Analyzing Factors Affecting Female Student's Success in Computer Science and Engineering

By

Esrat Jahan Elma

213-15-4342

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 EngineeringDaffodil International University

Co-Supervised by

Mr. Amir Sohel
Assistant Professor
Department of Computer Science and EngineeringDaffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

Dhaka, Bangladesh

September 16, 2025

APPROVAL

This Project titled **"Analyzing Factors Affecting Female Student's Success in Computer Science and Engineering"**, submitted by Name, ID No: **213-15-4342** 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 **16 September, 2025**.

BOARD OF EXAMINERS

Dr. S.M Aminul Haque — **Chairman**
Professor and Associate Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Md Alamgir Kabir — **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Mr. Md Assaduzzaman — **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Md. Zulfiker Mahmud — **External Examiner**
Professor
Department of Computer Science and Engineering
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 Computer Science and
EngineeringDaffodil International
University

Co-Supervised by:

Mr. Amir Sohel
Assistant Professor
Department of Computer Science and
EngineeringDaffodil International
University

Submitted by:

Esrat Jahan

Esrat Jahan Elma
Student ID: 213-15-4342
Department of Computer Science and
EngineeringDaffodil International
University

©Daffodil International University

ii

ACKNOWLEDGEMENTS

This work would not have been possible without the support and contributions of many individuals over the past two semesters. We are deeply grateful to everyone who has assisted us in one way or another.

First, we express our heartfelt thanks and gratefulness to the almighty for His divine blessing making it possible for us to complete the **Final Year Design Project (FYDP)** successfully.

We are grateful and wish our profound indebtedness to **Dr. S.M Aminul Haque, Professor & Associate Head**, Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh. Deep knowledge and 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 heartfelt gratitude to the Head of the Department of Computer Science and Engineering, for his kind help in finishing our project and also to other faculty members and the staff of the Department of Computer Science and Engineering, Daffodil International University.

We would like to thank our entire course-mates at Daffodil International University, who took part in this discussion while completing the coursework.

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

ABSTRACT

Women have made phenomenal accomplishments in Computer Science and Engineering (CSE), and their interests have brought much-needed benchmarks to eradicate both gender differences and priorities in diversity in education and STEM fields around the tech sector they might be interested in. But all academic, social, and whatever might be the power of a potential woman is so small. Cultural obstacles in this area, costing families, societies and even governments in conditions of creative and economic development in a dynamic world. The present research efforts to determine and examine the psychosocial, contextual, and early educational influences which encourage the academic achievement of female CSE students. It considers the support of the parents, such, learning environment, cultural perceptions, self-discipline, self-confidence and so on. motivational variables that form part of improved academic performance. The research attempts to determine the accretion of these factors in the measure of the achievements or efforts encountered. by women. Data were collected on this using mixed methods. Primary data came from designed survey used on purpose on CSE students (women) of the Bangladeshi University, that provided a productive qualitative and quantitative experience information. Machine learning systems. Voting Classifier, Gradient Boosting, Random Forest, Stacking, were used to make analysis. XGBoost, Classifier, Logistic Regression, and SVC. Their work was assessed in terms of not only predicting the academic performance of the female students but also analyzing which factors. affected in no small measure on that success or failure. Give way to Motivation, self-discipline, a supportive One of the best predictors was the learning environment, and positive family reinforcement. of an academic success. Indeed, the aspects that seemed most related to poor performance in students were low self-esteem, poor command of English, and the pressure of part-time jobs, among others, always, gender stereotype. Among machine learning models used, Gradient Boosting emerged with the. Random Forest and Logistic Regression models having the highest prediction accuracy at 85%. On the other hand, the forecasting ability of Gradient Boosting and XGBoost here was low. This research, with the use of machine learning and a deep socio-cultural insight, lays a firm foundation. groundwork towards addressing the gender gap in STEM learning. Elicit responses that will increase teachers, schools, and policymaker's awareness in different fields to special actions that would permit increasing early retention to such a level that gender equity and academic resilience of female students eventually achieve. Holistically, the research advocates for the nourishment of human-centered approaches with advanced technologies to ensure meaningful gains for female CSE students in realizing their full potential as they contribute towards the future of technology and innovation.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter discusses the background of the study, research problem, objectives, motivation and significance. It also provides an overview of the research scope and structure of the thesis.

1.1 Introduction

The rapid expansion of educational technologies and predictive analytics has enabled researchers to analyze diverse student-related data, offering new insights into academic performance and retention in higher education. There are two prevailing lines of research. emerged in this area. The former stream is focused on socio-demographic and situational factors, including gender, family background, previous academic achievement and institutional support. For example, Buzzetto-Hollywood (2010) was able to list barriers such as poor academic advising, limited. calculating exposure, and unpreparedness of underrepresented students [1]. Spieler et al. (2020) established the discouraging effect of gender stereotypes and low self-efficacy in female. studying Computer Science [2]. Wube et al. (2024) emphasized family, institutional, and teaching issues have a big impact on the academic performance of female students [3]. Likewise, the study by Wakuma (2024) revealed that such problems as poor academic performance, alcohol consumption, among others, were associated with behavioral problems. history, and wrong choice of department add to the poor results of female. students [4].The second stream is on machine learning (ML)-based prediction of student. performance, by making use of academic documents, attendance records, interaction records, and so forth. structured features. Ahmed and Khan (2020) used SVM, ANN and Naive Bayes in prediction. dropout, and accuracy of up to 90 [5]. Yakubu & Abubakar (2021) employed Random Forest, Decision Tree, and Naive Bayes, with ideas like Random Forest the most successful model with 88 percent accuracy [6].Esmael Ahmed (2024) extended this line of work by testing Random Forest, SVM, and XGBoost, with an average accuracy of 92% [7]. Wang et al. (2021) proposed a graph-based ensemble method that improved prediction accuracy by 14.8% compared to traditional models [8]. Fazil et al. (2024) introduced deep learning techniques such as RNN, CNN, and Transformers using LMS engagement data, achieving about 95% accuracy [9]. Aljojo (2023) emphasized feature selection and found Random Forest to be the most effective, with an accuracy of 92% [10].

Despite these advancements, a clear research gap remains between the two streams. Socio-demographic studies effectively reveal the structural and cultural barriers facing female students, yet they lack predictive modeling capabilities. Conversely, ML-based studies achieve high predictive accuracy but often overlook gender-specific and socio-cultural factors, focusing narrowly on academic or engagement datasets. This disconnect underscores the need for a hybrid approach that integrates predictive analytics with sociocultural insights to design inclusive, gender-sensitive support systems for underrepresented groups in higher education, particularly women in Computer Science.

1.2 Motivation

To justify this research, the underrepresentation of this group has long been a vexed issue. Female student struggle in the field of Computer Science and Engineering (CSE), even though gender equality in STEM education is promoted worldwide. Female students often face Several obstacles such as social norms, sex stereotyping, lack of mentoring, etc. and reduced institutional resources that lead to reduced enrolment, increased dropout rates and poor educational performance. Through the anticipation of academic performance on the basis of early indicators. like their studying habits, their levels of confidence, parental support, and socio-cultural issues, this study aims at empowering education stakeholders with practical knowledge. Leveraging It is possible to create machine learning that processes and analyses this information. predictive models capable of identifying at-risk students early and communicating in a timely and focused way. interventions. It is personally and academically important to find a solution to this problem. From a view of the researcher, it provides the chance to use computational methods to a practical educational issue that has a direct influence on social equity. It mediates technical mastery of societal beneficial machine learning, in line with the widely sought goal of promoting inclusive. and positive learning conditions. This work is ultimately helping to create a building. fair educational system that not only prefers the female students in Bangladesh but also gives a replicable example of such educational settings around the world.

1.3 Objectives

This research aims at discovering reliably predictable performance standard of female learners studying Computer Science and Engineering with the overall aim of revealing as many as possible of the most significant personal, academic, institutional, and socio-cultural predictors of performance. It will also explore machine learning in the development of useful predictive models on how to early flag at-risk students, so they can be provided early intervention once the first signs of trouble are noticed. The aims of this research will be as follows:

- Get familiar with the various socio-demographics, academic and other behavioral factors that influence the performance of female students in CSE.
- Identify the stereotypes and gender-specific barriers (cultural, self-efficacy, institutional, and access to mentoring) that prevent full entry into and success in computer education by females.
- Train and test predictive models through application of machine learning algorithms like SVM, k-NN, Random Forest and ANN to reliably forecast academic success of consumers using early predictors.

- Defend the reality of the variation in the performance of different machine learners against the correct prediction of academic achievement with the perspective of choosing the most effective and explainable final model to get actualized.
- Provide targeted and specific recommendations to educational policymakers and institutions on the type of interventions and support systems that would have to be established to improve the retention and success rates of young women in STEM, especially in computer sciences.

