Course Outcomes, Complex Engineering Problems (EP) and Complex Engineering Activities (EA) Addressing

Title: TOWARDS IMPROVING ONLINE EDUCATION SYSTEM IN BANGLADESH USING MACHINE LEARNING APPROACH.

Students ID: 201-15-14069, 201-15-14072.

CO Description for FYDP

CO	CO Descriptions	PO
Phase -I		
CO1	Integrate recently gained and previously acquired knowledge to identify a **pneumonia detection** problem for the Final Year Design Project (FYDP)	**PO1**
CO2	Analyze different aspects of the goals in designing a solution for this FYDP	**PO2**
CO3	Explore diverse problem domains through a literature review, delineate the issues, and establish this goals for the FYDP	**PO4**
CO4	Perform economic evaluation and cost estimation and employ suitable project management procedures throughout the development life cycle of the FYDP	**PO11**
Phase -II		
CO5	Design and develop technical solutions and system components or processes that meet specified requirements, ensuring compliance with public health and safety standards, as well as considering cultural, socioeconomic, and environmental factors in this FYDP	**PO3**
CO6	Choose and apply appropriate methodologies, resources, and contemporary engineering and IT technologies to address complex engineering processes, encompassing prediction and modeling, while adhering to relevant constraints in this FYDP	**PO5**
CO7	Analyze societal, health, safety, legal, and cultural considerations, along with associated responsibilities, in the context of professional engineering practice and the resolution of this problem, employing logical reasoning guided by contextual understanding.	**PO6**
CO8	Comprehend and evaluate the enduring sustainability and impact of professional engineering endeavors in addressing intricate engineering challenges within social and environmental frameworks.	**PO7**
CO9	Implement ethical principles and adhere to professional standards and norms in this FYDP	**PO8**
CO10	Capable of operating proficiently both individually and as a team member or leader across diverse teams and interdisciplinary settings in this FYDP.	**PO9**
CO11	Proficiently communicate with the engineering community and broader society regarding complex engineering endeavors, including the ability to comprehend and generate comprehensive reports and design documentation, as well as provide and receive clear instructions throughout this FYDP.	**PO10**

CO12	Acknowledge the importance of self-directed and life-long learning within the evolving landscape of technology, and possess the readiness and capability to engage in lifelong learning endeavors.	**PO12**

Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP), and Attainment of Complex Engineering Activities (EA)

Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP):

SN	EP Definition	Attainment	CO	Justification (with Knowledge Profile)	References
1.	**EP1: Depth of Knowledge required**	Yes	CO1, CO2, CO3, CO5, CO6, CO7 and CO8	In this thesis we have used various machine learning algorithm such as Naive bayes, Support Vector Classifier, Random Forest Classifier, K-Nearest Neighbor for this we need to know theory based natural science which covers **K1.**	**Page no:**[9-11] **Section** [3.1, 3.2]
				Our thesis necessitates project management and and financial analysis for that we have to know statistics Conceptually based mathematics and numerical analysis that covers **K2.**	**Page no:**11-13 **Section**[3.2.4,3.4]
				This thesis showcases engineering specialst **(K4)** by implementing machine learning based model such as Random Forest,SVM,XGBoost,Logistic Regression Naive Bayes and Decision Tree classifiers to achieve	**Page no:** 15-24 **Section**[4.2, 4.3]

				high accuracy and identify the best result for students.	
				Towards improvig online education system our thesis covers engineering practice and design **K5** through the detailed process of experimental setup and simulation	**Page no:** 26-34 **Section**[5.2, 5.4]
				We used Google Colab,Python programming language, Jupyter NoteBook as IDE, GPU, windows 10, Google doc & sheet which is engineering practices that covers **K6.**	**Page no:** 12 **Section**[3.3]
				For our thesis we studied existing literature which are written on Bangladeshi education system, online education system and existing models with similar goals like improving online education we do comparison between existing works, gap analysis that covers **K8 .**	**Page no:** 5-7 **Section**[2.2, 2.3]
2.	**EP2: Range of Conflicting Requirements**	No	CO2		

3.	**EP3: Depth of analysis required**	Yes	CO2	Towards improving online education system it is difficult to come up with a definitive solution for our thesis because of the volume and variety of students data. And also the solution vary from student to student. For the soluation we required depth of analysis. We carefully considered a number of options and carried out a thorough investigation before selecting a particular algorithm	**Page no:** 15-25 **Section** [4.2,4.3,4.4]
4.	**EP4: Familiarity of Issues**	Yes	CO3	To improve online education system in Bangladesh we had to address socio-economic disparities, identify challenges and opportunities, data collection and so on for that we need to look into various aspects with which we were unfamiliar. Specially when we collect data we have to follow data protection law, user privacy and terms of services of Bangladesh government. For that we had to study The Right to Information Act 2009 from Bangladeshi Constitution	**Page no:** 02, 37-39 **Section** [1.3,6.2,6.3,6.4]
5.	**EP5: Extends of application codes**	No	CO3	N/A	N/A
6.	**EP6: Extends of stakeholders involved and conflicting requirements**	No	CO3	N/A	N/A

