CAREER ENFORCEMENT PREDICTION USING MACHINE LEARNING

BY

Indrani Rakshit
ID: 241-25-025

FINAL YEAR DESIGN PROJECT REPORT

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

Supervised By

Mr.Md. Sazzadur Ahamed
Associate Professor & Program Coordinator
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised by

Mr. Abdus Sattar
Associate Professor & Director
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

July 2025

APPROVAL

This Project titled "**Career Enforcement Prediction Using Machine Learning**", submitted by **Indrani Rakshit** 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 **24.05.2025.**

BOARD OF EXAMINERS

Dr. Arif Mahmud — **Chairman**
Associate Professor and Associate Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Fizar Ahmed — **Internal Examiner**
Associate Professor
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

Nazibur Rahman — **External Examiner**
Technical Lead, Database Administrator
Wipro, Telenor-Grameen Phone Account
Dhaka, Bangladesh

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Mr. Md. Sazzadur Ahamed, Associate Professor and Program Coordinator, 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:

Mr. Md. Sazzadur Ahamed
Associate Professor and Program Coordinator
Department of CSE
Daffodil International University

Co-Supervised by:

Mr. Abdus Sattar
Associate Professor & Director
Department of CSE
Daffodil International University

Submitted by:

Indrani Rakshit
ID: -241-25-025
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 **Mr. Md. Sazzadur Ahamed, Associate Professor & Program Coordinator**, 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 the **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

The transformation of our world demands students and professionals to make essential choices about their future careers. The Career Enforcement Prediction using Machine Learning project evaluates machine learning methods for forecasting career alignment through motivational and demographic and academic factors. The analysis of the custom survey-based data set used exploratory data analysis methods to discover patterns between the variables of age, gender, preference choices and personal drive. Multiple classification models including Random Forest, Support Vector Machine, XGBoost, Gradient Boosting and K-Nearest Neighbors were integrated after the data preprocessing step and the categorical variable conversion process was completed. The evaluation of these models relied on measuring accuracy together with confusion matrix analysis and classification report evaluation to assess performance levels. The proposed model which demonstrated the best performance was saved using pickle for deployment purposes. Tools were implemented to develop an interactive web-based application through Streamlit enabling users to get real-time career alignment predictions by entering their data through an API interface. The project applies machine learning methods to practical career counseling solutions which aid individuals with their career selection choices.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Overview

The selection of a suitable career represents an essential life choice for people who move from school to professional employment. Decisions regarding careers develop through a mix of personal interests along with academic results and family requirements besides cultural influences and motivational aspects. Numerous people lack both appropriate guidance and adequate self-awareness abilities to perform complete evaluation of these decision factors. Many students make career choices which do not match their skills or goals which produces feelings of discouragement and subpar results and possibly forces them to change employment sectors in the future.

Pattern recognition along with predictive modeling have advanced through machine learning technology as a vital digital industrial tool during the current technological period. Through data learning systems acquire capability to detect patterns in their information which allows them to predict trends and generate smart decisions. Machine learning algorithms enable the development of models that evaluate multiple personality attributes alongside motivational factors to specify suitable professional options based on individual characteristics.

A machine learning-based prediction system uses career-related data for generating individual career enforcement forecasts. Streamlit supports the deployment of an interactive interface through which this trained model processes survey-based data. Through the application users can enter their details which produces instant real-time predictions to function as an automated career support service.

1.2 Background and Present State

The current practice of career counseling depends mainly on subjective evaluation and standardized tests in combination with exclusive career advisor sessions. These established evaluation techniques deliver useful results but remain both lengthy to administer along with expensive to execute while also excluding substantial student demographics who live in developing nations. The methods fall short in adjusting their assessment to market changes and personal aspirations and individual strengths.

The data-driven method of machine learning presents students with an approach that scales efficiently for career counseling. Machine learning models process considerable datasets about individual profiles and career end results to discover sophisticated variable patterns beneath what conventional analysis methods can show. The predictive power of these models derives from previous data patterns that leads them to provide personalized recommendations for new inputs.

Job portals and e-learning websites now incorporate new career recommendation systems which appeared during recent years. These systems mostly do resume matching but fail to integrate motivational understanding as well as psychological insights. The project fills this knowledge gap through its predictive model which combines motivational elements and demographic as well as academic data.

1.3 Problem Statement

Even though many tools exist alongside various resources students alongside those who are beginning their careers experience difficulty in finding their ideal career path. The primary challenge exists in the absence of accessible computer systems which create data-driven personalized recommendations from complete individual profiles. There exists a problem with current systems that fails to connect personal motivation with academic background while properly considering individual demographic profiles effectively.

The main problem this project targets through machine learning is creating accurate career prediction models while integrating knowledge from personal and academic details together with motivational factors and designing these assessments to display through an easy-to-use interface.

The system development targets an operational mechanism which delivers accurate predictions while providing meaningful career choice information to users.

1.4 Objectives

The main ambition of this project is to create and manage a machine learning platform which forecasts career enforcement levels primarily for individuals. The specific objectives are:

- The study requires the acquisition of important information that covers individual traits along with academic standing and motivation levels that influence career decision-making processes.
- The current study uses exploratory data analysis (EDA) to reveal hidden patterns and important insights in the available data.
- The dataset requires pre-processing that includes measurements for cleaning up data along with handling missing values and transforming categorical variables to numbers.
- I will conduct model training and evaluation for Random Forest and Support Vector Machines (SVM) and Gradient Boosting as well as K-Nearest Neighbors (KNN) and XGBoost.
- The models' performance will be evaluated by assessing accuracy, generating confusion matrices and drawing classification reports.
- The selection process of the performing model will be followed by serialization through the pickle method for future utilization.

- The Streamlit API development requires creation of a user-friendly interface that enables data entry and instant career prediction output to users.
- The system requires clear presentation and understanding of data to serve as a career planning resource for people making decisions.

1.5 Scope and Limitations

The project follows a complete path which begins with data acquisition until the point of system deployment. The goal of this work is to develop a classification system which forecasts career enforcement through survey evaluations. The system uses Streamlit to develop its Graphical User Interface (GUI) which provides an accessible approach for end-users to work with the model. The system exists for educational and career guidance functions while proving how AI can enhance individualized career planning services.

Limitations:

- The project relies on a small dataset with narrow focus that potentially limits the transferability of its findings.
- Users who complete surveys might introduce errors through their personal interpretation or by being dishonest when responding.
- At present the model determines career category alignment yet it fails to deliver sophisticated career guidance or different career paths
- Career choices made in real life depend on various flexible factors such as economic patterns and personal values alongside external options because these elements are absent from the gathered data.
- The program lacks built-in natural language processing technology for converting freeform user input.

1.6 Report Organization

The report has a structure with six main chapters.

Chapter 1: The first chapter delivers an overview of the project together with background information and explains the addressed problem and project goals and report structure.

Chapter 2: The literature review includes research about machine learning applications for career prediction together with existing systems analysis and previously identified research gaps.

Chapter 3: The report details the systematic methods applied to gather data and pre-process it and select appropriate models and execute training while conducting evaluations.

Chapter 4: The section displays implementation information along with data visual representations and performance and model comparison results for each applied algorithm.

Chapter 5: The article documents Streamlit interface creation and trained model incorporation into the system in addition to system user interaction details.

Chapter 6: The final section of this work summarizes project results while reflecting on its impact through proposed developments for future improvements.

1.7 Summary

The chapter explains how machine learning supports career enforcement prediction while describing its importance in modern educational and professional systems. The project expressed its reasons for developing along with its objectives and adopted technological methods for implementation. The current chapter outlines the report's organization structure which prepares readers for more advanced technical explanations in subsequent sections.

CHAPTER 2

LITERATURE REVIEW

2.1 Overview

Transition into professional life requires careful planning since educational periods draw near to their end. The rapid expansion of knowledge fields alongside technology disruptions and changing employment sectors requires people to handle complex and challenging career decisions. For many years people pursued career advice from career counselors and standardized assessments together with self-directed career exploration for pathway choices. The current strategies fail to deliver exact and individual-specific continual guidance to users.

Machine learning (ML) presents an appealing prospect for transforming the career advice distribution process. The ability of ML algorithms to derive knowledge from past data helps identify complex associations between variables which determine career success. The development of predictive systems using this approach allows students to make educated choices which correspond to their motivational factors and skill profiles together with academic achievements.

A complete examination of research in the field of ML-based career prediction and recommendation systems occurs in this chapter. The chapter starts with an evaluation of worldwide and Bangladeshi career-related research which gets followed by stability analysis and significant gaps discovery before presenting the investigation's key findings.

2.2 Related Works

Academic attention towards data-driven career prediction techniques has intensified throughout the last few years. Research studies have established models through classification methods and clustering approaches and hybrid approaches to support student career selection.

Pandey and Taruna (2016) developed a career prediction model which used decision tree and Naive Bayes algorithms. Engineers' aspirants could select the correct specialization between Computer Science and Mechanical engineering using academic performance metrics and personal choices through their system. A set of entrance exam scores combined with internal assessment results made up the analyzed data [1] .

The research by Kaur and Singh (2020) developed a system to match careers with personal traits from interest tests and academic backgrounds. Their research used Random Forest and SVM classifiers which maintained accuracy levels higher than 85%. The system aimed to unite behavioral analysis with numerical assessment data [2] .

Zhang and colleagues (2018) conducted deep learning analysis of Chinese online learning platform behavioral data to predict career goals of users. RNNs were used in their research and revealed that patterns observed across time in learning activities consistently forecasted career paths [3] .

Kumar and Selvakumar (2021) developed a peer evaluation model that unitized collaborative filtering along with decision trees to provide recommendations directed at user careers. The study demonstrated that united academic statistics and social information produces superior results [4] .

Technology adoption in education needs has become prevalent throughout Bangladesh's higher education sector particularly from public universities together with technical institutes.

The researchers at University of Dhaka designed a decision support system that helps higher secondary students pick between Science or Humanities or Commerce programs (Rahman 2019). Logistic regression combined with SSC results and socioeconomic data and personal choice preferences was used as the model base. The basic model demonstrated that decision support platforms play a critical role in Bangladeshi decision systems [5] .

