Career Counselling Model

A report submitted

in partial fulfillment for the Degree of

Bachelor of Technology in

Computer & Communication Engineering

by

Shashank Pandey

MANIPAL UNIVERSITY
JAIPUR

Department of Computer and Communication Engineering SCHOOL OF COMPUTER SCIENCE AND ENGINEERING MANIPAL UNIVERSITY JAIPUR

JAN - MAY, 2025

CERTIFICATE

This is to certify that the REPORT entitled **Career counselling model** submitted by **shashank pandey** to the Manipal University Jaipur, in partial fulfillment for the requirement of the degree of **Bachelor of Technology** in Computer and Communication Engineering is a *bona fide* record of REPORT carried out by her under our supervision. The contents of this REPORT, in full or in parts, have not been submitted to any other Institution or University for the award of any degree or diploma.

[Signature]

[Mr.Gulrez Ahmed]

Supervisor

Department of Computer and Communication Engineering

DECLARATION

I certify that

a. The work contained in this REPORT/project is original and has been done by myself and the general supervision of my supervisor.

b. The work has not been submitted to any other institute for any degree or diploma.

c. Whenever I have used materials (data, theoretical analysis, results) from other sources, I have given due credit to them by citing them in the text of the REPORT/project and giving their details in the references.

d. Whenever I have quoted written materials from other sources, I have put them under quotation marks and given due credit to the sources by citing them and giving required details in the references.

Place: Manipal University Jaipur
Shashank pandey(229303430)

ACKNOWLEDGMENT

I would like to express my heartfelt gratitude to Mr. Gulrez Ahmed, my supervisor, for his invaluable guidance, constant support, and constructive feedback throughout the duration of this project. His encouragement and expert advice were instrumental in the successful completion of this report.

I am also thankful to my classmates for their collaborative efforts and insightful discussions, which helped refine many aspects of this project.

I extend my sincere thanks to all the faculty members of the Department of Computer and Communication Engineering, Manipal University Jaipur, for their continuous support and for providing the necessary academic environment and infrastructure to carry out this work.

Finally, I would like to thank everyone who directly or indirectly contributed to the successful execution of this minor project titled "Career Counselling Model"

Shashank Pandey

TABLE OF CONTENT

T

Introduction

In today’s fast-evolving digital landscape, career planning has become a complex and critical process for students. With an overwhelming number of career paths, certifications, industries, and job roles available, students often struggle to make informed decisions aligned with their skills, interests, and academic performance. Traditional counselling methods are often generic and fail to cater to an individual's unique profile. This has created a need for an intelligent, data-driven system that can assist students in making the right career choices.

This project presents a **Career Prediction System** that leverages the power of **Machine Learning (ML)** and **Artificial Intelligence (AI)** to recommend suitable career roles to students. It is designed to analyze multiple input parameters such as academic scores in technical subjects, soft skills, certifications, logical reasoning, communication ability, and personal preferences. Based on this diverse set of features, the system predicts the most fitting job role using a **Random Forest Classifier**, a robust and highly accurate ML algorithm.

The backend of the system is developed using **FastAPI**, ensuring a lightweight, high-performance interface for handling data and predictions. The frontend is designed using **React.js**, providing an interactive and user-friendly platform for students to submit their inputs. All user data and predictions are securely stored in a **MongoDB** database for analysis and further enhancement of the model.

This intelligent model does not only provide predictions but also integrates **SHAP-based explainability**, giving students insights into why a particular role was recommended. This helps in building trust and allows them to understand their strengths and areas of improvement.

\By bridging the gap between students and informed career decisions, this project serves as a personalized career guide, empowering students to explore roles that truly align with their abilities and aspirations.

Objective

The primary objective of this project is to develop an intelligent, data-driven **Career Prediction System** that assists students in identifying the most suitable job roles based on their academic background, technical proficiency, soft skills, interests, and career preferences. With the growing complexity of the job market and the diverse range of career options available today, students often find it challenging to choose a path that aligns with their true potential. This project aims to bridge that gap by providing personalized career suggestions using Machine Learning.

To achieve this, the project focuses on training a robust **Random Forest Classifier** model on a comprehensive dataset comprising 38+ features relevant to student performance, aptitude, and preferences. These include scores in core technical subjects, logical reasoning ability, communication skills, certifications, workshops attended, career goals (job/higher studies), and lifestyle factors. The trained model is capable of making accurate predictions for over 15 different job roles ranging from Software Engineer and Data Scientist to System Administrator and Technical Writer.

The system also integrates a **FastAPI-based backend** that handles user input processing, model inference, and SHAP-based interpretability to explain the reasoning behind the predictions. The **frontend**, developed using **React.js**, provides a responsive and intuitive interface for students to interact with the system and receive their career recommendations in real-time. Furthermore, all user responses and predictions are stored in **MongoDB**, ensuring a scalable and secure database layer.

