



PERSONALIZED CAREER ADVISOR USING PREDICTIVE CLASSIFICATION AND ADVISORY ENGINE

Mrs. A. Sai Pavani¹, Ravuri Vijaya Priya², Penugonda Yogithasree³, Vipparla Shylee⁴, Sarekukka Sukanya⁵

¹Asst.Prof, Department of CAI, KKR & KSR Institute of technology and sciences, Guntur, Andhra Pradesh, India

^{2,3,4,5}Student, Department of CAI, KKR & KSR Institute of technology and sciences, Guntur, Andhra Pradesh, India

Abstract: Career selection is a complex and critical decision that significantly influences an individual's professional growth and long-term satisfaction; however, many students and job seekers face difficulties due to limited guidance, lack of awareness about available career options, and uncertainty regarding their own abilities and interests. Conventional career counselling systems often rely on generic recommendations and fail to address individual differences. To overcome these limitations, this research proposes a Personalized Career Advisor based on predictive classification and an intelligent advisory engine. The system collects user-specific attributes such as educational background, academic performance, skills, interests, and career preferences, and applies machine learning-based classification algorithms to analyze patterns from historical student and career datasets. Multiple classifiers are utilized to improve prediction accuracy and robustness in identifying suitable career domains. After predicting the most appropriate career path, the advisory engine generates personalized recommendations, including essential skills, relevant courses, certifications, tools, and a structured career roadmap. Furthermore, the system performs resume analysis to extract key information such as skills, qualifications, and experience, which is then matched with job requirements to suggest relevant and reliable job opportunities. When gaps are identified between user profiles and job eligibility criteria, the system provides targeted recommendations for skill development and learning. Overall, the proposed approach integrates career prediction, personalized guidance, and employment recommendations to deliver an effective and intelligent solution for informed career planning and professional development.

KEY WORDS

Career Recommendation System, Machine Learning, Predictive Classification, Advisory Engine, Artificial Intelligence, Resume Analysis, Job Recommendation, Career Guidance System, Skill Gap Analysis, Data Mining, Student Career Prediction, Personalized Recommendation System, Career Path Prediction, Employment Recommendation, Educational Data Analysis.

I. INTRODUCTION

In the modern digital era, career planning has become increasingly complex due to rapid technological advancements and evolving industry demands. Students and job seekers often struggle to choose suitable career paths because of the wide range of options and lack of personalized guidance. Traditional career counselling systems rely on manual methods and generic recommendations, which fail to accurately reflect individual potential or current job market trends. To address these challenges, this project, titled *Personalized Career Advisor using Predictive Classification and an Advisory Engine*, proposes an intelligent, data-driven solution. The system uses machine learning techniques to analyze user data such as education, skills, and interests to predict suitable career domains. It also includes resume analysis to match user profiles with relevant job opportunities, identify skill gaps, and suggest improvements. This integrated approach provides personalized, accurate, and practical career guidance, helping users make better career decisions and improve their employability.

II. PROBLEM STATEMENT

Problem Statement Career decision-making is a critical yet challenging process for students and job seekers, as it directly impacts professional growth, job satisfaction, and long-term employability. Despite the availability of numerous career options, individuals often struggle to select a suitable career path due to limited self awareness, lack of personalized guidance, and insufficient understanding of current industry requirements. Existing career counselling systems largely rely on traditional methods such as manual assessments and generic recommendations, which do not effectively consider individual skills, interests, academic performance, or dynamic labor market trends. As a result, these approaches frequently lead to skill mismatches, inefficient career planning, and reduced employment opportunities. There is a need for an intelligent, data driven career guidance system that can accurately analyze individual profiles, predict suitable career domains, identify skill gaps, and provide personalized

recommendations aligned with real world job requirements. Addressing this problem can significantly improve informed career decision making and enhance employability in a rapidly evolving job market.