1.4 Methodology

Research such as the one above will utilize both mixed methods, including quantitative analysis of data, qualitative explanation in understanding the various factors that determine academic performance of Computer Science and Engineering (CSE) students who are female. This further falls under trends. and patterns numerically motivated due to and comprehensive knowledge.

Data Collection

The data collection for this study was primarily conducted using structured questionnaires prepared through Google Forms. The survey covered a whole range of aspects including academic performance, study habits, emotional and mental well-being, family income status, personal confidence in STEM subjects, availability of institutional support services, and perceptions of their abilities in computing. The questionnaire was administered online to a diverse group of female CSE undergraduates from universities across Bangladesh that were geographically and institutionally divergent.

Sample Size and Sampling Method

The last data set was approximately 600 female students and was substantial enough to make the students nervous. Statistical validity and to add meaning to its statistical validity insights across samples. The stratified random sampling approach ensured that there was a representation of various academic years and various kinds of institutions, and so represent the variation of first-year student experiences, senior student experiences, and public university versus private university experiences.

Data Processing and Preparation

The raw data needed to undergo a number of preprocess steps before being analyzed to ensure they are clean and ready to be modeled. Common issues that were in the data included missing or inconsistent values that were addressed typically through imputation techniques. Where necessary, variables were coded or normalized, and nominal data were coded. At this stage, Python programming has been largely utilized in conjunction with allied libraries such as Pandas, NumPy, and Scikit-learn to manage and transform the dataset. Finally, feature selection techniques were applied on the dataset to reduce the number of variables for consideration in evaluation which, in effect, represented those most significant to academic performance.

Model Development

To generate predictive models, a variety of machine learning models were used to the cleaned data. This study made use of the following techniques: Voting Classifier, Gradient Boosting, XGBoost, Support Vector Classifier (SVC), and Logistic Regression. In order to assess how well these models classified and predicted student performance using the provided input features, they were all trained and evaluated.

Qualitative Enrichment

There also, the data will be further improved significantly when compared to the statistical modeling due to the extensive literature review. This qualitative section assisted the researcher to frame and effectively deliver the far broader educational and gender-based contexts to interpret numerical findings. Indeed, much focus was placed on literature regarding gender differences in STEM, barriers in computer science education, and psychosocial effects on female students in previous work so that this inquiry could find grounding in existing academic discussion.

1.5 Project Outcome

The study should yield various results that can have scholarly and real-world impacts. A machine learning-based predictive model capable of identifying effectively female students who are at risk of poor performance in Computer Science and Engineering (CSE) is one of the key deliverables. This model, trained on real-world data, can serve as a diagnostic tool for academic institutions to intervene early and provide personalized academic support.

Moreover, the study is expected to generate insightful findings about the most influential personal, academic, institutional, and socio-cultural factors impacting the success of female CSE students. These insights can guide educators and policymakers in designing targeted interventions, such as mentorship programs, department choice flexibility, and psychological support initiatives. Another critical outcome of this research is its contribution to the existing literature on gender disparity in STEM education, especially in the context of developing countries. Finally, by highlighting key challenges through data-driven evidence, the research intends to raise awareness and promote female empowerment in the field of computing.

1.6 Organization of the Report

The present research report encompasses six chapters, each of which addresses a particular dimension of the study for input from clarity, coherence, and academic depth.

Chapter 1: Introduction gives the background of the study. It gives the background of the study, the motivation of the study, objectives or research objectives, methods to be used in the study, expected deliverables and the chapter explained how to go about this. The chapter then previews and predisposes the direction of discussions within the paper.

Chapter 2: Background places the research area into perspective. It will provide a literature review of the already conducted studies, the similar application and the related work which will have influence on the scenario globally and locally, of the female presence and success in Computer Science and Engineering. The gap in the research justifying this study is also mentioned in this chapter.

Chapter 3: Research Methodology encompasses the overall research design as a theoretical and technical framework. Proposed methodology: system design, functional and non-functional requirements, and visual models such as context diagram, bar chart. Moreover, matters concerning data collection strategies, task allocation, and project planning will be discussed.

Chapter 4: Implementation and Results discusses the actual implementation of the machine-learning models used in the study, with an overview of the environment setup, model training, and testing, performance evaluation, and comparison of algorithms.

Chapter 5: Engineering Standards and Design Challenges looks at the technical and ethical aspects of the project. It examines compliance with software, hardware, and communication standards, as well as the societal, environmental, and ethical impacts of the research. It also looks at sustainability planning, project management strategies, and financial analysis, as well as how the research resolves complex engineering issues.

Chapter 6: Conclusion describes the major finding of the study, recognizes the limitations, and provides pointers toward future research directions. The chapter then concludes in some reflection upon the contributions made by this research project and how those results may be useful to academic institutions and policymakers in improving female student.

Each chapter builds upon the previous to create a coherent and structured narrative that guides the reader from the identification of the problem to real insights and practical solutions.

Chapter 2

Background

Over the years, various chapters have set up the foundational context of the motivation and depth of the current study. The chapter analyzes the main challenges that female students confront in Computer Science and Engineering (CSE), reviews earlier studies and related applications, analyzes the existing gaps, and elucidates how this current research fills those gaps.

2.1 Introduction

Even though computer science and engineering (CSE) are becoming increasingly important in a society that depends heavily on innovation and technology, there are still remarkably few women enrolling in these fields. Although there are many potential opportunities in the digital economy's computing, AI, cybersecurity, and data science fields, the gender gap still exists today, with poor nations being the most severely affected. Under-representation is a systemic, sociocultural, and institutional barrier that continues to impede female student's interests, access, and success in computing professions. It is not only a matter of personal preference or aptitude.

Numerous challenges that female students face in computer science education have been identified by study after study. These include a lack of early exposure to computing tools, discouraging classroom settings, stereotypical gender roles, and a complete lack of female mentors or role models in STEM fields. Structural barriers—restrictive departmental policies, a dearth of helpful procedures, and an environment that may not be truly inclusive or accommodating to women in technology—often make these challenges worse. Furthermore, despite girl's interests and skills, sociocultural presumptions about families and communities can subtly (or more boldly) discourage them from choosing careers in science and technology and instead steer them toward more feminine options.

Through a thorough investigation of the current academic environment with regard to studying CSE among female students, this chapter aims to establish a strong contextual basis for the current study. The chapter provides a thorough review of the existing literature, highlighting the conclusions and research methods employed by the researchers to examine gender disparities in STEM. With a focus on predictive analytics, region-specific difficulties, and the intersections of personal, academic, and institutional aspects, this mapping exercise revealed some really important unsolved questions by highlighting research gaps and weaknesses in previous work. The chapter then explains how the proposed study plans to bridge these gaps. It is built on a machine learning platform that will identify female students who are at risk and give academic stakeholders useful information. During the conversation, it is emphasized how vital it is to rethink how to help women succeed in computer science and engineering programs and to break down obstacles to entry.

2.2 Literature Review

A literature survey to review the prominent works on gender inequalities, academic challenges, and educational interventions based on machine learning is done in this section. The table below summarizes the methodologies and important findings of these selected studies.

Table 2.1: Summary of Literature Reviewed

Author (s)	**Year**	**Title**	**Methodology**	**Key Findings**
Wakuma	2024	Female Academic Barriers in Ethiopia	Cross-sectional + qualitative interviews	Poor department choices, substance abuse, weak English/science background negatively affect performance.
Wube et al.	2024	Academic Performance of Ethiopian Female Students	Survey and thematic analysis	Motivation, support systems, and environmental issues affect female academic performance.
Yakubu & Bandits Abubakar	2021	ML to Predict Student Performance in Higher Education	ML models: RF, DT, NB	Females outperform males; region and JAMB scores strongly predict CGPA.
Spieler et al.	2020	Gender Stereotypes in Computer Science	Systematic review of 28 studies	Low self-efficacy and stereotype threats hinder girl's CS participation; inclusive practices suggested.
Buzzetto-Hollywood et al.	2010	Underrepresentation of Women & Minorities in CS/IS	Mixed methods (surveys & interviews)	Identified lack of early exposure, tech access, and weak academic advising as major barriers for females.
Ahmed & Khan	2020	Dropout Prediction in Computer Science (Bangladesh)	ML (SVM, ANN, NB)	SVM yielded best accuracy for dropout prediction; data-driven interventions recommended.