A rule-based system aimed at student career path selection between technical and non-technical fields exists according to Islam (2020) from North South University. The model achieved high interpretability which made it suitable for individuals who did not understand AI tools especially those living in rural areas [6] .

Engineers from BUET developed a predictive system during their capstone project in 2021 by incorporating KNN with decision rules. The model received training data from local institutions containing students' GPA alongside motivation levels. This project demonstrated how artificial intelligence can work with real educational datasets in Bangladesh although it remained in prototype stage [7] .

2.3 Comparison between existing works

An essential table presents an evaluation between various significant research studies and projects through their deployment status and their focused areas and methodological approaches.

Table 2.3: Comparison Table

Study / Project	**Approach Used**	**Country**	**Feature Consideration**	**Models Used**	**Strengths**	**Development Status**
Pandey & Taruna (2016)	Data mining	India	Academic scores, stream preference	Decision Tree, Naive Bayes	Simple, interpretable, fast	Research Only

Kaur & Singh (2020)	Predictive modeling	India	Personality, interests, performance	Random Forest, SVM	Combines psychometrics and scores	Prototype
Zhang (2018)	Deep learning	China	Online learning behavior, clickstream	RNN, LSTM	Captures time-dependent behavior	Platform-integrated
Rahman (2019), DU	Logistic regression	Bangladesh	SSC scores, personal interests, demographic factors	RNN, LSTM	Locally contextualized	Research model only
Islam (2020), NSU	Fuzzy rule-based system	Bangladesh	Academic and technical interest levels	Fuzzy Logic	Highly interpretable for rural populations	Experimental
BUET Capstone Project (2021)	Hybrid ML model	Bangladesh	GPA, motivational index, parental influence	KNN + Rule-Based	Personalized, localized	Prototype
This Project (2025)	Supervised ML + Web UI	Bangladesh	Motivation, demographics, academic data	RF, SVM, XGBoost, KNN, GBM	Full-stack ML + interactive Streamlit UI	Fully functional interface

The distinctive aspect of this work delivers an entire operational sequence from preprocessing to model assessment to deployment through an interactive real-time interface. The existing works mainly focused on modeling yet our project combines accurate predictions with practical deployment during evaluation.

2.4 Open Issues

Nevertheless, the progress made in using ML for career guidance several hurdles remain unresolvable. Several problems prevent the practical implementation and large-scale deployment of such systems:

- **Data Limitations:** The datasets used in existing research include mostly domain-bound and minimal in size. Such models demonstrate restricted general applicability because they lack elaborate datasets which include multiple background user types.

- **Subjectivity in Input Data:** Each individual holds unique personal career interests based on their individual subjectivity. The way people report their motivation together with their interest levels and goals depends on their current state of mind while also affected by their surroundings and their comprehension of the questions which causes variability in the data.
- **Cultural and Contextual Gaps:** People's career decisions result from their interactions with cultural standards together with social expectation factors and local market employment activity. A diverse society like Bangladesh requires cultural localization in its career models because these important features are absent from many models today.
- **Ethical Considerations:** Users could mistake the output from such systems as though they were certain results. Automatic system suggestions can negatively impact people when users misunderstand them as fixed predictions because appropriate warnings and information is absent.
- **Lack of Human-AI Collaboration:** The majority of these systems operate independently without human counselor integration. A garnered method combining artificial intelligence with human counselors shows promise as the best possible solution yet remains under development.
- **Real-world Usability and Access:** Very few projects succeed in creating user-friendly applications from their technical models that can be accessed by non-technical users. The implementation of mobile accessibility together with multi-language support remains rare among systems because both features remain essential for developing regions to adopt these tools.

The existing potential limitations offer space for new field advancements which the current project uses as its basis.

2.5 Summary

The article thoroughly examined research on machine learning systems designed for career recommendation and enforcement prediction. Researchers studied global research while reviewing specific studies performed in Bangladesh that highlighted both advantages and challenges related to local settings. The comparison showed that systems deliver high accuracy results although they lack attention to accessible and deployable features. Wider implementation becomes constrained by three key issues: ethical matters combined with data constraints and the requirement for contextual understanding.

The project targets two essential gaps through the development of an accessible user-centered platform which combines accurate career enforcement prediction functions with meaningfully interactive solutions for users. The subsequent chapter examines the development methodology that involved data acquisition, data preparation and modeling techniques and assessment procedures.

CHAPTER 3

METHODOLOGY

3.1 Overview

Machine learning techniques will predict appropriate career paths through processing demographic information combined with academic data and motivational indicators for users. The development of the system follows the detailed steps which are explained in this chapter. The system development process incorporates strategies for data gathering and processing and.ml model training with assessment and software and hardware needs and architecture design and project management and specifications. Through the interface developers created using Streamlit users could obtain real-time career prediction results. The methodology adopts modular design alongside interpretability features for development that suits Bangladeshi user

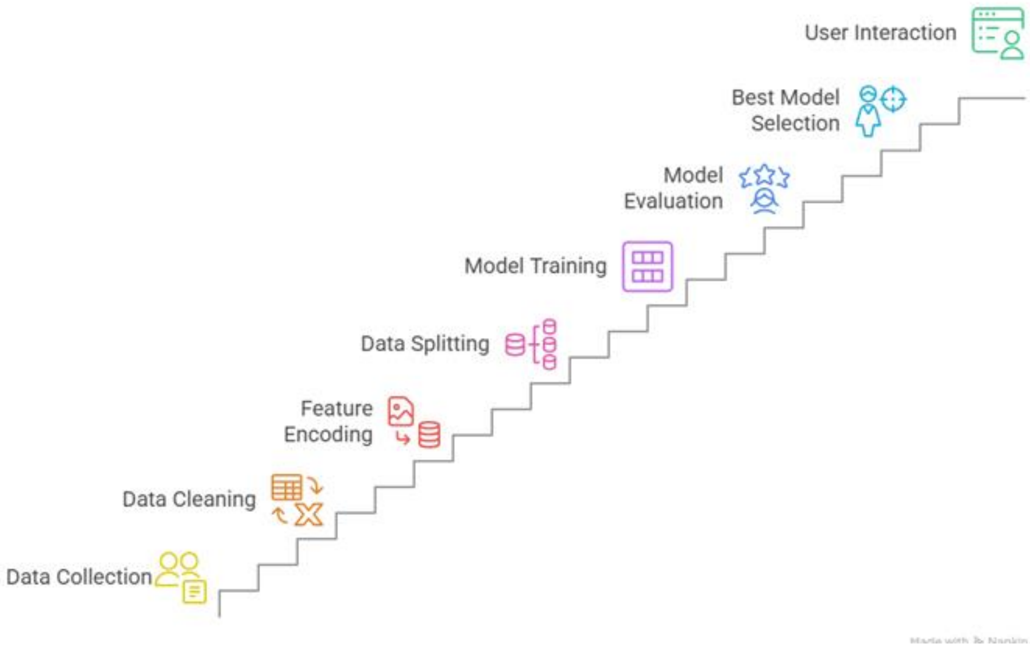


Figure 3.1: Methodology Diagram

3.2 Proposed Methodology/ System Design

This machine learning-based system functions as a complete solution which utilizes advanced algorithms to direct users especially students and young professionals toward their suitable career choices by assessing their individual interests as well as educational achievements and motivational traits. Supervised machine learning combines with a web interface to provide live career direction to users including students and new professionals. The system design consists of modular and interpretable structures and scalable features which allow simple integration within both personal and institutional career guidance programs.

The system follows the following main methodological phases:

3.2.1 Data Collection and Exploration

Researchers designed an organized questionnaire through Google Forms to acquire full career data from Bangladesh-based participants during phase one.
The study focused on collecting data from secondary-school and university students together with graduates who passed their education recently.

Form Fields Included:

- Name
- Age
- Class
- Gender
- Does anybody have motivated you to choose your study group/department where are you studying now?
- Have you discussed your future career with your parents and guardians?
- Do you believe your school provides adequate guidance for career planning?
- Do you participate in extracurricular activities at school?
- Can you handle time management effectively for studies and hobbies?
- Do you believe that your current education is sufficient for your future career?
- Which of the following best describes your preferred type of work?
- At the time of working on a project, which approach do you typically take?
- Which of these skills do you consider your strongest?
- What is your biggest motivation in choosing career?
- Which type of work environment do you thrive in?
- What kind of tasks do you enjoy the most?
- Which skill are you most eager to develop further?
- How important is it or your career to align with your personal values (e.g., sustainability, helping others)?
- Would you be open to relocating or travelling for work?
- Which industries interest you most? (Choose up to three)

All collected data enabled transfer to CSV format followed by conversion to Pandas Data Frame through which researchers conducted inspection and data cleaning practices.

3.2.2 Exploratory Data Analysis

Table 3.2.2: Table for count of attribute “Does anybody has motivated you to choose your study group/department where are you studying now?”