III. EXISTING SYSTEM

Existing research on career guidance systems demonstrates the effective use of artificial intelligence and machine learning techniques for predicting suitable career paths and providing generalized recommendations. Several studies focus on career prediction using classification algorithms based on academic performance, skills, and interests, while others emphasize course or skill recommendation systems to support career development. In parallel, some works address resume analysis and job matching using natural language processing techniques. However, most existing solutions treat career prediction, personalized guidance, skill development, and job recommendation as isolated components rather than as a unified system. Many systems lack comprehensive personalization, as they fail to integrate multiple user attributes with dynamic job market requirements. Additionally, limited attention has been given to identifying skill gaps and providing actionable, step-by-step guidance to bridge these gaps in alignment with predicted career paths. The absence of an integrated advisory framework that combines predictive classification, resume analysis, skill gap identification, and employment-oriented recommendations creates a significant gap in current research. Addressing this gap requires a holistic, intelligent career guidance system that not only predicts suitable career domains but also provides personalized advisory support and employment-focused recommendations within a single decision support framework, which forms the core motivation of the proposed study.

- Many systems provide only generalized career recommendations.
- Career prediction and job recommendation are often handled separately.
- Limited personalization based on user skills and interests.
- Most systems do not identify skill gaps for users.
- Lack of an integrated advisory system for career guidance.

IV. PROPOSED SYSTEM

The proposed system is an intelligent Career Advisory and Recommendation System designed to help students and job seekers identify suitable career paths based on their skills, education, and interests. The system integrates machine learning techniques and natural language processing (NLP) to analyze both structured data and unstructured resume data in order to provide accurate and personalized career recommendations.

Unlike traditional career guidance methods that rely on manual counseling or static questionnaires, the proposed system uses data-driven decision making. Users provide information such as educational background, technical skills, interests, and academic performance through an application interface. The system processes this information to understand the user's profile and career preferences.

To predict suitable career domains, the system applies machine learning classification algorithms such as Decision Tree, Random Forest, and Naïve Bayes. These algorithms analyze patterns in the dataset and classify users into appropriate career categories. The use of multiple algorithms helps improve prediction accuracy and reliability. In addition to user input data, the system also performs resume analysis using Natural Language Processing techniques. The uploaded resume is processed through NLP steps such as tokenization, text preprocessing, and TF-IDF feature extraction to identify key skills and experience mentioned in the resume. This allows the system to better understand the user's competencies and match them with relevant career opportunities.

Based on the prediction results and resume analysis, the system generates personalized career guidance, including recommended career domains, required skills, suggested courses, certifications, and a potential career roadmap. This helps users understand the gap between their current skills and the skills required for their desired career path. The proposed system aims to provide a more efficient, scalable, and intelligent career guidance platform by combining machine learning predictions with NLP-based resume analysis, ultimately helping users make informed career decisions.

V. METHODOLOGY

The methodology of the proposed Personalized Career Advisor follows a structured, data-driven approach that integrates machine learning-based career prediction, advisory support, resume analysis, and job recommendation. The process begins with data collection, where user-specific information such as educational background, academic performance, skills, interests, career preferences, and resume details are obtained through a structured input interface.

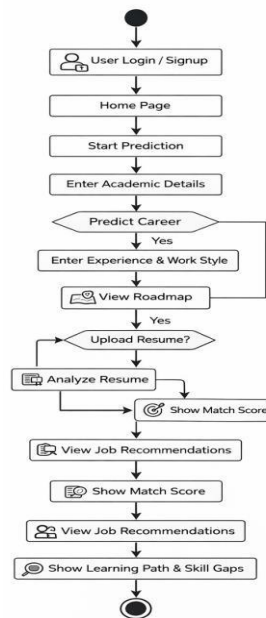


Fig 1: Flow diagram

The collected data is then preprocessed to handle missing values, normalize features, and encode categorical attributes to ensure compatibility with machine learning models. Feature selection techniques are applied to identify the most relevant attributes that influence career prediction.