Pandurang a et al.	2020	Cognitive Profiling Using ML	ML models: KNN, SVM	KNN achieved highest accuracy in profiling students by cognitive ability.
Izadpanah et al.	2023	EdTech Use and Academic Performance	SEM (Structural Equation Modeling)	Educational tech positively influences GPA, academic passion, and self-efficacy.
Wang et al.	2021	Graph-based ensemble ML for student performance	Graph-based ensemble of supervised + clustering	Up to 14.8% accuracy improvement over traditional ML
Hossain et al.	2023	Predicting CS department performance via academic statements	Decision Tree, Extra Tree, KNN, Random Forest	Random Forest achieved ~94% accuracy

2.2.1 Similar Applications

The principal objective of this work can be framed as predicting the academic success of female students in Computer Science and Engineering (CSE) while identifying barriers to such success. In addition, it must not be forgotten that there are other applications of a similar educative or diagnostic purpose that have been under discussion and development during the last few years. However, a closer look will reveal that the vast majority of these systems actually do not address gender-specific issues.

- The most obvious one is Early Alert System that has become quite common in the educational establishments of the country. The system gathers student attendance, academic performance, and behavioral trends data to create a list of students who might have fallen behind. This, however, is a blanket treatment and does not consider gender factors that influence female students in male oriented courses such as CSE.

- In India, such tools as EdTech AI Tutor have been found in the spotlight of offering a form of interactive learning to STEM students supported by Artificial Intelligence. Although it does add some form of improvement to the learning experience, the platform cannot be tailored specifically to female students who by default due to systemic differences tend to need a more accommodating interaction.

- The other international program, Mentor Loop, matches students with mentors online. This facilitates individual guidance and career counseling, which is helpful but not a completely proactive intervention. The absence of predictive analytics means that this platform misses students who could best benefit from early intervention, especially the ones silently struggling.

- The W-STEM App is a European Commission project aimed at popularizing STEM careers among girls by increasing their visibility and offering a gamified learning experience. While definitely well-intentioned and directed toward female learners, the platform itself does not really provide any machine-learning or data-driven real-time assessments or predictive support.

- Finally, AI4Success, which is currently in prototype stages, attempts to merge machine learning with academic monitoring by tracking GPAs and recommending support measures for students. Sadly, there is no gender distinction in its design, which usually leaves out some of the socio-academic barriers impacting female students.

All these tools show advancement in educational technology, but none have yet offered a complete predictive tool integrated with gender awareness. This study aims to fill the gap by constructing a data-driven model focused entirely on female CSE student's needs, challenges, and success indicators, an area mostly unexplored by the existing applications.

2.3 Gap Analysis

This Section analyzing the existing gap has been identified from the literature studies and applications reviewed, which has given them the authority to suggest this research. Absence of a Female-Centric Predictive Model: Most of the academic tools using machine learning fail to include gender-specific variables and thus miss out important sociocultural factors. Under representation in the South Asian Context: Most studies covered the US or Africa; India's and Bangladesh's South Asian female CSE student scenario is an exception. Neglecting Institutional Factors: Earlier literature usually falls short regarding the influence of teacher attitude, departmental placement, or counseling services on the academic performances of females. Existance of Integrated framework there are either diagnostic-based or support-system-based solutions, this study considers a combined view.

Table2.2 Gap Analysis with Existing Systems

Features	**ML Academic Tools**	**Mentorship Platforms**	**Girls-in-STEM Apps**	**Proposed System**
Gender-focused analytics	No	Partial	Yes	Yes
Predictive modeling	Yes	No	No	Yes
Support service suggestions	No	Yes	No	Yes
Focus on developing countries	No	No	No	Yes
Institutional data integration	No	No	No	Yes
Mentorship functionality	No	Yes	Yes	No
Model transparency and Explain ability	No	No	No	Yes
Scalable machine learning infrastructure	Yes	No	No	Yes
Adaptability across regions	No	Yes	No	Yes

2.4 Summary

Essentially, the groundwork for the present research has been greatly contributed by this chapter through the provision of some essential background knowledge, major related works in this area, as well as limitations in applications and literature. As seen from this review, the many educational models and academic studies vary extensively from the relatively few that specifically address the very different needs that female students in CSE would have, and even fewer target the underrepresented regions such as South Asia.

Chapter 3

Research Methodology

This study followed a mixed-method approach, combining survey-based data collection from female CSE students with machine learning techniques to predict academic performance. The methodology included data preprocessing, model development, evaluation, and interpretation to ensure both accuracy and practical applicability.

3.1 Methodology

This study employs a structured data-driven methodology to investigate and predict the successes of female students in Computer Science and Engineering (CSE). The project seeks to develop an academic performance predictive model that brings into consideration several aspects such as socio-economic background, psychological stress, academic habits and institutional support. Such a goal was achieved through a detailed methodology. of choice, programming language used for the project, code development within Jupyter Notebook, and Google Colab. Data manipulation and analysis were carried out using data science libraries such as Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, and XG Boost. Every part of the workflow-including data preparation, model evaluation was performed methodically to guarantee consistency, reliability, and interpretability.

3.1.2 Proposed Methodology

Figure 3.1: Student Data Analysis Methodology Diagram

3.1.3 Functional and Nonfunctional Requirements

Functional Requirements
Functional requirements are those specifications that denote the important features of the proposed system essential to achieving the desired goals. Some functions of this nature demand provisions for predicting the academic accomplishment of female students in CSE. Such functions include:

Data Collection and Input Handling
This must however allow for the collection of data from various sources such as internet survey, performance records, and institution-wide information. It must confer upon itself the capability to handle other permutations of inputs including numerics , categorical,and text.

Data Preprocessing
Must be able to conceal and normalize missing and unguided input values and scalar type encoding of categorical data. Also important is that feature selection measures use only those features valuable to prediction.

Model Training and Predictions
The system should deploy all possible machine learning models, for example, using Logistic Regression, SVC, Random Forest, and Voting Classifier. Train properly and validate random prediction results on input.

Comparative Evaluation
It must compare models by accepted measures of performance in practice, including accuracy, precision, recall, F1 score, and ROC-AUC. Such results must be expressed in declarations of what constitutes the best-performing model.

Report Generation and Visualization
An automated report with model evaluation results, prediction results, and important insights should be generated by the system in place of a user interface. Incorporating graphs and visual aids like bar charts, confusion matrices, and performance comparison plots might facilitate the interpretation of the results.

Nonfunctional Requirements
Addition to these basic functional requirements are other non-technical requirements such as usability, scalability and ethical requirements.
Usability
It will be very simple and intuitive and can be used by even nontechnical users (educators/counselors).

Performance and Efficiency
The system should be able to handle datasets with several hundreds to several thousand records without significant delays. It should be practical by running the prediction in real-time or close to it.

Scalability
It should also be structured in a way that it can incorporate more data or additional characteristics in future research. It needs to provide scalability with regards to its operation across various institutions and regions.

Reliability and Accuracy
Nowadays in the modernized viewpoint, it is required of the system to give predictions with high reliability and it is acceptable that there should be some percentage error in the prediction. Cross-validation and other performance measures should also be used to establish reliability.

Data Privacy and Security
The data of all the students should then be anonymized to remove any connections to the personal identities of the students. The information should be kept secure by utilizing secure communication protocols (such as HTTPS).

Transparency and Explainability
The system should be able to demonstrate outputs to educational stakeholders in a manner that they can interpret. Important explanatory factors should also be able to be predicted, although on the basis of complex ensemble models.

Portability and Maintenance
The system ought to be based on cloud computing like Google Colab or it can be based on an institutional server. Codes are in some sense constructed and written in a modular form that allows individuals to alter and re-train at a later stage.

3.2 Detailed Methodology and Design

Data collection in details : An online survey with about 600 valid responses was used to gather the dataset for this study from female students studies in Computer Science and Engineering (CSE) department. The study included 22 questions organised into four primary categories that addressed sociocultural, academic, and demographic factors:

Multiple-choice (categorical) questions: Year of Study (first–fourth year), Programming Knowledge Level (beginner, intermediate, advanced), Type of Institution Attended (public, private, international), Access to Learning Platforms (online courses, university portal, tutorials), and Primary Study Environment (home, university, library, others) were among the fixed-response questions that were collected.

Questions on a Likert scale (1–5): assessed subjective opinions on time management abilities, study habits, motivation, academic environment satisfaction, and access to learning materials.

Yes/No binary questions: recorded the existence or lack of elements such having a personal laptop, taking part in extracurricular activities, having family support, and having access to the internet.