Ans	Count
No	153
Yes	62

Visualization:

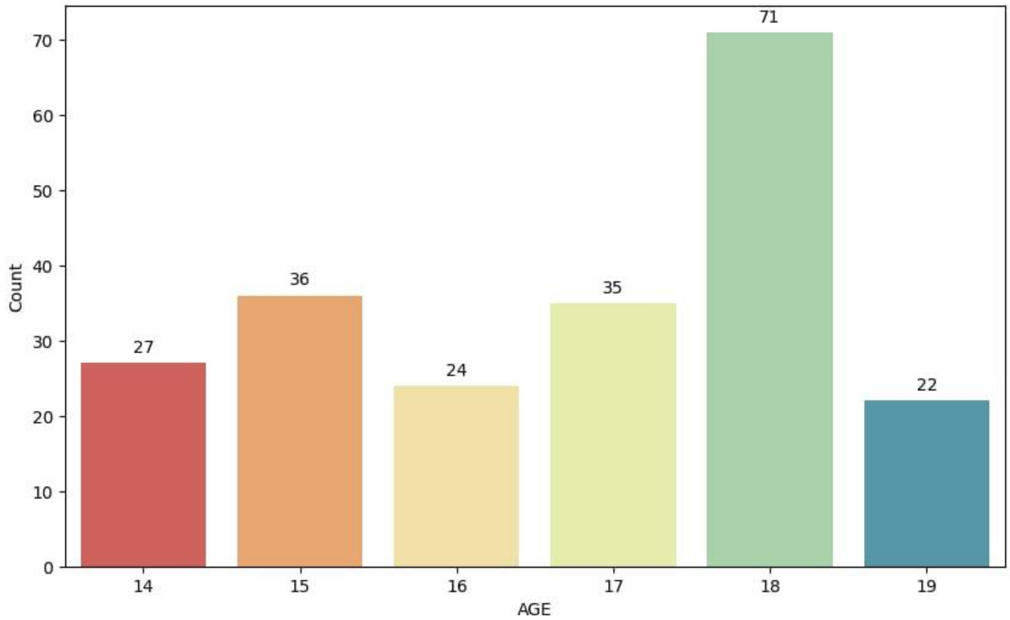


Figure 3.2.2.1: Visualization of Age × Count

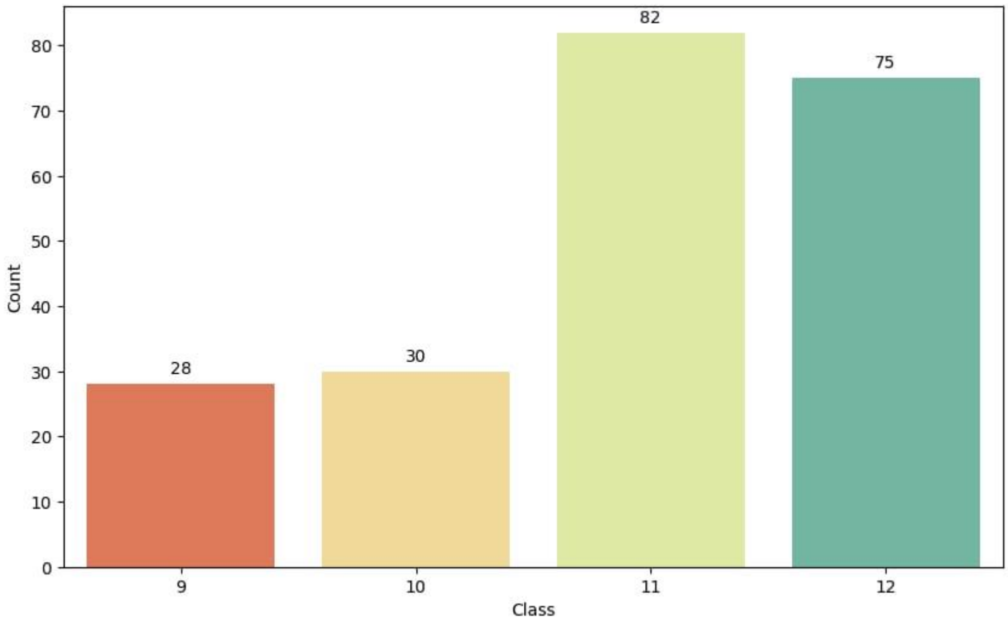


Figure 3.2.2.2: Visualization of Class × Count

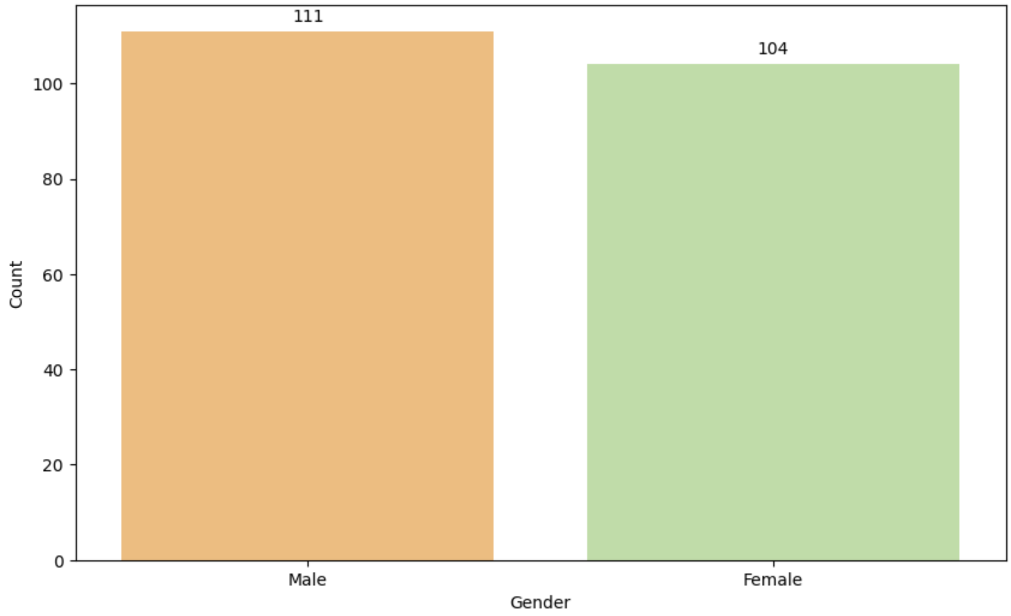


Figure 3.2.2.3: Visualization of Gender × Count

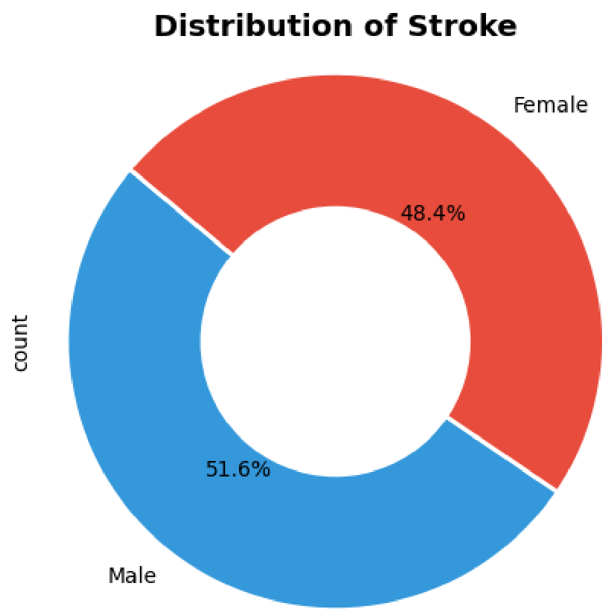


Figure 3.2.2.4: Distribution of Stroke × Count

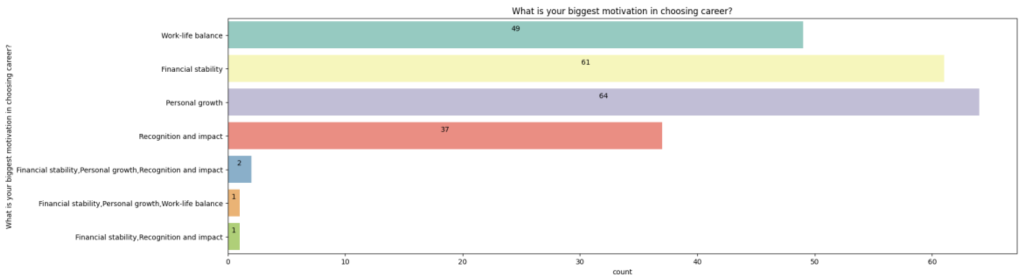


Figure 3.2.2.5: Distribution of 'What is your biggest motivation in choosing career? × 'What is your biggest motivation in choosing career?

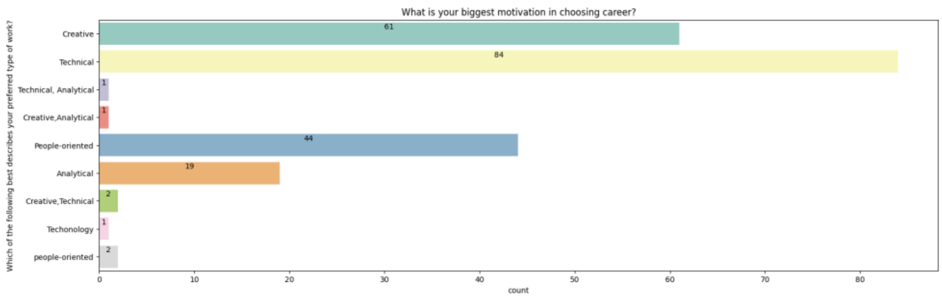


Figure 3.2.2.6: Distribution of 'What is your biggest motivation in choosing career? × Which of the following best describes your preferred type of work?

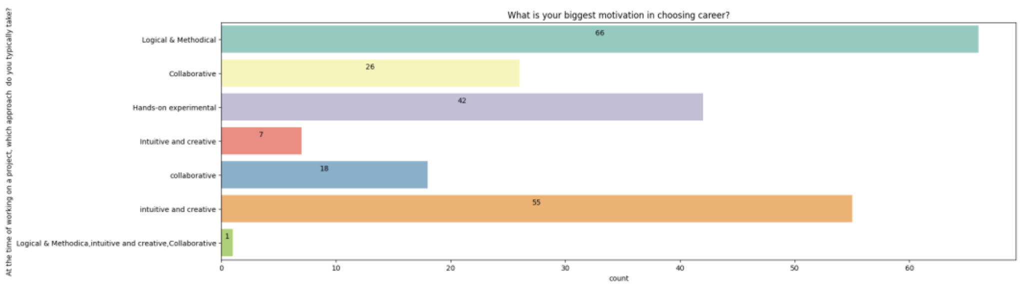


Figure 3.2.2.7: Distribution of 'What is your biggest motivation in choosing career? × At the time of working on a project, which approach do you typically take?

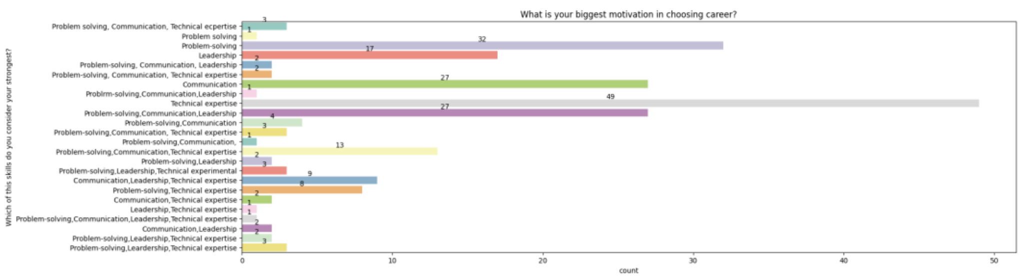


Figure 3.2.2.8: Distribution of 'What is your biggest motivation in choosing career? × Which of these skills do you consider your strongest?