7.	**- EP7: Interdependence**	Yes	CO4	Our thesis addresses various smaller challenges such as data collection,geographical reach, educational levels, stakeholder Involvement,socioeconomic factors,technology Infrastructure. This comprehensive approach ensures the attainment of EP-7.	**Page no:** 2 **Section** [1.3]

Addressing CO11 with Complex Engineering Activities (EA) [Some or all of the following]:

SN	EA Definition	Attainment	CO	Justification	References
1.	**EA1: Range of resources**	Yes	CO11	In our research a diverse range of resources was academicutilized. Key resources included access to a wide array of academic journals, conference papers and books from databases like IEEE Xplore, ACM digital Library and Google Scholar. Educational datasets were sourced from various schools, colleges, universities. Essential software tools facilitated model development. High-performance computing resources including cloud platforms like Google Cloud	**Page no:** [12] **Section:** [3.3]

				supported the computational needs. Collaboration educational technology along with the involvement of educators and students provided valuable insights.	
2.	**EA2: Level of interaction**	No		N/A	N/A
3.	**EA3: Innovation**	No		N/A	N/A
4.	**EA4: Consequences for society and the environment**	Yes		It enhances educational access and quality especially for students in remote and underserved areas.online education reduces the carbon footprint by minimizing the need for physical infrastructure and transportation decreases the use of paper and other physical resources and promotes energy efficiency through the use of digital platforms and cloud computing.	**Page no:** [36-39] **Section:** [6.1,6.2,6.3,6.4]

5.	**EA-5: Familiarity**	Yes		Towards improving online education our thesis expands upon existing research by examining a novel approach familiarity with several key areas to ensure its success. Such as Machine Learning and Artificial Intelligence, Educational Technology, Online Education Systems It introduces new methodologies and conducts a detailed comparative analysis, showcasing preliminary terminologies and providing fresh insights into the field	**Page no:** [6-8] **Section:** [2.3, 2.4,2.5]

Addressing CO (4, 9, 10, and 12):

SN	COs	Attainment	Justification	References
1	CO4	Yes	Our thesis addresses **CO4** by integrating effective project management and financial oversight, ensuring meticulous planning, resource allocation and budget estimation for optimal resource utilization throughout the research lifecycle.	**Page no:** [12-13] **Section:** [3.4]
2	CO9	Yes	Ensuring data privacy and security and the anonymity of participants were key priorities. Compliance with data protection regulations and the ethical use of machine learning to prevent biases were rigorously maintained. Cultural and social sensitivities were respected and	**Page no:** [38] **Section:** [6.3]

			collaboration with stakeholders was conducted ethically ensuring all actions were culturally appropriate and responsible which comply with **CO9**.	
3	CO10	Yes	We collected data and we developed machine learning models, train models by using that dataset and these tasks were divided among our team members which ensures the attainments of **CO10**	**Page no:** [15-24] **Section:** [4.2,4.3]
4	CO12	Yes	Our thesis dedication to continuous learning **(CO12)** by personalizing learning experiences and integrating cutting-edge technology it empowers students to take charge of their learning paths. The system fosters critical thinking and problem-solving skills essential for adapting to technological advancements comprehensive data collection, rigorous statistical analysis, meticulous methodology development and thorough experimental results and analysis, showcasing a commitment to staying updated and refining techniques to address modern challenges.	**Page no:** [37,38,39] **Section:** [6.2, 6.3, 6.4]

TOWARDS IMRPOVING ONLINE EDUCATION SYSTEM IN BANGLADESH USING MACHINE LEARNING APPROACH

ORIGINALITY REPORT

15%	5%	7%	9%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	2%
2	Submitted to Daffodil International University Student Paper	2%
3	doctorpenguin.com Internet Source	1%
4	www.researchgate.net Internet Source	1%
5	Pramod Gupta, Naresh Kumar Sehgal, John M. Acken. "Chapter 2 Machine Learning Algorithms", Springer Science and Business Media LLC, 2025 Publication	1%
6	Sonali Mishra, K. Priyadarsini, Arpit Namdev, S. Venkataramana, Varun, Sabyasachi Pramanik, Ankur Gupta. "chapter 11 Analysis Model at the Sentence Level for Phishing Detection", IGI Global, 2024 Publication	<1%