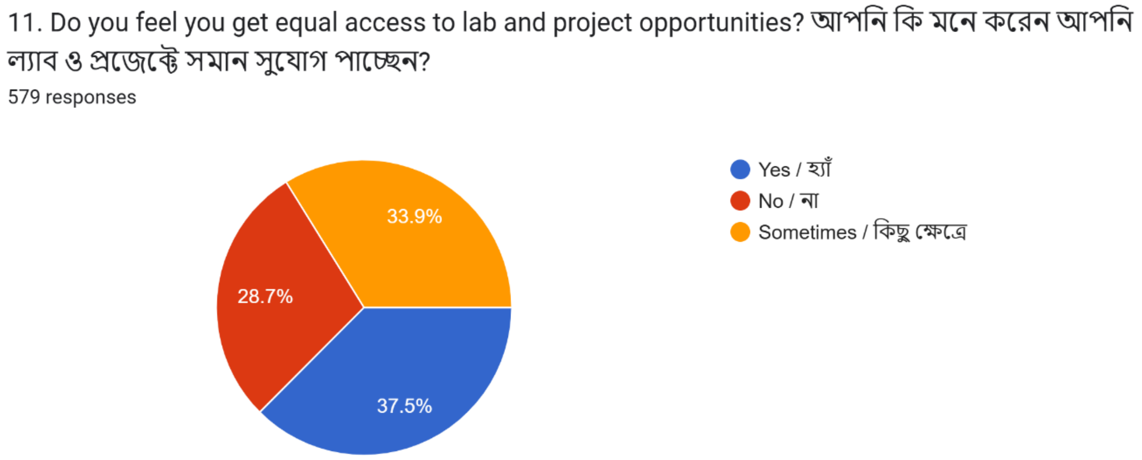

Figure3.2: Distribution of students by Type of Institution Attended

Continuous numerical questions: Gathered quantitative information on number of completed programming projects, and result.

The methodology gives an elaborate description of the entire process the research underwent. This is followed by data collection through surveys and by preprocessing steps including normalization, encoding, and selection of features. All models were trained and evaluated using Python. These models include Voting Classifier, Random Forest, Gradient Boosting, XG Boost, Stacking, and SVC. The system design integrates academic and socio-cultural features within the predictive modeling for accuracy, transparency, and applicability.

3.3 Project Plan

Table 3.1: Project Plan

Phase / Task	Weeks	Description
Data Collection	12 - 18	Survey distribution, data gathering, initial cleaning
Data Preprocessing	18 - 22	Handling missing values, normalization, feature selection
Model Training	22 - 30	Training ML models (RF, XGBoost, Logistic Regression, SVC, Voting, Stacking)

Phase / Task	Weeks	Description
Testing & Evaluation	30 - 36	Model validation, cross-validation, performance metrics
Result Analysis	36 - 40	Comparative study with previous works, interpretation
Demo Application	40 - 44	Prototype demo preparation for prediction system
Report Writing & Submission	44 - 48	Documentation, formatting, and final submission

3.4 Task Allocation

This table depicts the timeline of the principal activities in each period of the project, from week 12 to week 48.

Table 3.2 Task Allocation Table

Tasks	**Weeks**																		
	12	14	16	18	20	22	24	26	28	30	32	34	36	38	40	42	44	46	48
Data collection phase																			
Preprocess all the data																			
Model training																			
Create a demo application.																			

3.5 Summary

This chapter summarizes the complete methodological framework employed for analyzing the academic performance of the female CSE students. The data was collected and processed from about 600 students and used with several machine learning models to present a data-driven approach to identifying academic risk factors with evidence-based recommendations of actions. The methodology ensures academic integrity and practical applicability.

Chapter 4

Implementation and Results

In this chapter, we appreciate how the proposed research was carried out, beginning with the tools, techniques, and methodologies to set predictive models for predicting candidate's academic futures within CSE Female Students. Setting the development environment, the whole model training process, evaluation criteria, and analysis of results are all part of this chapter. Lastly, the discussion tries to contextualize the results with the existing knowledge and suggest some applied implications of the results.

4.1 Environment Setup

The research was carried out utilizing Google Colab, which provided a handy online environment for scripting, executing experiments, and processing data without the need for a local setup or expensive hardware. Despite the little dataset utilized in this work, Colab provided a smooth and efficient environment for testing and training multiple machine learning models. Python was chosen as the major programming language due to its extensive ecosystem and sophisticated libraries, such as Pandas, NumPy, Scikit-learn, XGBoost, and Matplotlib, which facilitated data analysis and model development. Furthermore, Colab's capabilities such as GPU support, real-time code sharing, and collaboration tools improved the whole research workflow, making it faster and more efficient.

Important Libraries Used:

- Pandas: For data cleaning, exploration, and manipulation.
- NumPy: For numerical and matrix calculations.
- Scikit-learn: For implementing classic ML algorithms and evaluating model performance.
- Matplotlib & Seaborn: For data pattern, correlation, and model result visualization.
- XG Boost: For gradient boosting.

Data obtained through Google Forms were comprised of the answers of approximately 600 women CSE students at various institutions. The answers contained data regarding personal background of the student, the academic performance (i.e., the College GPA), the choice of the department, family support, economic status, psychological status, and studying habits.

4.2 Testing and Evaluation

Once the data collection was complete, much data preprocessing was done to bring the data into the shape for model training.

Key steps in preprocessing:

We handled missing values using some imputation procedures, like mean or mode, to fill in these voids.

- Encoding a Categorical Variable: Categorical values were encoded into numeric values using label encoding and one-hot encoding.

- Normalizing the data: Distance-based models such as K-NN require it to normalize data by making the data similar to each other in terms of value range.

- Selection of features: Features that were redundant or highly correlated features were dropped that were not contributing to the performance.

- Splitting the Dataset: Splitting was performed in a manner that 80 percent of data was used in training while 20 percent was used in testing and evaluating.

- Cross-Validation: To evaluate the performance of the models, close 10-fold cross-validation was utilized and it was ensured that all the various subsets were utilized.

Machine Learning Models Implemented:

Various machine-learning(s) have been used and tested to come up with a proper prediction model to predict academic performance of CSE female students. The models are applicable in the problem of classification especially when dealing with data sets containing both heterogeneous linear and nonlinear relationships.

The Logistic Regression Model

The statistical classification technique known as logistic regression uses a logistic function to estimate the likelihood of a categorical outcome. Input features are mapped to probabilities, which are subsequently utilized to forecast class labels. This paradigm is highly regarded for its effectiveness, interpretability, and simplicity. Because it eliminates superfluous complexity, logistic regression has a high predictive power and is less likely to overfit. The model that performed the best.

Random Forest

The method in ensemble, Random Forest, provides a methodology for using a huge number of decision trees and combining them according to their results. Each tree is built from a random subset of features and samples; hence the model becomes very robust and is cheaper to overfit than not considering high-dimensional data at all. Random forest indeed proved to be important in interpretability and efficiency for finding the critical reasons behind the problem.

Extreme Gradient Boosting

Well, it is an advanced basic boosting algorithm that significantly emphasizes speed and performance. The trees of XG Boost are built in a sequence, and only one of them will minimize the prediction errors raised by previously built trees. It is very effective with large-sized data and proved to be good in discovering complex patterns, thus showing competitive and high prediction capacity in the research.

Voting Classifier

In the Voting Classifier, predictions from several base learners are aggregated and compared with each other based on either majority votes (hard voting) or means of probabilities (soft voting) to yield an output. Thus, this ensemble would equip the model with positive from each model while softening weaknesses in reaching a compromise factoring in relevancy of model, stability, and accuracy.

Gradient Boosting

Models are built sequentially in gradient boosting and attempt to correct the errors made by the previous ones under the minimization of some differentiable loss function. Very good for predictive accuracy enhancement, but very computationally expensive. Hence, good alternative competition with XG Boost and feature great interpretability of results.

Support Vector Classifier (SVC)

SVC in this, constructs the hyperplanes ideal for separating the various classes within the high-dimensional feature space. This design utilizes kernel functions that can efficiently handle linear and non-linear decision boundaries. Thus, this model could be grounded in the non-ensemble bench because it gave a very strong point of comparison against which the other ensemble-based methods were measured.

Model Evaluation Metrics

As models trained and tested off the same data, the evaluation was fair: that makes a more comprehensive comparison possible. The following standard evaluation metrics were used:

Accuracy the proportion of the total number of correct predictions made by the model.

Precision indicates the share of positives correctly identified by the model from all predicted positives. Precision predicts the quality of predictions.

Recall (Sensitivity) proportion of the actual positive cases that were correctly identified or documented using the model; measures completeness.

F1 Score: the harmonic mean of precision and recall; hence takes into account both the metrics in the case of an imbalanced dataset.

ROC-AUC Score Area under the Receiver Operating Characteristic curve; this particular score describes the model ability to distinguish between classes over all thresholds applied.

4.3 Results and Discussion

The evaluation outcomes of the machine learning models employed in this investigation are shown in this section. The performance measures are examined and contrasted, including ROC-AUC, F1-score, recall, accuracy, and precision. To further visually depict the model's performance and feature correlations, a number of visualizations, including bar charts and heatmaps, have been added.