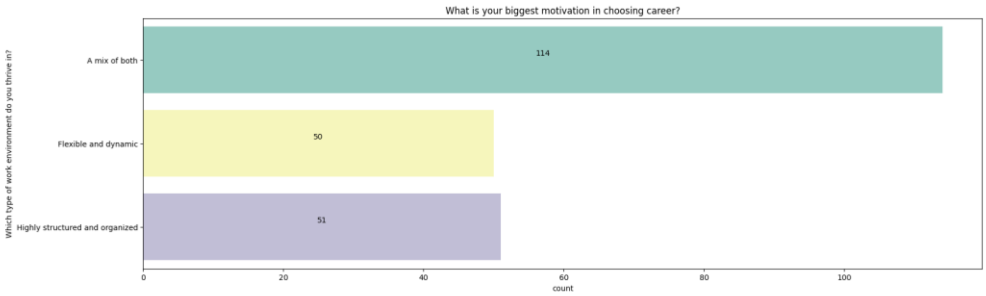


Figure 3.2.2.9: Distribution of 'What is your biggest motivation in choosing career? × Which type of work environment do you thrive in?

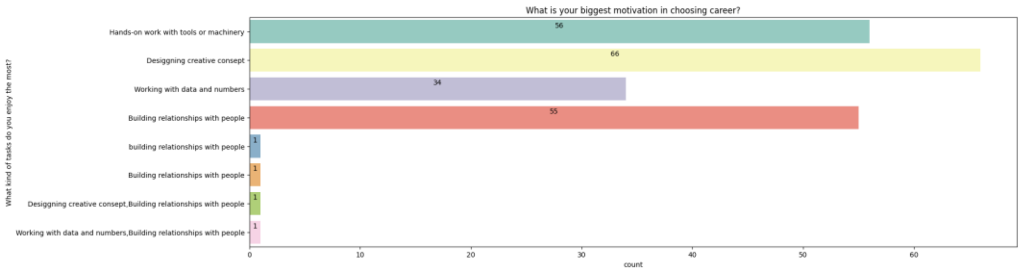


Figure 3.2.2.10: Distribution of 'What is your biggest motivation in choosing career? × What kind of tasks do you enjoy the most?

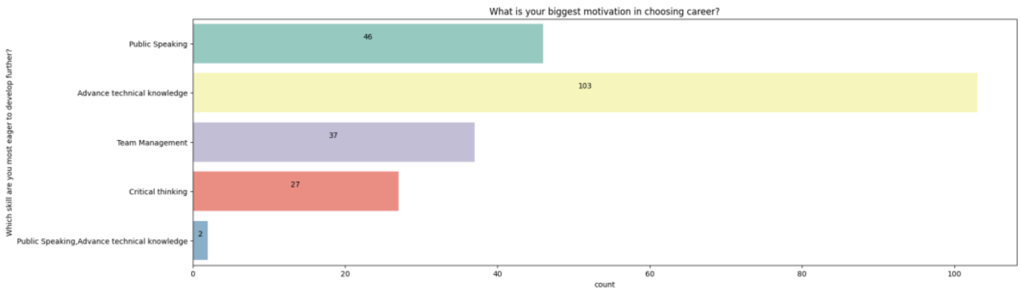


Figure 3.2.2.11: Distribution of 'What is your biggest motivation in choosing career? × Which skill are you most eager to develop further?

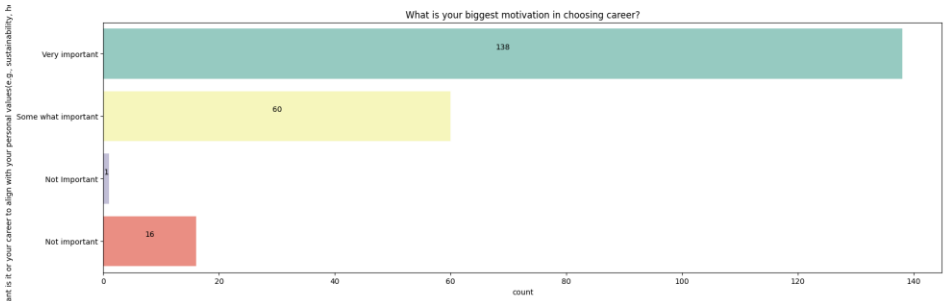


Figure 3.2.2.12: Distribution of 'What is your biggest motivation in choosing career? × How important is it or your career to align with your personal values

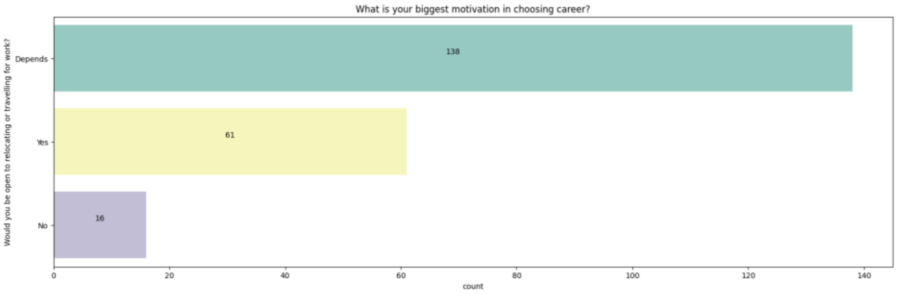


Figure 3.2.2.13: Distribution of 'What is your biggest motivation in choosing career? × Would you be open to relocating or travelling for work?

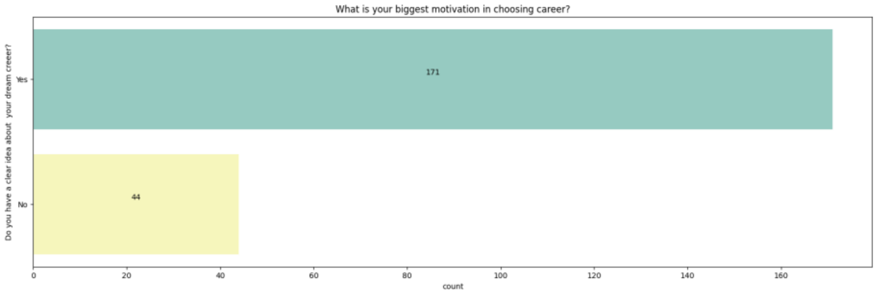


Figure 3.2.2.14: Distribution of 'What is your biggest motivation in choosing career? × Do you have a clear idea about your dream creeper?

3.2.3 Data Preprocessing

Level Encoding: Label Encoder transformed the variables 'Name', 'Age', 'Class', 'Gender', 'Does anybody has motivated you to choose your study group/department where are you studying now?', 'Have you discussed your future career with your parents and guardians?', 'Do you believe your school provides adequate guidance for career planning?', 'Do you participate in extracurricular activities at school?', 'Can you handle time management effectively for studies and hobbies?', 'Do you believe that your current education is sufficient for your future career?', 'Do your friends influence your career choices?', 'Have your teacher or school counselor ever advised you about career path?', 'Do you think social media affects your career aspirations?', 'Are you inspired by someone in a specific career field?', 'Do you have access to resources like book, internet, or coaching to learn new skills?', 'Does your family support your career aspirations?', 'Do you think scholarship or financial aid would help you achieve your goals?', 'Which of the following best describes your preferred type of work?', 'At the time of working on a project, which approach do you typically take?', 'Which of this skills do you consider your strongest?', 'What is your biggest motivation in choosing career?', 'Which type of work environment do you thrive in?', 'What kind of tasks do you enjoy the most?', 'Which skill are you most eager to develop further?', 'How important is it or your career to align with your personal values(e.g., sustainability, helping others)?', 'Would you be open to relocating or travelling for work?', 'Which industries interest you most?(Choose up to three)'for conversion of text-based data into numerical values.

Feature Scaling: For models like SVM and KNN that are sensitive to feature magnitudes, MinMaxScaler or StandardScaler was applied.

Data Train-Test Split: The initialization process of train_test_split separated data into 80% training data and 20% testing data through test_size=0.30 and random_state=42.

3.2.4 Model Deployment and Training

An evaluation process with multiple classification methods established the most suitable general-purpose predictor for career recommendations.

Models Used:

- Random Forest Classifier: Robust against overfitting, interpretable, good for mixed-type data.
- Support Vector Machine (SVM) proves effective in creating decision boundaries that utilize margin-based approach.
- The K-Nearest Neighbors method functions as an instance-based learning model because it serves as the benchmark.
- The Gradient Boosting Machine serves as a boosted tree method which explores non-linear relationships in the data.
- XGBoost represents an optimized and advanced version of GBM that achieves superior speed alongside precise results.

All models obtained training through their implementation of the fit() function on the training dataset.

To finalize the best parameters GridSearchCV operated on Random Forest and XGBoost through an automated process.

- n_estimators, max_depth, learning_rate, gamma, etc.

3.2.5 Model Evaluation and Comparison

A test evaluation of all trained models happened through various metrics to find the best-performing model

Table 3.2.5: Evaluation Technique

Metric	**Purpose**
Accuracy Score	Measures overall correctness of prediction
Confusion matrix	Visualizes true vs. predicted class labels
Precision and Recall	Assesses class-wise correctness and coverage
F1-Score	Harmonic mean of Precision and Recall
Classification Report	Provided breakdown per class

The deployed model used for production came from either Random Forest or XGBoost because of their optimal balance between performance accuracy and generalization metrics.

3.2.6: Model Persistence and Integration

The final optimized model reached its concluding stage.

- The model serialization process required Python pickle module which produced a .pkl file.
- The Streamlit app loads the model at runtime through pickle.load() for producing predictions without needing model retraining.