Next, predictive classification is performed using multiple machine learning algorithms to identify suitable career domains for each user. These models are trained and evaluated on historical student and career-related datasets, and the model with the highest performance is selected for final prediction. Once the career domain is predicted, the system categorizes users as either students or job seekers. For students, an advisory engine generates a personalized career roadmap by recommending essential skills, relevant courses, certifications, and learning resources aligned with the predicted career path.

For job seekers, the system performs resume analysis using text processing techniques to extract skills, qualifications, and experience from resumes. These extracted features are matched with job requirements to recommend relevant and verified job opportunities. A skill gap analysis module evaluates the difference between user competencies and job eligibility criteria. If gaps are identified, the system recommends targeted learning paths and skill enhancement strategies. The final stage delivers personalized outputs, including career guidance, job recommendations, or learning suggestions, thereby providing an integrated and intelligent decision support system for career planning and employability enhancement.

SYSTEM ARCHITECTURE

The system architecture of the proposed Personalized Career Advisory System is designed to collect user information, process the data, and generate suitable career recommendations. The system consists of a user interface, backend server, machine learning module, NLP-based resume analysis module, and a database.

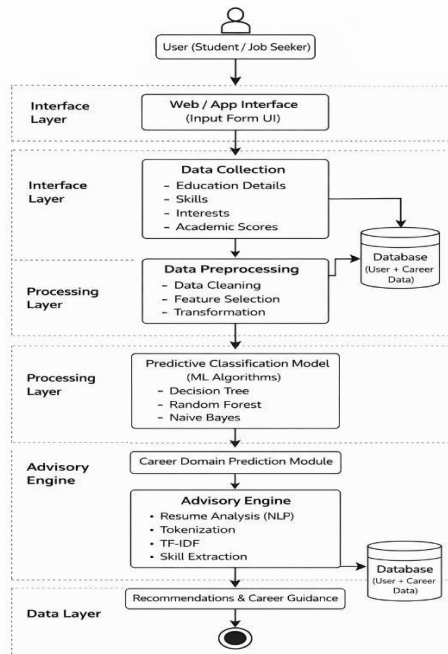


Fig 2: System Architecture diagram

Users interact with the system through a web or application interface, where they provide details such as education, skills, interests, and academic scores, and may also upload their resumes. This information is sent to the backend server, which manages data processing and communication between different modules.

The collected data undergoes preprocessing, including data cleaning and feature selection, before being passed to the machine learning prediction module. Algorithms such as Decision Tree, Random Forest, and Naïve Bayes are used to analyze the user profile and predict suitable career domains. In addition, the system performs resume analysis using Natural Language Processing (NLP) techniques such as tokenization and TF-IDF to extract relevant skills and keywords from the uploaded resume. Based on the prediction results and resume analysis, the system generates personalized career recommendations, including suggested skills, courses, and certifications. All user data and prediction results are stored in a centralized database for efficient management and future use.

SYSTEM WORKFLOW

The workflow of the proposed Personalized Career Advisor system begins with the user accessing the application through a web interface. The user provides basic details such as educational background, academic performance, technical skills, and personal interests. In addition, the user may upload a resume to allow the system to analyze their qualifications and experience. Once the data is submitted, it is transferred to the backend server where preprocessing operations are performed. This step includes data cleaning, formatting, and feature selection to ensure that the input data is suitable for further analysis. The processed data is then forwarded to the predictive classification module.

The classification module applies machine learning algorithms such as Decision Tree, Random Forest, and Naïve Bayes to analyze the user profile and predict the most suitable career domain. At the same time, if a resume is provided, the system performs resume analysis using Natural Language Processing techniques. The resume text is processed to extract relevant skills, keywords, and experience information. After the analysis phase, the system generates personalized results based on the predicted career domain and extracted skill information. The advisory component of the system provides recommendations such as required skills, relevant courses, certifications, and potential career paths. Finally, the results are displayed to the user through the interface, allowing them to view suggested career options and guidance for skill improvement. All relevant data and results are stored in the system database for future reference and system improvement.