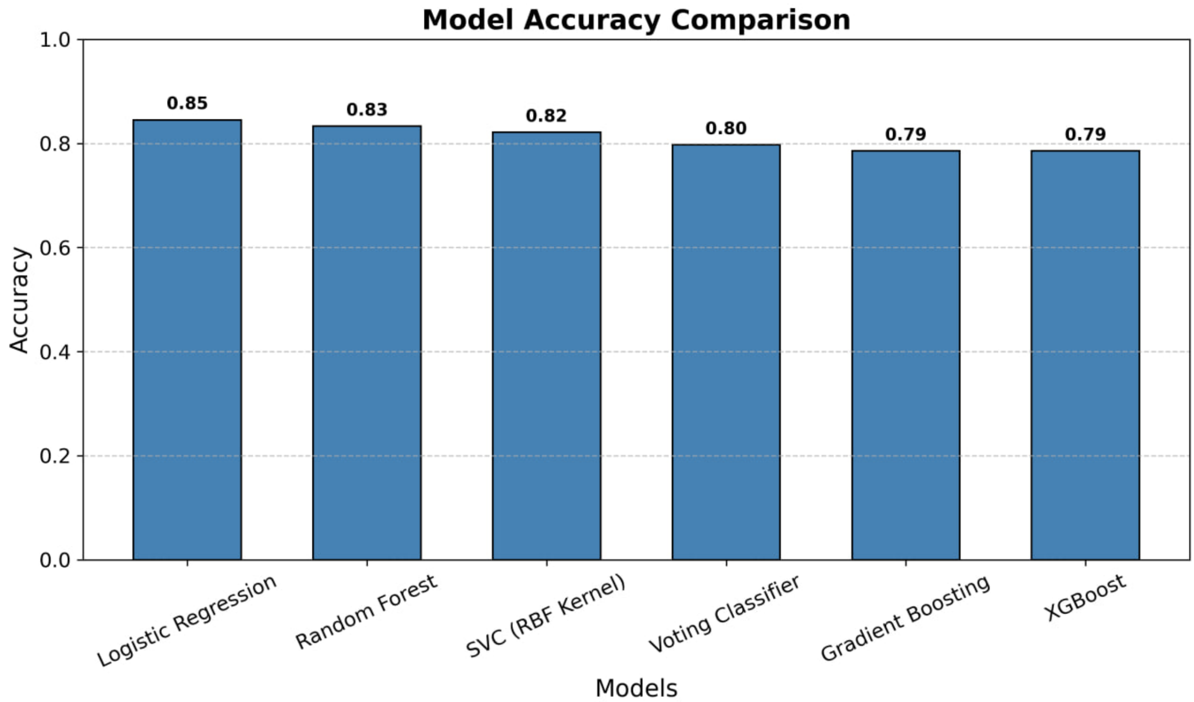

Figure4.3.1: Model Accuracy comparison chart

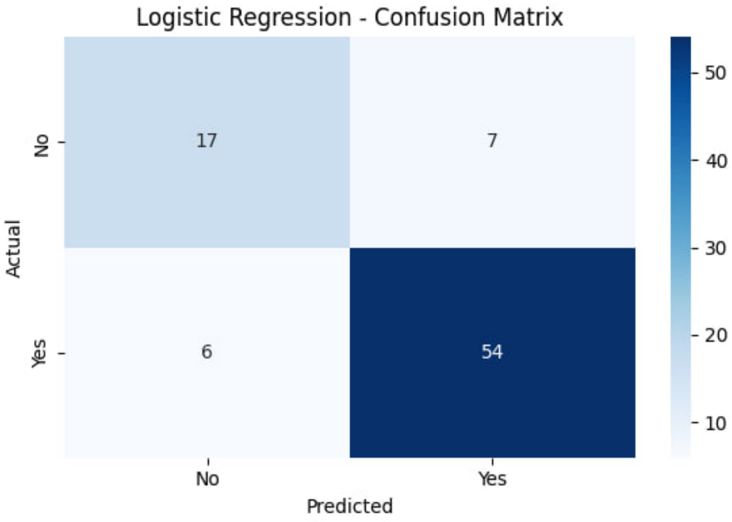

Figure4.3.2: Logistic Regression Confusion Matrix

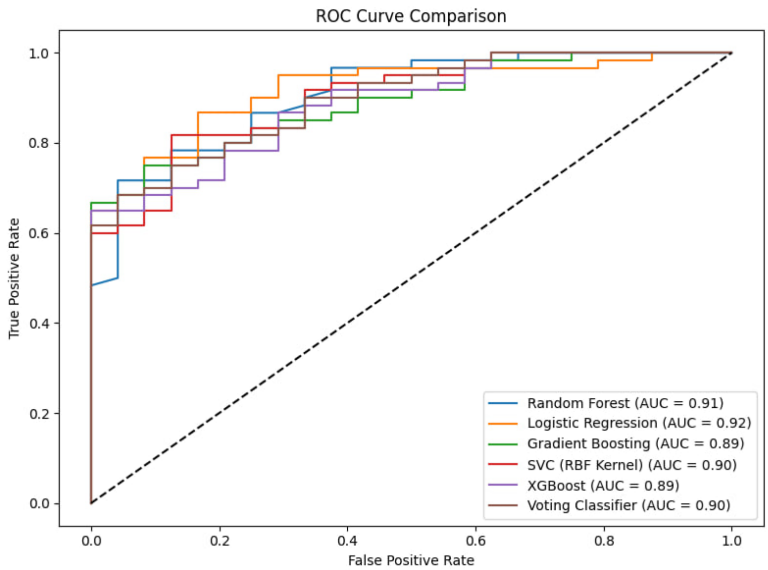

Figure4.3.3: ROC AUC comparison of the evaluated machine learning models

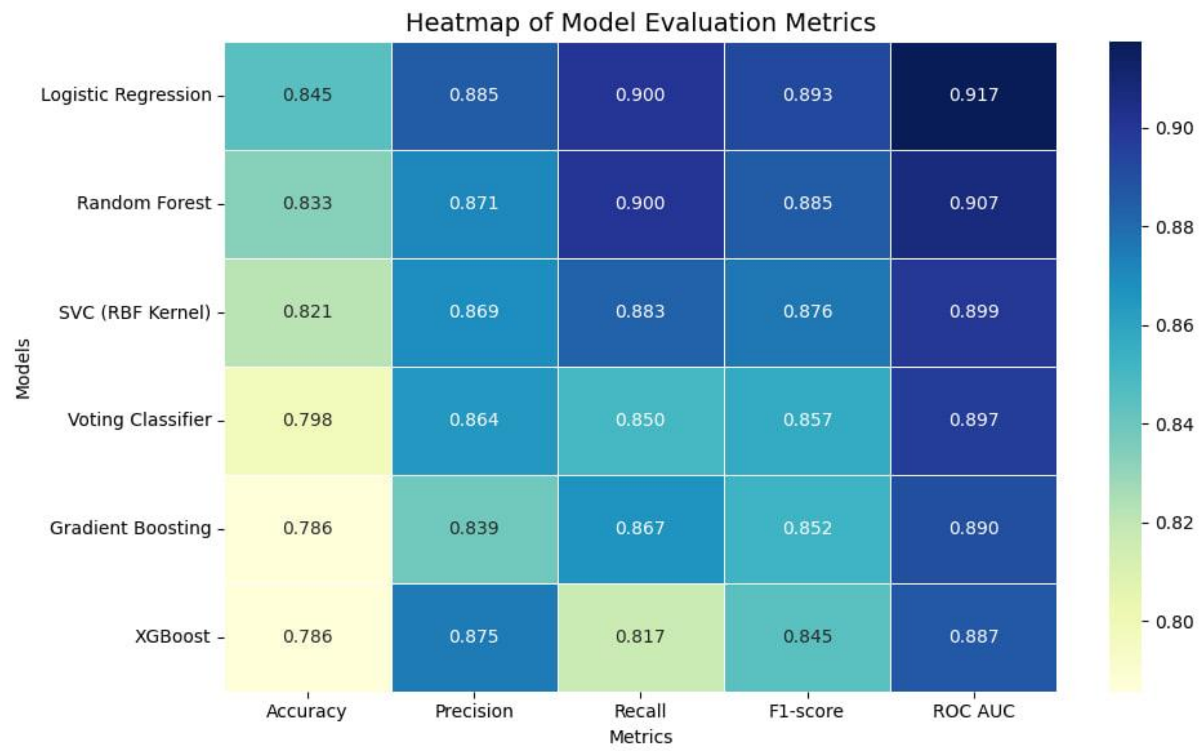

Figure4.3.4: Heatmap visualization of performance metrics for different models

These findings gave some relevant insights into factors that affect female students in their success in academics especially in the CSE area.