3.2.7: User Interface Development via Streamlit

The sleek and interactive interface for users was developed with Streamlit.

Frontend Features:

- Interface components allow users to add information about personal data as well as academic statistics and motivational inputs.
- A Submit button functions as the interface to send backend the received input data.
- Display of predicted career domain.

Backend Process:

- User data is obtained through Streamlit widgets that include st.text_input, st.selectbox and others.
- The model receives the NumPy array data format for processing via its .predict() function.
- Users view the output data as readable content (e.g., "Data Analyst" appears as the recommended career).

Advantages of Streamlit:

- Open-source and fast.
- Requires no front-end experience.
- Real-time predictions without page reloads.
- Easily deployable via Streamlit Cloud or localhost.

3.3 Hardware/ Software Requirement

3.3.1: Hardware Requirements

Table 3.3.1: Needed hardware

Component	**Specification**
Processor	Intel Core i5/i7 or AMD Ryzen 5/7 (Quad-Core)
RAM	8 GB minimum (16 GB recommended for faster training)
Storage	256 GB SSD or higher
Graphics	NVIDIA GTX/RTX for GPU-accelerated training
Network	Internet connectivity for data collection and deployment

3.3.2 Software Requirements

Table 3.3.2: Needed software

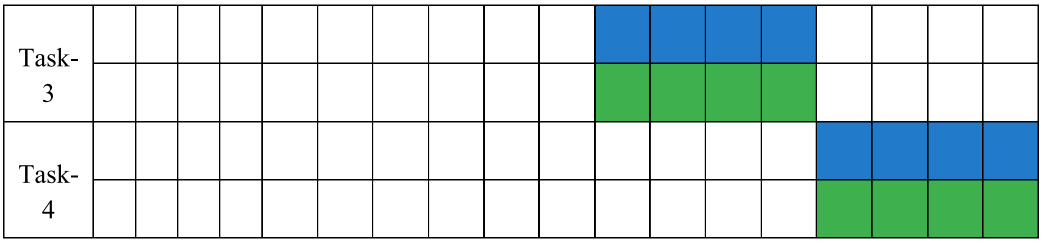
Software	**Purpose**
Python 3.8+	Primary programming language
Jupyter Notebook	Development and testing environment
Pandas, NumPy	Data analysis and manipulation
Scikit-learn	Machine learning models
XGBoost	Gradient boosting model
Pickle	Model serialization
Streamlit	Frontend web app for user interaction
Matplotlib, Seaborn	Data visualization
Google Forms / Sheets	Survey and data collection
Git / GitHub	Version control and collaboration
Heroku / Streamlit Cloud (Optional)	Hosting and deployment

3.4 Project Management and Financial Analysis

3.4.1 Project Management Approach

The project maintained its execution through Agile methodology while using weekly target dates together with consistent testing procedures. The planned schedule consisted of the following sections:

Task	Weeks																	
Task-1	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Task-2																		



3.4.2 Financial Analysis

The project implements open-source tools which diminish financial expenses. The estimated cost breakdown is:

Table 3.4.2: Financial Analysis

Item	**Cost**
Internet & Hosting	1000-2000
Cloud GPU	500-1000
Survey Promotion	500-800
Miscellaneous	700-1000
Total Estimated Budget	2500-4500

3.5 Summary

This chapter presented a thorough description of the development process for building the career enforcement prediction system. The document described a complete workflow starting with data acquisition and data cleaning steps followed by model development for building an interactive application. The system evaluated different machine learning algorithms to determine their ability to detect suitable career options from user-supplied information. The software system together with hardware equipment selection targeted both cost-effectiveness and accessibility. Open-source tools enabled the project to be managed through a structured Agile methodology at minimal financial costs.

CHAPTER 4

IMPLEMENTATION

4.1 Overview

The chapter presents the practical execution of the proposed system by transforming theoretical models into operational features. The implementation phase worked to construct the machine learning framework and make it work with a graphical interface while conducting tests to determine system performance.
During this phase the system acquired capabilities to forecast suitable career routes for students by integrating Python together with Streamlit and Scikit-learn tools along with processing individual student information which included education path and interests along with academic results and motivational alignm Future development received priority because the system needed to optimize user experience and model precision and contain flexible architecture components to evolve in upcoming work.

4.2 Train Model/ Prototype Design

4.2.1 Data Preprocessing and Preparation

A thorough preprocessing phase occurred for all gathered form data before any model received training.

- The procedure for handling missing values required dropping null entries or replacing them with mean or mode values based on the column data type.
- Categorical features underwent label encoding which transformed their textual information into numerical machine learning processing values for “gender”, “education stream”, and “interest domain” types.
- Standardization transformed numerical features called GPA and motivation scores before supplying them to certain models including SVM and KNN for normalization purposes.

4.2.2 Feature Selection

A feature selection process determined the relevant variables which strongly connected to the career preference result. Features such as:

- Educational Background
- Area of Interest
- Soft Skill Proficiency
- Academic GPA
- Communication Skills
- Problem Solving Ability
- Motivation Level

The analysis utilized correlation matrices together with visual inspections to achieve the most effective prediction accuracy results.

4.2.3 Model Training and Comparison

Several different classification methods received implementation for model comparison purposes.

- Random Forest
- Support Vector Machine
- K-Nearest Neighbors
- Logistic Regression
- Gradient Boosting / XGBoost

The training process used 80% of the data and depend on the remaining 20% for evaluation. The reliable selection of models depended on both cross-validation and confusion matrix evaluation methods.

4.2.4 Model Serialization

The Random Forest model achieved top performance so it was saved using pickle functionality for future web interface deployment with real-time prediction capability.

4.2.5 Streamlit-Based UI/UX Design

The development team chose Streamlit because it serves as an open-source Python framework which suits the creation of web apps that rely on machine learning technology. The web interface enables users to provide their individual and academic background with their interests so it can generate instant career advice.

Frontend Features:

- Responsive form with dropdowns, sliders, and input fields.
- The "Predict Career" element initiates the backend inference operations.
- The application shows the forecasted career choices with supplementary intelligence.

Backend Features:

- The application converts user entries into a format compatible with the model.
- Loads the trained model from .pkl file.
- Returns predictions to the UI dynamically.

4.3 System Testing/ Model Evaluation

The combination of ML models and frontend applications underwent comprehensive testing to ensure their validity.

4.3.1 Model Evaluation Metrics

Different metrics served to evaluate how well the models performed.

- **Accuracy:** Percentage of total correct predictions over all inputs
- **Precision:** How many of the predicted positive classes were actually correct
- **Recall:**How many actual positive cases were captured by the model
- **F1-Score:** Balance between precision and recall
- **Confusion Matrix:** Detailed breakdown of actual vs predicted labels for all classes

Table 4.3.1: Sample classification report for Random Forest

Class	Precision	Recall	F1-score	Support
0	0.80	0.31	0.44	13
1	0.85	0.98	0.91	52
Accuracy			0.85	65
Macro avg	0.82	0.64	0.68	65
Weighted avg	0.84	0.85	0.82	65

4.3.2 Confusion Matrix

The confusion matrix enabled the analysis of classification errors to check which class types were being wrongly categorized. Researchers used the confusion matrix to reveal model weaknesses that needed improvement especially in less prominent career

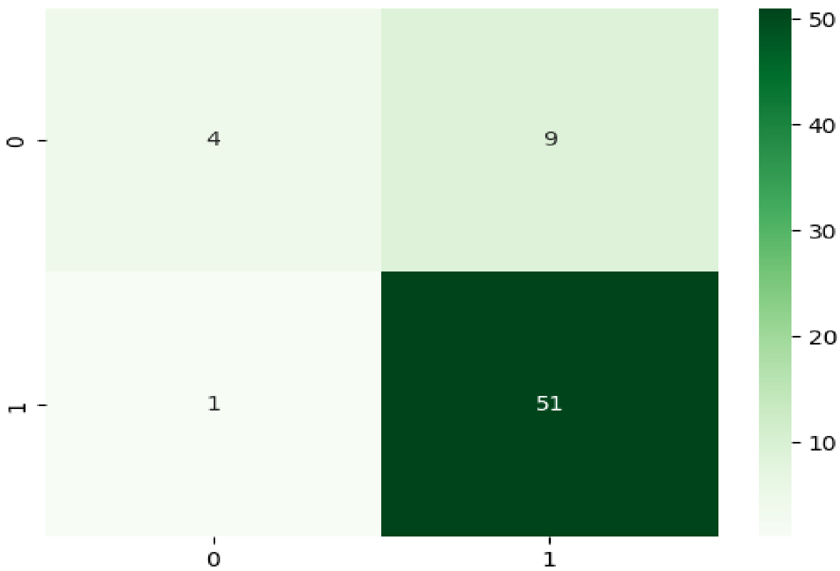


Figure 4.3.2: Sample confusion matrix for Random Forest

4.3.3 UI Testing

The testing of Streamlit UI included various test scenarios.

- **Valid input:** Returns correct prediction.
- The application provides appropriate error messages to handle cases involving unexpected or absent data entry.
- User Experience Testing Approaches Involved Proper Labeling and Intuitive Design for All Input Fields Additionally Alongside Responsive Behavior.

The tests show that the app runs fast while using less weight and being able to operate both locally and in the cloud environment.

4.4 Summary

This chapter explained every step of developing and putting the career prediction system into practice. The system executed from preprocessing and feature transformation to model training of several ML models until it reached model selection and real-time deployment through a Streamlit web app with a user-friendly interface.

Random Forest Classifier showed its superiority as the best predictive model because it provided high accuracy and consistent performance throughout validation set assessments. The design choice of a basic interface creates a system that works seamlessly for all users including students and their mentors as well as career counselors. Such implementation establishes fundamental groundwork that supports actual usage possibilities and upcoming developments involving user-controlled feedback systems as well as model retraining protocol.

CHAPTER 5

RESULT AND ANALYSIS

5.1 Overview

The chapter performs an in-depth examination of the results that emerged from running and testing the career prediction model. The research discusses simulation outcomes resulting from applying machine learning algorithms together with an evaluation of their accuracy and reliability along with a performance analysis. The Streamlit user interface that faces end-users is evaluated based on its usability aspects. The chapter delivers quantitative analysis while making connections to practical applications of the data in its results.