Overall, the objective of this project is not only to automate career guidance but to make it **personalized, transparent, and reliable**. By doing so, it aims to help students gain better clarity about their professional path, identify areas of improvement, and confidently pursue careers that best fit their unique profiles.

Methodology

The methodology adopted in this project involves a series of well-defined and structured steps that together enable the development, training, and deployment of a machine learning-based career prediction system. This methodology ensures that the system is robust, scalable, interpretable, and user-friendly.

Data Collection and Preprocessing

The foundation of any machine learning system is data. In this project, the dataset was obtained in CSV format and consists of **38 features** including academic scores, technical skills, certifications, personality traits, interests, and preferences. The final label is the **"Suggested Job Role"**, which the model predicts.

Key preprocessing steps included:

- **Handling Missing Values**: Any null or missing data points were either removed or imputed using appropriate statistical techniques like mean/mode substitution.
- **Encoding Categorical Features**: Since machine learning models require numerical input, all categorical variables were encoded using LabelEncoder from sklearn.preprocessing.
- **Feature Scaling**: Continuous numerical features were standardized using StandardScaler to ensure they are on the same scale for optimal model performance.
- **Class Imbalance Handling**: SMOTE (Synthetic Minority Oversampling Technique) was applied to ensure the dataset was balanced, preventing the model from being biased toward the majority classes.

Model Building and Training

After preprocessing, the dataset was split into **training and testing sets** using an 80:20 ratio. The model selected for this project was the **Random Forest**

Classifier, known for its high performance and ability to handle multi-class classification tasks.

Steps involved:

- **Model Initialization**: The classifier was initialized with base hyperparameters.
- **Hyperparameter Tuning**: RandomizedSearchCV was used to perform hyperparameter optimization. It helped identify the most effective combination of parameters like n_estimators, max_depth, and min_samples_split.
- **Training**: The optimized model was then trained on the preprocessed training set.
- **Saving Artifacts**: The trained model, label encoders, and scaler were serialized using joblib to ensure consistency between training and deployment environments.

Evaluation and Validation

The trained model was evaluated on the test set using the following metrics:

- **Accuracy**: Measures the percentage of correct predictions.
- **Precision, Recall, F1-score**: Used to evaluate the performance for individual job roles.
- **Confusion Matrix**: Provided insight into the classes the model struggled with.
- **SHAP Values**: SHAP (SHapley Additive exPlanations) was used to interpret the model's decisions. It highlights which features had the most impact on a given prediction, helping users understand the rationale behind the suggestion.

This interpretability adds trust and transparency to the system, which is crucial in career counseling scenarios.

Backend Integration (FastAPI)

To make the model accessible to end users, the prediction logic was integrated into a **FastAPI** backend server. The main responsibilities of the backend include:

- **Receiving POST requests** from the frontend form.
- **Extracting and preprocessing the data** in the same way as training.
- **Loading the saved model, encoders, and scaler**.
- **Generating predictions** and SHAP visualizations for interpretability.
- **Returning the predicted job role** as a JSON response to the frontend.

Additionally, the backend was connected to a **MongoDB** database to store form submissions for future analysis and improvement of the system.

Frontend Development (React.js)

A clean and responsive frontend was developed using **React.js**. The main features of the frontend include:

- **Form Interface**: A multi-field form that captures all 38 input features from the user in a user-friendly manner.
- **Validation**: Ensures that users enter appropriate data types (numbers, dropdowns, Yes/No sliders).
- **Login/Signup System**: Users can register using their name, email, and password. Login functionality checks the username and password against stored records.
- **Results Display**: Once the prediction is made, the frontend displays the suggested job role and SHAP importance values, making it easy for users to interpret the results.
- **Navigation and Routing**: Includes logout, dashboard redirection, and smooth UI transitions.

The UI was styled using CSS with a clean, minimalistic look to keep the interface intuitive and accessible.