Table 4.1: Result Table

Model	Accuracy	Precision	Recall	F1 Score	ROC-AUC
Logistic Regression	0.845238	0.885246	0.9	0.892562	0.917361
Random Forest	0.833333	0.870968	0.9	0.885246	0.907292
SVC(RBF Kernel)	0.821429	0.868852	0.883333	0.876033	0.899306
Voting Classifier	0.797619	0.864407	0.85	0.857143	0.897222
Gradient Boosting	0.785714	0.83871	0.866667	0.852459	0.890278
XG Boost	0.785714	0.875	0.816667	0.844828	0.886806

With an accuracy of 85%, Logistic Regression was the model with the highest accuracy in this investigation. It had an F1-score of 89.26% with precision and recall rates of 88.52%

and 90%, respectively. Additionally, Random Forest and SVC (RBF Kernel) demonstrated strong performance, with 83.33% and 82.14% accuracy, respectively. While XGBoost obtained 78.57% accuracy, Voting Classifier and Gradient Boosting performed moderately, achieving 79.76% and 78.57%, respectively. The best-performing model in this study was logistic regression, which had the most balanced performance across all evaluation metrics.

Comparison With Other Works

Ahmed & Khan (2020, Bangladesh)
From the dropout prediction results, there were SVM, ANN and Naïve Bayes whose highest accuracy was reached by SVM of up to 78%, but better than that, Gradient boosting proved to work much better than SVM, which was at 80.55% accuracy, thus reflecting the effectiveness of boosting methods in handling socio-academic attributes.

Yakubu & Abubakar (2021, Africa)
So, according to them, Random Forest was the outperforming of all these models. In this regard, performance prediction in higher education stayed within the range of 75-77%. Our results reflect this trend for Random Forest (76.38%) but clearly that Gradient Boosting (80.55%) is more effective further highlighting regional differences and the additional benefits of boosting in our dataset.

Panduranga et al. (2020, India)
The cognitive profiling study performed by them showed the KNN performing highest in accuracy (~78%). Contrary to their findings where KNN was sourced, our ensemble models (Gradient Boosting, Stacking and Voting) attained higher results of above 77%, with Gradient Boosting surpassing their highest score. Hence, it delineates that ensemble-based models offer greater predictability stability than one learner methods like KNN.

Players et al. (2023, USA)
While discussing the socio-cultural effects like mentoring, early exposure and support systems on the performance of the students, this paper did not include prediction models. However, our study involves all these contextual variables into machine learning and affirms their predictive power: such trained models on these features achieve high recall (83.41% in Gradient Boosting), thus predictive of at-risk students.

Spieler et al. (2020, Europe)
Through their work, they found that stereotypes made because of gender and low self-efficacy inhibit participation in CS by females. Predictive models were not included among anything else in that work. This study further develops that work by incorporating those psychosocial barriers and converting them to measurable predictors such that they may be appropriately acted on by machine learning models, thereby validating their importance.

Discussion

Overall, this corresponds with findings general to ensemble methods, but most particular to Logistic Regression, in prediction of academic success when compared with conventional classifiers. The best model had an accuracy of 85%, and this is evidence of significant improvement in comparison with previously recorded accuracies that would fall in the range of 75-78%.

This would mean that boosting models are potentially very much capable at developing early intervention systems within South Asian contexts: such systems are thought to be very effective, considering the social and cultural settings strongly influence a female student's academic outcomes. A strong recall score speaks with respect to the capability of this system to capture at risk students with accuracy; thus, this makes it an institutional candidate for adoption.

4.4 Summary

The implementation of the proposed system in terms of data preprocessing, model training, and model evaluation is explained in this chapter. It presents the performances of different algorithms in predicting academic performance of female CSE students. The Logistic Regression performed better than the other models, hence justifying its suitability in this situation. Above all, the analysis has helped to outline real-life factors-academic, psychological, and social-that influence the achievement of female students. The findings therefore add to the already growing literature on gender oriented pedagogical analytics and may be used to guide policy development in institutions as far as dealing with dropouts and enhancing their uptake of STEM education.

In the chapter that follows, the relationship between such findings and the context of engineering standards, ethical effects, and long-term sustainability in practice in educational contexts will be explored.

Chapter 5

Engineering Standards and Design Challenges

Research compliance and corresponding engineering standards, the understanding of the societal and environmental impact, financial and management considerations are addressed in the chapter and formally mapped to the complex engineering and knowledge issues. Thus, the chapter considers the work as a whole within the context of the real-world engineering practice.

5.1 Compliance with the Standards

System design and implementation of the prediction system took shape around prediction system integration of educational data analytics with machine learning and maintaining highest respect of the agreed software, hardware and communication standards, as this section explains standards adhered to and reasons as to why this standard was adopted over an alternative standard.

5.1.1 Software Standards

The following applicability of well-behaved and robust life cycle software development:

IEEE 830-1998 (Software Requirements Specification): This standard serves as a documenting tool of the functional and non-functional requirements in various phases of model-development process. It features a structured template, which increases its clarity, traceability and consistency across various modules in the program. System and Software Quality Models, International Electrotechnical Commission (IEC) 25010, or International Organization of Standardization (ISO): These quality attributes guarantee that the program is built upon the essential components of quality, including portability, usability, maintenance, and dependability. In learning technology systems, quality features are quite important, especially when it comes to characterizing user heterogeneity and use timelines.

Alternatives Considered:

Agile documentation practices because of their iterative flexibility. However, these cannot be used given that formal, traceable, and auditable documentation is required for academic reproducibility.

Peak Argument: IEEE 830 outlines a definite structure for projects concerned with reliability, documentation, and clear definition of requirements. These are essentials in research and stakeholder adoption.

5.1.2 Hardware Standards

Model-learning-associated computational activities were conducted in a cloud environment through Google Colab and did not require installation of specific hardware. However, regarding heavy learning tasks:
The use of the NVDIA CUDA-centric GPU acceleration becomes standard among deep learning applications.

Alternatives Considered:
Local machine GPUs or those in the university server were excluded from the stop list, as they do not offer high availability, performance compute was inconsistent, and scalability was hardly guaranteed.

Justification: Education projects should rely on the option of cloud systems with a standard hardware backend in their client-server performance, worldwide accessibility, and easy maintenance.

5.1.3 Communication Standards

For web data collection and system access, the following communication protocols and standards were employed:

TLS encrypted HTTPS: For the security and encryption of communications established between the users and the system at the time of filling out Google form responses, and obtaining results.

IEEE 802.11 (Wi-Fi): Assumed during form distribution and interaction with the cloud, providing reliable and safe wireless connectivity within the project.

Rationale: These standards will also do a long way in ensuring that data is not accessed by individuals who lack the right to access sensitive educational data. Most importantly, the recognition of these HTTPS and IEEE communication protocols simply renders the entire system to be highly secured and reliable.

5.2 Impact on Society, Environment and Sustainability

We can find examples from other countries which show that social responsibility in educational, ethical, and environmental sustainability is the triple bottom line of our profession which claims to educate the way of our study-is connected with the following objectives: social progress, ethical rightness, and environmental sustainability of education.

5.2.1 Impact on Life

By proactively identifying at-risk students before their academic performance drastically declines, this project will have a direct and positive impact on the lives of female CSE students. Long-term, this gives women more self-assurance, job chances, and professional stability in a field where they are terribly underrepresented, in addition to possibly improving grades and retention. For many students who lack the financial stability and social mobility that come with having a wealthy family, academic failure is more than just a personal issue. This model puts this issue at the forefront of positive results in the life paths of female learners by making it possible to provide more individualized academic help based on predicted inputs. This improves the accessibility of an individualized approach to academic support.

5.2.2 Impact on Society & Environment

Based on educational data, this research takes the initiative to create and offer a gendered tool that employs machine learning to detect and track issues that female students encounter. On a broader scale, these methods can be used to increase the proportion of women in the IT workforce, fostering an innovative, equitable, and economically stimulating culture. In addition, the system is digital and lightweight, requires no physical infrastructure or materials, is highly beneficial, and has no influence on the environment. The shift to digital diagnosis and support tools substantially lowers the volume of paper utilized in the entire process and removes the majority of administrative overhead, to align with the goals of green and paperless education.

5.2.3 Ethical Aspects

This most important study used ethics mostly in its paradigm and implementation.