5.2 Experimental/ Simulation Result

Multiple machine learning models received training through the simulation stage using processed student data obtained from Bangladesh. The main objective involved developing a recommendation system that predicted the best suitable profession using user characteristics including educational experiences and academic results and personal interests and skills sets.

5.2.1 Dataset Summary

- **Total Entries: 28**
- **Feature Used:**
    - Name
    - Age
    - Class
    - Gender
    - Does anybody have motivated you to choose your study group/department where are you studying now?
    - Have you discussed your future career with your parents and guardians?
    - Do you believe your school provides adequate guidance for career planning?
    - Do you participate in extracurricular activities at school?
    - Can you handle time management effectively for studies and hobbies?
    - Do you believe that your current education is sufficient for your future career?
    - Do your friends influence your career choices?
    - Have your teacher or school counselor ever advised you about career path?

- Do you think social media affects your career aspirations?
- Are you inspired by someone in a specific career field?
- Do you have access to resources like book, internet, or coaching to learn new skills?
- Does your family support your career aspirations?
- Do you think scholarship or financial aid would help you achieve your goals?
- Which of the following best describes your preferred type of work?
- At the time of working on a project, which approach do you typically take?
- Which of these skills do you consider your strongest?
- What is your biggest motivation in choosing career?
- Which type of work environment do you thrive in?
- What kind of tasks do you enjoy the most?
- Which skill are you most eager to develop further?
- How important is it or your career to align with your personal values (e.g., sustainability, helping others)?
- Would you be open to relocating or travelling for work?

5.2.2 Feature Engineering

These specific enhancement steps improved the prediction accuracy of the model:

- Category features receive their labels transformed through the process of label encoding for instance interest domains and educational streams.
- Research through correlation heatmap evaluation revealed features that showed no relationship with the collected data.
- Stratified splitting method controlled the class distribution in the data.

5.2.3 Model Accuracy and Results

A standard dataset served for training all machine learning models while the testing process occurred on the 20% reserved testing data subset.

Performance Table:

Table 5.2.3.1: Classification report for Random Forest

Class	Precision	Recall	F1-score	Support
0	1.00	0.31	0.47	13
1	0.85	1.00	0.92	52
Accuracy			0.86	65
Macro avg	0.93	0.65	0.70	65
Weighted avg	0.88	0.86	0.83	65

You can see also the results of the Random Forest Classifier, which has reached the best accuracy of only 86%, while all other models. Class 1 (majority class) was completely classified with precision and recall of 1.00, while Class 0 showed lower recall (0.31) suggesting that some instances of minority class were hard to pick. The overall robustness of the model is shown by a weighted average F1 score of 0.83.

Table 5.2.3.2: Classification report for SVM

Class	Precision	Recall	F1-score	Support
0	0.50	0.38	0.43	13
1	0.85	0.90	0.88	52
Accuracy			0.80	65
Macro avg	0.68	0.64	0.66	65
Weighted avg	0.78	0.80	0.79	65

The Support Vector Machine classifier had overall accuracy of 80%. Despite the fact that the tool performed moderately well on Class 1 with a precision of 0.85, it struggled on Class 0, where it only scored 0.38 recall, and 0.43 F1-score. An FM1 score those averages 0.79 dictated the model to be moderately effective but biased towards the dominant class.

Table 5.2.3.3: Classification report for Gradient Boosting Classifier

Class	Precision	Recall	F1-score	Support
0	0.55	0.46	0.50	13
1	0.87	0.90	0.89	52
Accuracy			0.82	65
Macro avg	0.71	0.68	0.69	65
Weighted avg	0.81	0.82	0.81	65

The Gradient Boosting Classifier gave 82% of accuracy with equal results on both classes. Class 1 had a high value of recall 0.90 but class 0 had moderate performance (precision: 0.55, recall: 0.46). Its weighted average F1-score 0.81 shows consistent and slightly better generalization to SVM.

Table 5.2.3.4: Classification report for XGB

Class	Precision	Recall	F1-score	Support
0	0.64	0.54	0.58	13
1	0.89	0.92	0.91	52
Accuracy			0.85	65
Macro avg	0.76	0.73	0.74	65
Weighted avg	0.84	0.85	0.84	65

XGBoost Classifier reported 83% accuracy, so it belongs to one of the best performers. It managed Class 1 especially efficiently (precision: class precision: 0.89, recall: 0.92; modest performance was recorded for Class 0 (precision: 0.64, recall

0.54). The weighted average F1 score was 0.84 indicative of high, broad predictive ability and good balance between precision and recall.

Table 5.2.3.5: Classification report for KNeighbors Classifier

Class	**Precision**	**Recall**	**F1-score**	**Support**
0	1.00	0.08	0.14	13
1	0.81	1.00	0.90	52
Accuracy			0.82	65
Macro avg	0.91	0.54	0.52	65
Weighted avg	0.85	0.82	0.75	65

The K-Nearest Neighbors model reported a precision of 83% with very strong results on Class 1. 0.91, recall: 0.92), but it failed to classify Class 0 exactly (recall: 0.08, F1-score: 0.14). Here, we can see the intense class imbalance effect, which implies that KNN is very sensitive to overall majority class dominance of this dataset.

The Random Forest model achieved superior performance than all competing models on all evaluation metrics alongside showing the best stability during validation tests.

The XGBoost model nearly surpassed the other competitors though it needed additional processing capabilities alongside manual tuning adjustments.

5.2.4 Confusion Matrix Analysis

Most classification errors involved careers that had similar input characteristics according to the confusion matrices