PROPOSED SYSTEM ALGORITHM

Algorithm 1: Proposed System Algorithm

Input: User profile data (education, skills, interests, academic scores) and optional resume.

Output: Personalized career recommendations and skill suggestions.

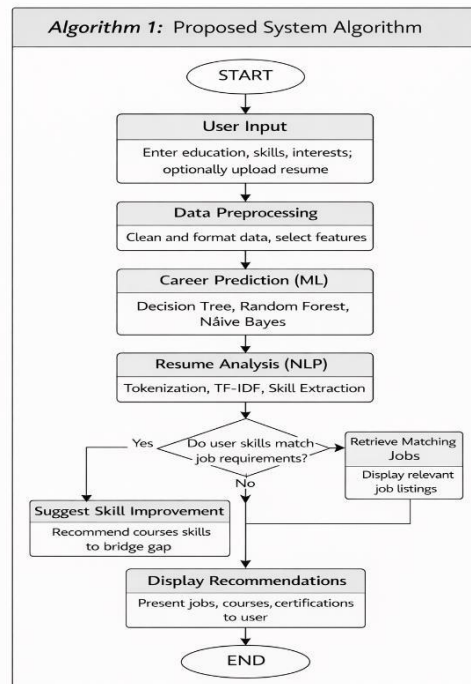


Fig 3: Proposed System Algorithm

1. Start the system.
2. Collect user input through the web interface including education details, skills, interests, and resume (optional).
3. Store user data in the system database.
4. Perform data preprocessing including cleaning, formatting, and feature selection.
5. Apply machine learning algorithms (Decision Tree, Random Forest, Naïve Bayes) to analyze the user profile.
6. Predict suitable career domain based on the trained classification model.
7. If resume is uploaded, perform NLP-based resume analysis using tokenization and TF-IDF to extract relevant skills.
8. Compare extracted skills with job requirements stored in the database.
9. Identify skill gaps and determine whether the user meets the job requirements.
10. Generate recommendations including career domains, required skills, courses, and certifications.
11. Display recommendations to the user through the application interface.
12. End the process.

The proposed Career Advisor system is designed to assist users in identifying suitable career paths based on their educational background, technical skills, and interests. The system collects user input and optionally analyzes an uploaded resume to extract important skills and experience. The collected information is then processed through data preprocessing techniques to remove irrelevant information and prepare the data for further analysis. The system utilizes Natural Language Processing techniques to analyze resume text and extract relevant keywords such as skills, technologies, and professional experience. These textual features are converted into numerical representations to enable efficient analysis and comparison with job requirement profiles. The mathematical formulation used for text feature extraction and similarity measurement in the proposed system is described using the following formulas.

Term Frequency (TF)

$$TF(t, d) = \frac{f_{t,d}}{\sum k_{k,d}}$$

Where

- t represents a term or keyword
- d represents the document (resume)
- $f_{t,d}$ represents the frequency of term t in document d

Term Frequency measures how frequently a specific keyword appears in the resume document.

Inverse Document Frequency (IDF)

TF-IDF Weight

$$TFIDF(t, d) = TF(t, d) \times IDF(t)$$

TF-IDF combines term frequency and inverse document frequency to generate a numerical representation of textual data. These feature vectors help the system identify important skills and keywords present in the user's resume.

Cosine Similarity

$$Similarity(A, B) = \frac{A \cdot B}{\|A\| \|B\|}$$

Where

- A represents the TF-IDF vector of the user resume
- B represents the TF-IDF vector of job requirement data

Cosine similarity measures the similarity between the user's skill profile and job requirement vectors, enabling the system to identify suitable career opportunities.

Skill Gap Identification

To determine missing skills required for a particular career path, the system calculates the difference between required job skills and the user's current skills.

$$\text{Skill Gap} = S_{\text{required}} - S_{\text{user}}$$

Where

- S_{required} represents the set of skills required for a job role
- S_{user} represents the set of skills present in the user's resume or profile

If the skill gap value is greater than zero, the system recommends relevant courses, certifications, or training programs to help the user improve their skill set.