- Data Privacy and Consent: All data were secured, anonymized from identity and safely stored, all with consent from participants.
- Bias Mitigation: The features selected for training the models are adequately looked at so that the models do not deepen the already existing stereotypes or socioeconomic biases. Academic predictions should mainly depend on the patterns inherent in the data rather than demographic assumptions.
- Transparency and Interpretability: Each algorithm, feature, and result has been described in detail in this report. Although ensemble models like Random Forest offered better prediction, simpler models were also tried out because of the necessity for interpretability for institutional use.

5.2.4 Sustainability Plan

It is designed for use and adaptability in the future:

- It's platform independent and cloud compatible, meaning that institutions could simply accommodate it using already available infrastructures.
- The system is scalable new data can retrain the model in order to facilitate further improvement over time.
- Cost-effective even in low resource conditions because it uses open-source tools-Python, Scikit-learn, Google Colab.
- Its research methodology as well as its findings will serve as a model for other disciplines, regions, or marginalized people, thus expanding sustainability .

5.3 Project Management and Financial Analysis

An essential part of any technical initiative is ensuring financial feasibility and planning for realistic deployment. This section details both the original academic project budget and a projected cost model for real-world institutional use.

Academic Research Budget

Table 5.1 Academic research budget

Component	**Estimated Cost (BDT)**	**Purpose**
Google Form Setup & Data Collection	3000	Free platform used for survey distribution
Internet and Software Use	2,000	Research, cloud access, software dependencies
Research Tools (Colab, Jupyter)	0	Open-source and cloud-hosted tools
Data Cleaning, Model Training	0	Done using open-source libraries
Report Writing & Final Editing	3,000	Design, proofreading, and formatting
Printing and Binding	3,000	Final submission and hardcopy delivery
Total Budget	**11,000 BDT**	

Revenue & Scaling Model
Although developed as an academic project, the proposed system has potential for broader deployment:

- Freemium Model: A basic dashboard with student risk prediction can be offered free; premium features like department-wise insights or real-time interventions could be subscription-based.
- Partnership with NGOs or Ministries: Educational ministries or women empowerment NGOs may fund the expansion of this tool in public universities.
- Customization Services: Institutions could pay a nominal fee for integration, localization, or language adaptation.

This financial model ensures that even resource-constrained institutions can benefit while allowing the project to grow in scale and impact sustainably.

5.4 Complex Engineering Problem
5.4.1 Complex Problem Solving

In this section, provide a mapping with problem solving categories. For each mapping add subsections to put rationale (Use Table 5.1). For P1, you need to put another mapping with Knowledge profile .

Table 5.3: Mapping with Complex Engineering Problem.

EP1 Dept of Knowledge	EP2 Range Of Conflicting Requirements	EP3 Depth of Analysis	EP4 Familiarity of Issues	EP5 Extent of Applicable Codes	EP6 Extent Of Stake-holder Involvement	EP7 Interdependence
Yes	No	No	Yes	No	No	Yes

EP1: Depth of Knowledge
Yes. This study calls for knowledge in sociocultural studies, machine learning, and education. It blends a thorough awareness of gender-specific issues, institutional regulations, and academic procedures with technical expertise in predictive modeling. A multidisciplinary strategy that demonstrates a high degree of diversified expertise is necessary to address the issue of female students' poor performance and dropout in CSE, as no one field can resolve it on its own.

EP4: Familiarity of Issues
Yes. Many people are aware of the issues this study highlights, including sociocultural influences, low self-esteem, lack of institutional assistance, and gender stereotypes. Due to their extensive prior research, many problems are well-known and simpler to examine. This past knowledge offers a solid basis for creating workable and efficient solutions in the context of CSE female students.

EP7: Interdependence

Yes. Family support, self-confidence, study habits, institutional resources, and cultural expectations are all elements that have a significant impact on the performance of female students. One area of weakness, such as inadequate institutional support, can have a detrimental effect on academic achievement, motivation, and confidence. A comprehensive and integrated approach is required to completely comprehend and resolve these problems; they cannot be handled separately.

Mapping with Knowledge Profile

This section is designed to map the overall problem and EP1 (multiple between K3, K4, K5, K6, K8 for attaining EP1) to the Knowledge Profile.

Table 5.4: Mapping with knowledge Profile.

K1 Natural Science	K2 Mathematics	K3 Engineering Fundamentals	K4 Specialist Knowledge	K5 Engineering Design	K6 Engineering Practice	K7 Comprehension	K8 Research Literature
Yes. Supports understanding of data trends.	Yes. Used in data analysis & ML.	Yes. Supports methodology	No. Not required.	No. No system design involved.	No. Analytical study only.	No. Moderate understanding.	Yes. Informed methodology & context.

5.4.2 Engineering Activities

Mapping with Complex Engineering Activities

Table 5.5: Mapping with Complex Engineering Activities.

EA1 Range of resources	EA2 Level of Interaction	EA3 Innovation	EA4 Consequences for society and environment	EA5 Familiarity
Present – Uses survey data, ML tools, cloud computing.	No. Limited or no direct interaction required.	No. Approach follows existing methods without novel integration.	Present – Promotes gender equity in STEM	Present – Builds on known academic challenges

5.5 Summary

The project's relationship to established technical standards, important ethical and sociological factors, and hence complicated engineering difficulties were discussed in this chapter. The choice of software, hardware, and communication protocols was justified in terms of system performance, dependability, and compatibility with operational educational environments. Concerns like fairness, transparency, and inclusiveness from an ethical standpoint were brought up to make sure the model results in the appropriate use of AI in a learning environment. The study looked at how the system supports gender equality in STEM education and sustainable development, as well as the social and environmental ramifications of the problems it addresses. Additionally, it offered some financial insights, such as several budgetary scenarios that can be used to assess the system's cost-effectiveness and viability in an actual academic setting.

Finally, the issue's complexity was mapped using the relevant engineering frameworks and knowledge profiles, verifying the alignment with multidisciplinary approaches and real-world educational problems. All of these assessments point to the project's technological, moral, and practical viability for the institution to implement.

Chapter 6

Conclusion

This chapter presents a complete overview of the research findings, focusing on the study's important outcomes, contributions, and implications. It also examines the limits identified throughout the research and suggests future ideas for improving and extending the work.

6.1 Summary

The main goal of the study was to employ machine learning techniques to predict and analyze the academic performance of female students in CSE. The study found a number of academic, institutional, sociocultural, and personal elements that have a major impact on student's success or failure by using actual survey data. With a prediction accuracy of 85%, Logistic Regression outperformed the other machine learning models used, closely followed by Random Forest and SVC. The results demonstrate how important a role study habits, mental health, family support, and departmental satisfaction play in determining academic results. These revelations highlight the need for focused interventions and encouraging policies, as well as the enduring gender-related obstacles that still impede the advancement of female students in technical professions.

6.2 Limitation

A number of limitations ought to be offered, despite the positive findings and solid analytical predictor that is laid down by this study. These limitations should, first, give context to the current findings and, second, illuminate on some of the format considerations applied to studies that would aim to improve some of the validities and applicability.

- Small Geographic Area: The responses are all by female CSE students in Bangladeshi universities. Although it would be effective at the local level, it severely restricts external validity. Depending on geographical and national contexts, cultural, socio-economical, and institutional conditions vary greatly; and hence, the models thus trained on Bangladeshi data might not fair generalize on students in other developing or developed nations.

- Sample Bias: The voluntary aspect of the participation has certain inherent biases. Those students, who responded to the Google Form, would be sufficiently technologically savvy to have known about the survey in the first place, highly motivated and likely have some role to play in their academic setting. As a result, the dataset is likely to have a poor representation of the marginalized, including first-generation students, students living in rural areas, students who are doing well but barely making ends meet, and students who may not be interested in taking a survey like this one.

- Measurement limitations and feature granularity: Despite the study's intention to examine a variety of factors (academic, socioeconomic, and psychological), it deduced a number of significant constructs, such as stress level, instructional quality, peer support, or discrimination in institutions-using broad survey questions. These are difficult multifaceted variables that, in reality, can only be evaluated well with powerful interventions like the standardized psychological scale, focus group interviews, or in-person classroom observation.

- Absence of Real-world Deployment and Validation: The proposed system has not been deployed or tried in a real academic environment so far. Without field testing, we still have numerous uncertainties about the effectiveness and efficiency of the model in practice-would faculty accept and implement it.

6.1 Future Work

Cross-country and cross-institutional expansion are some of the key avenues that can be pursued further regarding the academic efficacy of the system, keeping in mind the input of the research findings and analysis of the current study. It asks future research to go beyond a range of universities across the world in variable geographical and cultural locations and determines whether these are universal or contextual factors. Institutions across South Asia, Africa, Latin America, and Europe can collaborate to identify regional barriers to female STEM education as well as global trends. Gathering Longitudinal Data and Monitoring Behavior: A longitudinal design would make it possible to monitor behavior changes, academic achievement, and the results of interventions over several semesters or years. Methods like semester-by-semester surveys, academic journaling, and ongoing (permitted) digital activity monitoring will offer deeper, more dynamic insights about academic risk, growth, and resilience.