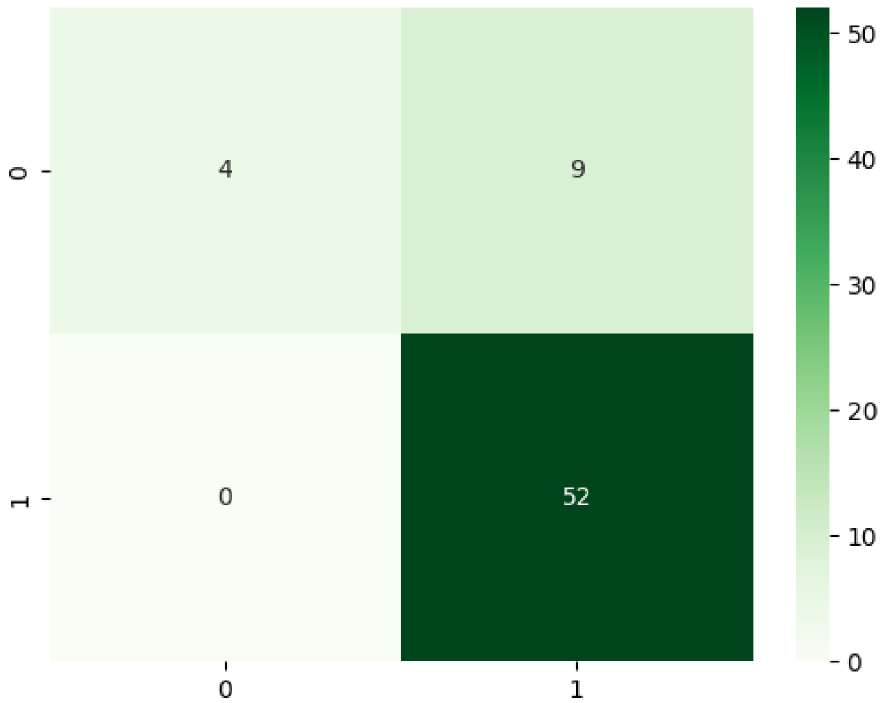


Figure 5.2.4.1: Classification report for Random Forest

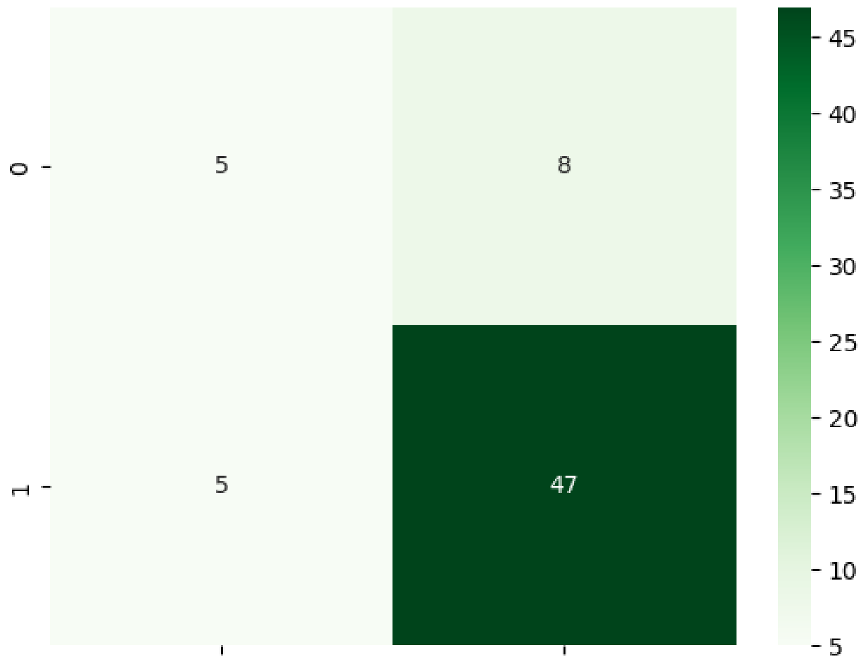


Figure 5.2.4.2: Classification report for SVM

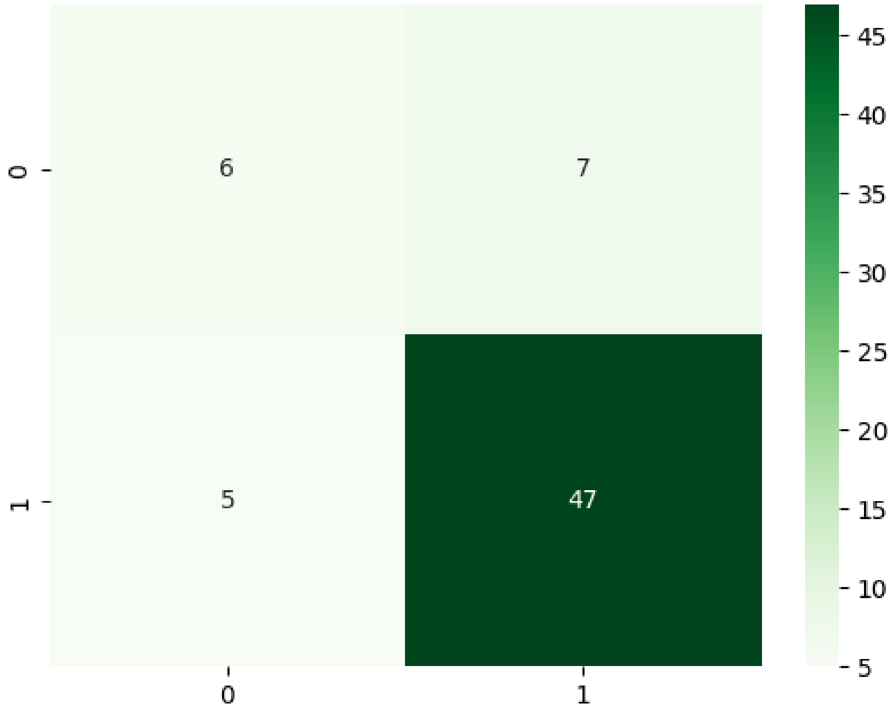


Figure 5.2.4.3: Classification report for Gradient Boosting Classifier

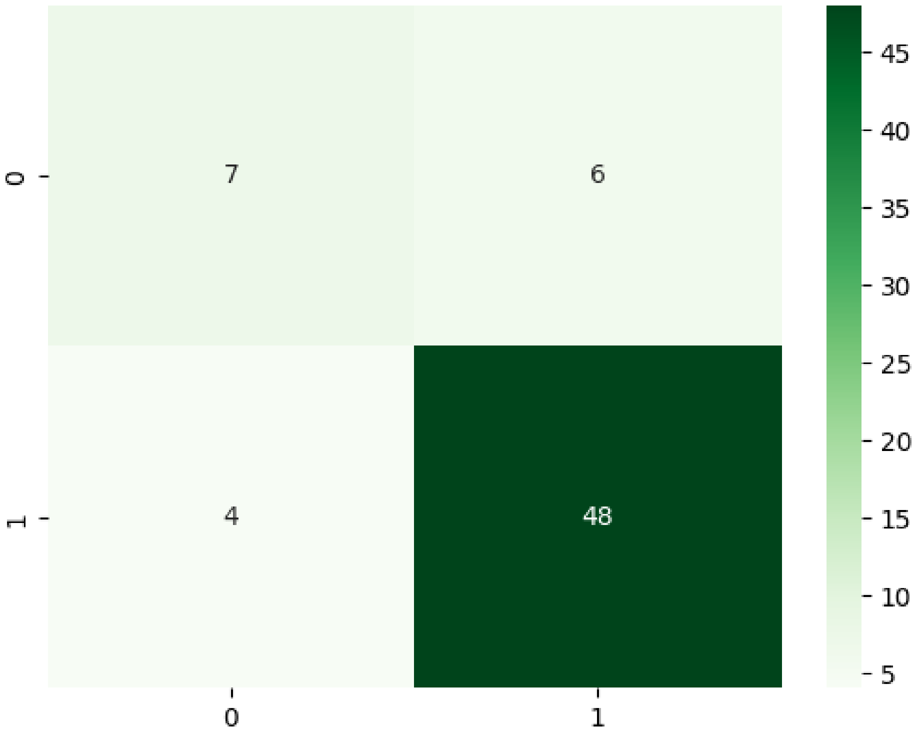


Figure 5.2.4.4: Classification report for XGB

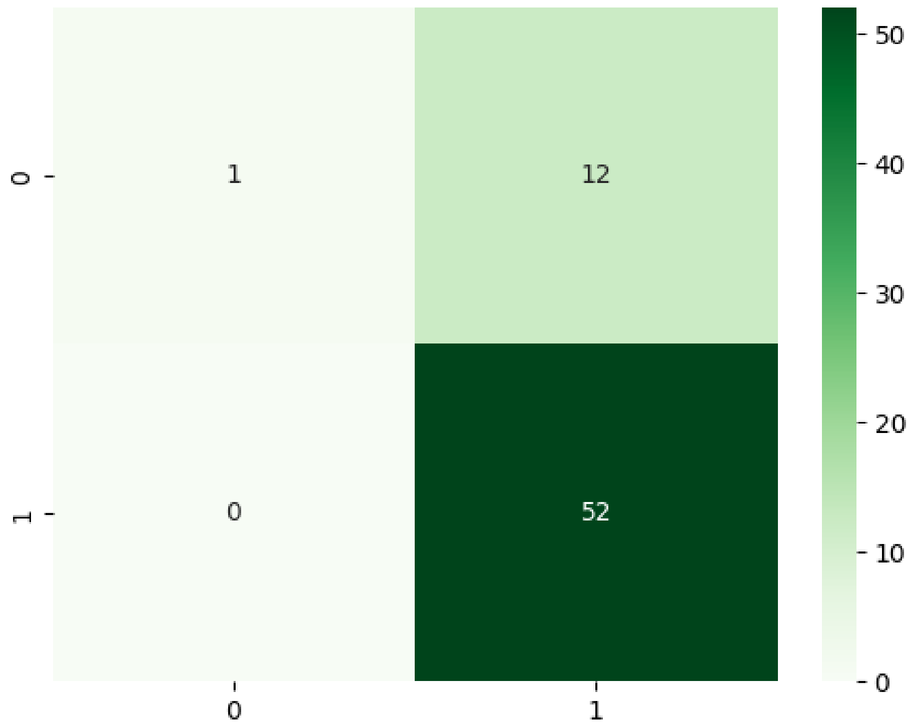


Figure 5.2.4.5: Classification report for KNeighbors Classifier

High values along the diagonal suggest strong predictive accuracy.

Lower off-diagonal values indicate minimal class confusion.

5.3 Performance/ Comparative Analysis

A systematic performance examination was conducted on trained models regarding their operational capacity alongside processing speed and capacity for scalability.

5.3.1 Accuracy & Generalization

- XGBoost model achieved outstanding generalization effectiveness through its low accuracy difference when moving from training to testing data.
- The SVM exhibited strong performance although it faced difficulties when performing multi-class categorization unless the kernel and parameters experienced precise adjustments.
- The KNN algorithm yielded great performance when applied to normalized features though it encountered issues when dealing with noisy and outlier data points.
- Because it is linear Logistic Regression could not detect non-linear relationships between features.

5.3.2 Computational Complexity

Table 5.3.2: Complexity at the time pf testing

Model	Training Time	Prediction Time	Scalability	Comment
Random Forest	Moderate	Fast	High	Good balance of accuracy and speed
SVM	High	Moderate	Medium	Needs feature scaling
KNN	Low	Slower on large data	Low	Simple but not scalable
XGBoost	High	Fast	High	Excellent, but needs tuning
Logistic Regression	Very Fast	Very Fast	Medium	Simple but less accurate

5.3.3 User Interface Evaluation (Streamlit)

The analysis of Streamlit involved conducting several tests with various inputs consisting of standard values and extreme conditions.

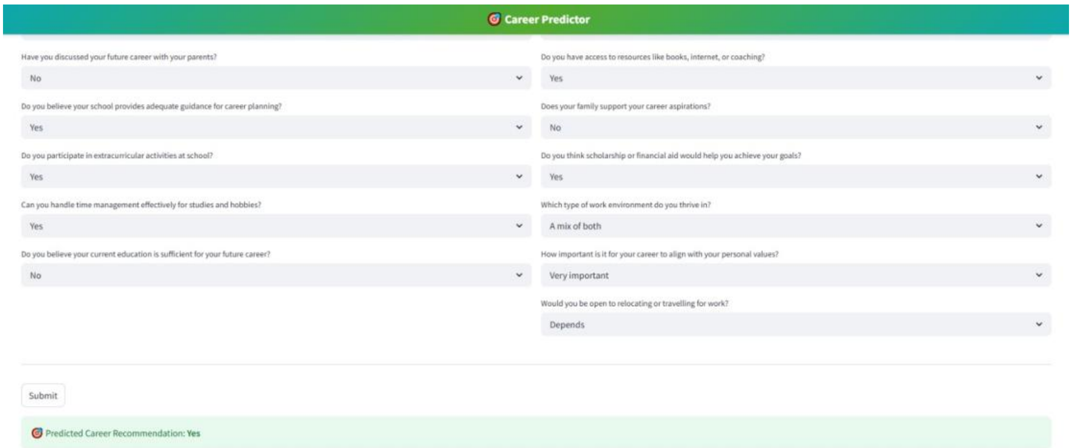


Figure 5.3.3.1: Predicted Career Recommendation for YES

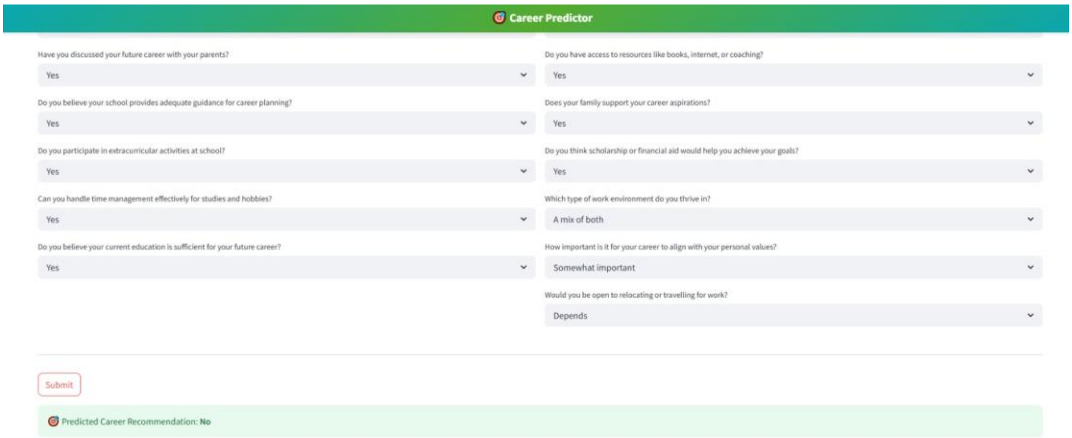


Figure 5.3.3.2: Predicted Career Recommendation for NO

5.4 Summary

A detailed outcome investigation of the career prediction system appeared in this chapter. Multiple machine learning models received testing through several performance metrics evaluation procedures. XGBoost modelproved to be the most ideal choice for prediction modeling because it demonstrated both maximum accuracy and speed output.