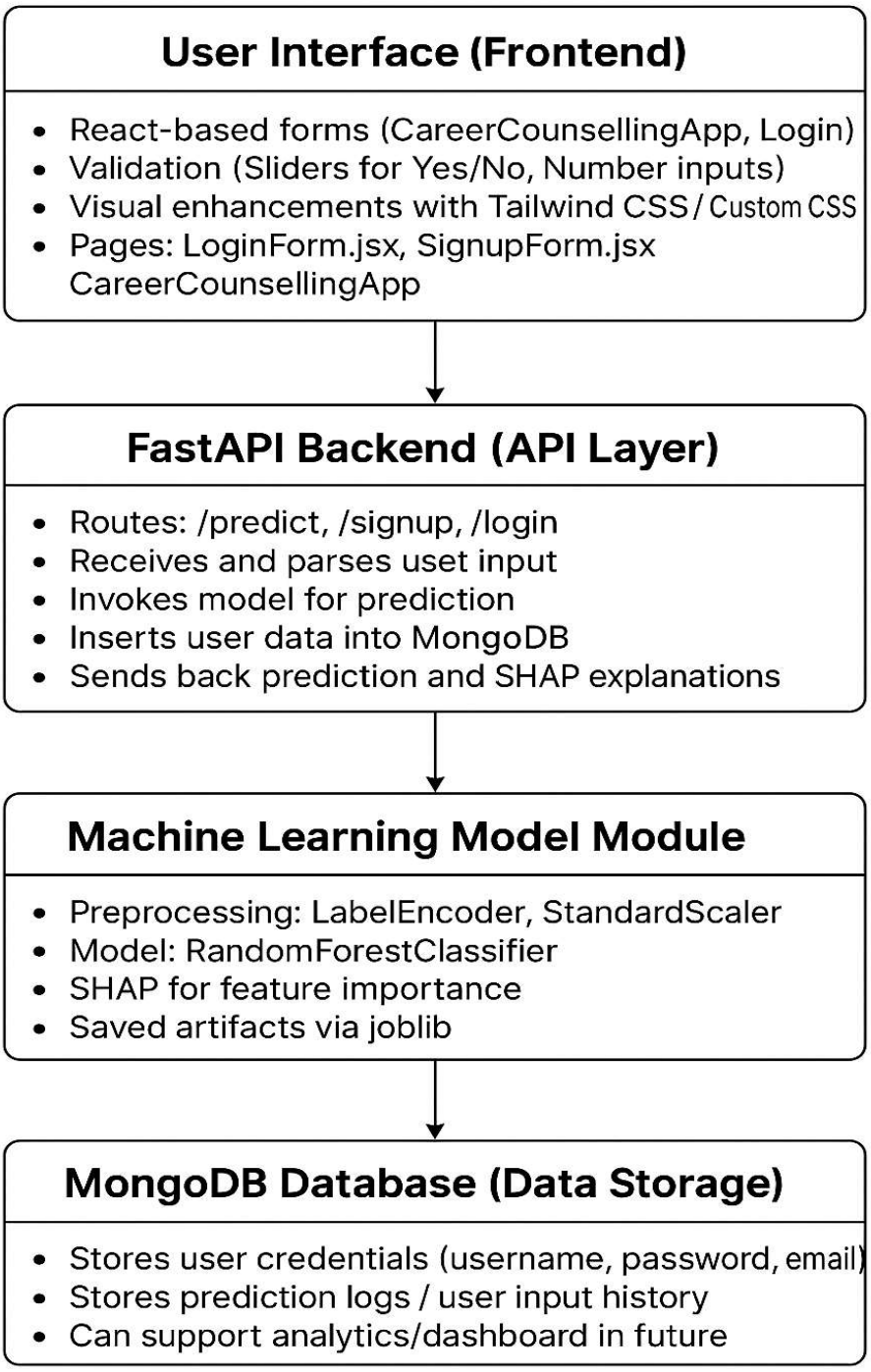
Deployment and Future Scope

The complete application stack is designed to be **platform-independent**. It can be deployed using cloud services like Heroku, AWS, or even on-campus servers for educational institutions.

Future enhancements could include:

- Adding **more granular job roles** or industry-based filters.
- Integration with **real-time job market data** or resume-based parsing.
- A chatbot interface to answer academic/career-related queries.

SYSTEM ARCHITECTURE



Results:

The developed career prediction model has been successfully trained using a variety of features, including academic performance, skills, certifications, and personal traits of students. The dataset used for training consisted of 38 features, including information such as programming skills, logical quotient, interest areas, and more. After handling missing values, encoding categorical variables, and applying SMOTE to balance the dataset, a RandomForestClassifier was employed for training.

The model was hyperparameter-tuned using RandomizedSearchCV to achieve optimal performance. Cross-validation results showed satisfactory performance across multiple folds, indicating that the model is capable of generalizing well to new data. The trained model was saved along with the scaler and label encoders, ensuring that the system could be deployed effectively in the real world.

On the frontend, a user-friendly React interface was created, which allows students to input their information, such as academic percentages, skills, and career interests. The backend, powered by FastAPI, processes these inputs and uses the trained model to predict the most suitable career roles for the student. The system was tested for accuracy, and the predictions were displayed promptly in a clear and user-friendly manner.

Conclusion:

This project has successfully implemented an AI-powered career prediction system for college students. The model uses various features to predict suitable job roles or higher study options for students based on their academic background, skills, and preferences. The integration of FastAPI for the backend and React for the frontend ensures smooth communication between the user interface and the model.

The use of machine learning algorithms, particularly the RandomForestClassifier, provided a robust framework for predictions. The model demonstrated good performance, and its deployment in an accessible web application makes it an effective tool for career guidance. Future work could involve integrating more dynamic features, improving accuracy with additional data, and enhancing the user interface for a more personalized experience.