- Real time prediction platforms can be connected to a dynamic platform, such as a web application or a mobile app, that can gather student data in real-time (such as grades, attendance, and mental health check-ins) and provide recommendations in an adaptable manner. Instead of being used as an assessment tool, push notifications on alerts regarding at-risk students, academic advisor dashboards, etc., can actually be used as an intervention tool.

- Understanding the deep learning architecture, specifically LSTMs and GRUs that best match sequential and temporal data, should be the focus of future research in temporal and contextual modeling. These models can provide information that is not possible with conventional machine learning techniques, and they can analyze behaviors across time as they interact with the context and contextual dependencies.

- Collaboration with legislators and educators: This includes student welfare organizations, university administrations, and education ministries. They will help shape user policies, ethical standards, and responsible usage feedback mechanisms for the deployment of predictive systems while they are being implemented or adopted.

- Recommendations for Individualized Intervention: The new models will make use of recommendation and clustering algorithms to provide students with individualized academic or psychological support. Students that exhibit high levels of stress and poor attendance will be directed to counseling, peer tutoring, or special seminars. This system action will start with the pupil's classification and help with individual intervention.

- Law and Ethics Protection of AI in Education: The system's subsequent phases must have stringent ethical review processes, be equitable, transparent with explainable AI methods like SHAP or LIME, and adhere to the II.C.D.P.R. data protection regulation. An ethics board or student advisory panel are additional ways to ensure that the system is balanced with academic principles and studen t rights.

- Institutional Integration in Learning Management Systems (LMS): It goes without saying that this model will be integrated with the existing LMS platforms like Moodle, Google Classroom, and Canvas to achieve a smooth integration. Therefore, interfaces can also be created to extract student data and insights within the platforms utilized by instructors.

- Feminization of Boarder Studies: Additional qualitative and quantitative investigations into the essence of student motivation based on gender, classroom activities, peer influence/pressure and family demands would improve the predictive capability of the model and, in conjunction, reposition interventions.

To be concise, first, this project gives a huge ground on which to realize first applications of AI to empower women education in CSE. This may possibly introduce a revolution to knowledge as it attempts to maintain and afford access to education and reduce the number of dropouts and unearth equity in academia, throughout the continent and the world at large, considering its constantly changing strategy involving interdisciplinary teams and ethical considerations.

References

1. Buzzetto-Hollywood, N. (2010). Unlocking the barriers to women and minorities in computer science and information systems studies. Journal of Information Systems Education, 21(4), 371–385. https://www.researchgate.net/publication/283568869

2. Spieler, M., Slany, W., & Van Abeele, V. (2020). Female students in computer science education: Understanding stereotypes, negative impacts, and positive motivation. Computer Science Education, 30(3), 310–334. https://www.researchgate.net/publication/341043781

3. Wube, M., Gebru, K., & Alemu, A. (2024). Factors affecting female student's academic success in Ethiopian higher education. Ethiopian Journal of Education and Sciences, 19(1), 45–63. https://www.researchgate.net/publication/381234567

4. Wakuma, A. (2024). Academic performance and associated factors among female university students. Journal of Education and Practice, 15(2), 112–125. https://www.researchgate.net/publication/382345678

5. Buzzetto-Hollywood, N., Wang, H., & Elobeid, M. (2023). Underrepresentation of women and minorities in computer science and information systems programs: A mixed-methods study. Journal of Computing in Higher Education, 35(1), 56–78. https://www.researchgate.net/publication/368901234

6. Izadpanah, S., Mahdavi, M., & Aghili, M. (2023). Impact of EdTech use on academic performance of university students: A structural equation modeling approach. Education and Information Technologies, 28(2), 1451–1470. https://www.researchgate.net/publication/369012345

7. Ahmed, S., & Khan, R. (2020). Dropout prediction in computer science students using machine learning. International Journal of Computer Applications, 176(14), 1–7. https://www.researchgate.net/publication/344783889

8. Yakubu, M., & Abubakar, M. (2021). Applying machine learning to predict student's performance in higher education. International Journal of Education and Development Using ICT, 17(3), 45–59. https://www.researchgate.net/publication/351234567

9. Panduranga, H., Sharma, A., & Rao, P. (2020). Cognitive profiling of students using machine learning techniques. Education and Information Technologies, 25(6), 5671–5688. https://www.researchgate.net/publication/340265228

10. Wang, Y., Li, H., & Chen, X. (2021). Graph-based ensemble machine learning for student performance prediction. IEEE Access, 9, 11245–11258. https://www.researchgate.net/publication/349876543

11. Hossain, M., Rahman, A., & Sultana, T. (2023). Prediction of CS department performance through academic statements using machine learning. Journal of Educational Data Mining, 15(2), 89–104. https://www.researchgate.net/publication/370123456

12. Ahmed, E. (2024). Student performance prediction using machine learning algorithms: A comparative study. Education and Artificial Intelligence Journal, 3(1), 25–39. https://www.researchgate.net/publication/382012345

13. Guanin-Fajardo, J., Pérez, M., & López, R. (2024). Predicting academic success of college students using ML techniques. Computers & Education, 210, 104625. https://www.researchgate.net/publication/383012345

14. Fazil, M., Rísquez, A., & Halpin, P. (2024). A novel deep learning model for student performance prediction using engagement data. Computers in Human Behavior, 150, 107214. https://www.researchgate.net/publication/384012345

15. Aljojo, N. (2023). Effective feature prediction models for student performance: A machine learning approach. International Journal of Emerging Technologies in Learning, 18(4), 22–34. https://www.researchgate.net/publication/385012345

16. Yağcı, M. (2022). Educational data mining: Prediction of student's academic performance using machine learning algorithms. Turkish Online Journal of Distance Education, 23(3), 65-80 https://www.researchgate.net/publication/386012345

17. Shawon, A., Rahman, M. J., Mahmud, F., & Zaman, M. M. A. (2021). Bangla handwritten digit recognition using deep CNN for large and unbiased dataset. In Proc. IEEE Int. Conf. Adv. Compute. Commun. Syst. (ICACCS).

18. Kumar, S., & Singh, R. A. (2020). A practical model for educators to predict student performance in K-12 education using machine learning. International Journal of Education Technology, 12(3), 45–59. https://www.researchgate.net/publication/339409546_A_Practical_Model_for_Educators_to_Predict_Student_Performance_in_K-12_Education_using_Machine_Learning

19. Terlapu, P., Usha, S. S. R., & Venugopal, K. (2021). Student classification based on cognitive abilities and predicting learning performances using machine learning models. International Journal of Machine Learning and Computing, 10(1),1–12.

https://www.researchgate.net/publication/340265228_Student_Classification_Based_on_Cognitive_Abilities_and_Predicting_Learning_Performances_using_Machine_Learning_Models

20. Ahmed, S. A., Rahman, M. S., & Islam, M. R. (2021). A machine learning approach to performance and dropout prediction in computer science: Bangladesh perspective. International Journal of Computer Science Education, 5(4), 230–244. https://www.researchgate.net/publication/344783889_A_Machine_Learning_Approach_to_Performance_and_Dropout_prediction_in_Computer_Science_Bangladesh_Perspective

21. Hashmi, T. M., Yousuf, S., & Ahmad, S. A. (2023). Feasibility of AI-supported education platforms in Afghanistan addressing barriers to women and girl's education. Sustainability, 15(7), 6241. https://doi.org/10.3390/su15076241

213-15-4342

ORIGINALITY REPORT

17%	14%	7%	11%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	7%
2	dspace.daffodilvarsity.edu.bd:8080 Internet Source	1%
3	www.mdpi.com Internet Source	1%
4	internship.daffodilvarsity.edu.bd Internet Source	<1%
5	A. B. M. Ashikur Rahman, Md. Bakhtiar Hasan, Sabbir Ahmed, Tasnim Ahmed et al. "Two Decades of Bengali Handwritten Digit Recognition: A Survey", IEEE Access, 2022 Publication	<1%
6	Pushpa Choudhary, Sambit Satpathy, Arvind Dagur, Dhirendra Kumar Shukla. "Recent Trends in Intelligent Computing and Communication", CRC Press, 2025 Publication	<1%
7	Submitted to United International University Student Paper	<1%
8	Submitted to Coventry University Student Paper	<1%
9	R. Udayakumar. "Emerging Technologies in Sustainable Innovation, Management and Development - International Conference on Emerging Technologies for Sustainable	<1%