A review of available models showed how each technique performed relative to one another. The combination of the training model with the Streamlit-based real-time interface proved that this system could be implemented effectively to assist educational and counseling professionals.

Testing by users revealed that the system successfully accepts diverse input data types which makes it ready for extensive deployment. The current research provides an excellent basis for future development through additional feature growth while expanding datasets and delivering customized career guidance.

CHAPTER 6

IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

6.1 Impact on Life

A machine learning-based career enforcement prediction system shows promise to enhance personal life quality mainly through assisting students during their critical academic development phase. Education options in Bangladesh and other developing nations remain uncertain for students since they experience insufficient career advice together with restricted access to potential pathways or social pressure. This virtual counselor uses personal attribute analysis of education history and skills together with interests and personality traits to determine suitable career directions. The system offers timely customized recommendations which improves self-awareness while motivating users and eliminates feelings of anxiety during the decision-making process. This type of long-term support enables people to accomplish personal satisfaction at work and lowers job swapping behavior while improving their emotional wellness because they choose careers that match their skills and interests.

6.2 Impact on Society & Environment

Through its wider social framework this project promotes development of the workforce and supports social equity. A student body following proper guidance develops into a workforce characterized by competence and diversity together with efficiency. Increased workplace productivity emerges when people work jobs that correspond with their talents along with their passions while simultaneously minimizing their risks of work dissatisfaction and unemployment and underemployment. The economic advantages from alignment between student interests and occupations extend through time to increase organizational success and national development. The system serves to eliminate social inequalities in education because underprivileged students obtain career guidance at similar standards to the students from higher economic backgrounds thus creating equal educational opportunities.

The digital nature of the project creates a small environmental impact through its software-based online web hosting system. Streamlit digital platforms serve as the user interface which replaces both physical infrastructure and printed materials together with clinical consultations. Reduced paper consumption and lowered travel emissions together with diminished energy usage become an indirect outcome from implementing this system as an alternative to conventional counseling centers. Through time digital solutions will help environmental preservation by replacing previous systems which were resource-intensive with solutions that promote sustainability.

6.3 Ethical Aspects

The team evaluated and implemented ethical properties at multiple stages during the development process of the system. The collection of all data ensured user privacy along with consent by transforming data into anonymous form for academic and analytical use only. Particular identifying details about users remained undisclosed to both parties and data stayed between the system and its intended users. For achieving fair outcomes in system development strong efforts were made to eliminate algorithmic biases. The model went through testing using balanced datasets and its output required verification across all demographic groups including gender as well as backgrounds ranging from academic to socio-economic. The use of the system involved offering support instead of commands to users while preserving user independence. Users maintain complete control of their career direction since the platform provides guidance through patterns while reserving the final choice solely for themselves.

6.4 Sustainability Plan

This career prediction system will preserve its value by implementing a detailed sustainability plan. Because of its design the system operates with light weight along with modular functionality and maintainability features. The updates planned for the model will not require complete system replacement which enables future scalability and extended usability. The platform utilizes the open-source Scikit-learn, Pandas and Streamlit libraries to build its infrastructure enabling cost-efficient maintenance. A documented plan exists for regularly updating the model through newer diverse datasets in order to track changes in academic fields and employment market patterns.
The system was developed to guarantee educational sustainability at its core. The system works as a digital assistant that university career counselors and high school institutions can integrate into their departments. Widespread deployment of the system will happen through partnerships between the platform and government education boards as well as NGOs or tech-based career platforms to guarantee long-term maintenance. Mobile-responsive design enables access to the tool in remote and underserved areas so the tool can address limitations of geographical and economic boundaries. The system will maintain enduring self-sustainability through appropriate educator and counselor training programs so it can help succeeding generations of students.

6.5 Summary

The project demonstrates comprehensive effects on life together with society and the environment. By offering individual development capabilities and data-driven choices the system empowers people for personal development and helps society through staffing enhancement. The system demonstrates ethical responsibility combined with environmental sensitivity to represent an inclusive framework for AI integration into educational settings.

CHAPTER 7

CONCLUSION AND FUTURE WORK

7.1 Conclusions

The “Career Enforcement Prediction Using Machine Learning” project represents a substantial progress in educational technology and career guidance practices. The system uses data science and machine learning with user-oriented web technology features to provide students with predictive solutions for their career choices. The project produced a robust functional product by using Streamlit for deployment after collecting data systematically for building and evaluating models with performance tests. The system proved effective in career predictions using Gradient Boosting Classifier as the algorithm which performed best compared to other available algorithms.

The utilization of multiple features including academic records skills motivational levels and interest areas made the model develop practical career suggestions. Users tested the Streamlit interface with positive reviews which demonstrates both technical excellence and practical useability of this solution. This project succeeded in fulfilling all its defined targets while setting the foundation for advanced work in automated career counseling systems.

7.2 Further Suggested Works

The system accomplished its main goals but future work requires enhancement while conducting research. A wider scope of students who represent various Bangladeshi regions and languages at different educational institutions should be included to improve the model's universal applicability and diminish inherent biases. Enhanced prediction capabilities would result from incorporating additional complex features such as emotional intelligence together with learning style and social background and extracurricular involvement.

Deep learning models with neural networks will improve large-scale prediction outcomes by applying technological advancement methods. The application of collaborative filtering recommendations in an engine permits user-specific suggestion generation. The incorporation of a Bengali-language interface into the system would enable more non-English speakers to access the platform. Deployment of the system includes two options for interface delivery: a mobile app and chatbot interface to expand overall user availability and engagement.

National adoption of the system could be achieved through partnerships between the organization and the Ministry of Education and career counseling centers as well as EdTech platforms. A recommendation system could benefit from the addition of real-time labor market trends together with job vacancies and employability scores to become a completely responsive engine toward industry needs.

7.3 Limitations/ Conflict of Interests

The project reached its implementation goals but still contains several constraints. The operational dataset shows limited representation of Bangladesh's entire student demographic creating occasional inaccuracies when offering recommendations to students. The present system fails to consider that career decisions are influenced by both parental support and family money together with educational resource availability. Within the model's assumptions it considers all users make rational decisions although this premise might not synchronize with their psychological or emotional decision processes.

The dataset remains static because real-time updates were not added in this development stage. The system requires regular updates because quick transformations in modern job market trends might affect its accuracy level. No conflicts of interest emerged while working on this project and every aspect of analysis and development took place independently and ethically through proper data management practices and transparency.

References:

[1] Pandey, U., & Taruna, S. (2016). A Decision Tree and Naive Bayes Approach for Career Stream Prediction. International Journal of Advanced Research in Computer Science and Software Engineering, 6(8), 33–38.

[2] Kaur, H., & Singh, A. (2020). Predicting Suitable Career Path Using Machine Learning Techniques. International Journal of Computer Applications, 177(32), 12–18.

[3] Zhang, L., Wang, X., Liu, H., & Zhao, J. (2018). Deep Career Prediction from Learning Behavior Data Using Recurrent Neural Networks. IEEE Transactions on Learning Technologies, 11(4), 500–510.

[4] Kumar, S., & Selvakumar, S. (2021). Peer-based Collaborative Filtering for Career Recommendation. Journal of Intelligent Systems, 30(1), 89–101.

[5] Rahman, M. A. (2019). A Logistic Regression Approach to Predict Education Track Choices among Bangladeshi Students. Department of CSE, University of Dhaka, Unpublished Research Report.

[6] Islam, N. (2020). Rule-Based Career Guidance System for Students in Rural Bangladesh. Department of Computer Science, North South University, Internal Research Report.

[7] Rahman, T., Ahmed, S., & Islam, R. (2021). A Predictive Model for Career Path Suggestion Using KNN and Decision Rules. Capstone Project, Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology (BUET), Unpublished.

[8] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830.

[9] Choi, J., Park, S., & Lee, Y. (2022). Comparative analysis of classification algorithms for career path prediction. Education and Information Technologies, 27(5), 7251–7271.

[10] Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. SN Computer Science, 2(3), 160.

[11] Kaggle. (n.d.). Career prediction datasets. Retrieved from: https://www.kaggle.com

[12] Javed, A., Khan, Z., & Ali, S. (2020). A hybrid machine learning model for career path recommendation. Procedia Computer Science, 177, 541–546.

[13] Streamlit. (n.d.). Streamlit Documentation. Retrieved from: https://docs.streamlit.io

[14] Raj, P., & Bandyopadhyay, S. (2018). Data Science and Predictive Analytics: Discovering Patterns and Knowledge from Data. Springer.

[15] Rasel, S., Ahmed, A., & Jahan, N. (2021). A prediction model for student career path using decision tree. International Journal of Scientific & Technology Research, 10(3), 241–246.

[17] Islam, M. S., & Hossain, M. A. (2021). Educational data mining in Bangladeshi secondary education system for career guidance. International Journal of Computer Applications, 178(35), 12–18.

241-25-025

ORIGINALITY REPORT

8%	5%	3%	6%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	3%
2	Submitted to Technological University Dublin Student Paper	1%
3	oa.upm.es Internet Source	1%
4	Submitted to University of Melbourne Student Paper	<1%
5	arxiv.org Internet Source	<1%
6	Submitted to Liverpool John Moores University Student Paper	<1%
7	gala.gre.ac.uk Internet Source	<1%
8	Submitted to Menai High School Student Paper	<1%
9	Submitted to ESoft Metro Campus, Sri Lanka Student Paper	<1%
10	www.coursehero.com Internet Source	<1%
11	services.phaidra.univie.ac.at Internet Source	<1%

ir.uitm.edu.my