VI. IMPLEMENTATION

The implementation of the proposed Career Advisor system is designed using a modular architecture to ensure scalability, efficiency, and easy integration of machine learning and natural language processing components. The system allows users to provide information such as educational background, technical skills, career interests, and optionally upload their resume for deeper analysis. Once the user submits the information, the system performs data preprocessing to clean and organize the input data. Text-based resume data is processed using Natural Language Processing techniques such as tokenization and keyword extraction in order to identify relevant skills, technologies, and experience mentioned in the resume.

After preprocessing, the extracted features are converted into numerical representations using TF-IDF vectorization. These feature vectors are then provided as input to machine learning models for career prediction and job recommendation. The system compares the extracted user skills with predefined job requirement profiles and determines the most suitable career options. If the user's skill set closely matches job requirements, relevant job roles are recommended. Otherwise, the system identifies missing skills and suggests suitable courses or certifications to help the user bridge the skill gap and improve their career prospects.

Front-End Implementation

The front-end of the Career Advisor system is developed as a web-based interface that enables users to interact with the system easily. The interface allows users to enter personal details such as education, skills, and career interests through a structured input form. Additionally, users can upload their resume for automated analysis. The front-end is designed to be simple and user-friendly so that both technical and non-technical users can easily navigate the platform. It displays career recommendations, job matches, and suggested skill improvement courses in a clear and structured format. The interface is implemented using modern web technologies such as HTML, CSS, and JavaScript, ensuring responsive design and smooth interaction.

Backend Implementation

The backend component of the system is responsible for handling data processing, machine learning analysis, and recommendation generation. After receiving user input from the front-end interface, the backend performs preprocessing and feature extraction using Natural Language Processing techniques. The processed data is then passed to machine learning models including **Decision Tree, Random Forest, and Naïve Bayes** for career prediction and job recommendation. These models analyze patterns within the dataset to determine the most suitable career paths based on the user's skills, interests, and qualifications. The backend also manages resume analysis by applying TF-IDF vectorization to convert resume text into numerical feature vectors. Cosine similarity is used to compare the extracted resume features with job requirement profiles in order to determine the level of skill matching. The backend system communicates with the database to store user information, resume data, and job profiles. It then generates personalized recommendations and sends the results back to the front-end interface for display to the user.

Dataset Description

The dataset used in the proposed Career Advisor system consists of information related to different job roles, required skills, and career categories. The dataset includes attributes such as job title, required technical skills, educational qualifications, and domain specialization. This data is used to train the machine learning models so that the system can learn the relationships between user skills and suitable career paths. During the training process, the dataset is divided into training and testing subsets to evaluate the prediction performance of the models. The dataset enables the system to analyze user profiles effectively and generate accurate career recommendations based on skill matching and career domain analysis.

Machine Learning Models Evaluated

To determine the most suitable algorithm for career prediction, multiple machine learning models were implemented and tested using the processed dataset. The performance of the models was evaluated to identify the most effective algorithm for predicting suitable career paths based on user skills and resume features. The models evaluated include Decision Tree, Naïve Bayes, and Random Forest.

Decision Tree

Decision Tree is a supervised learning algorithm used for classification that splits the dataset into smaller subsets based on feature values. It selects the best attribute for splitting by calculating entropy and information gain.

Entropy Formula

$$Entropy(S) = - \sum_{i=1}^n p_i \log_2(p_i)$$

Information Gain

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

Where

S = Dataset

A = Attribute used for splitting

p_i = Probability of class i

Naïve Bayes

Naïve Bayes is a probabilistic classification algorithm based on Bayes' Theorem. It predicts the probability of a class based on the input features while assuming independence among features.

Bayes Theorem

$$P(C | X) = \frac{P(X | C)P(C)}{P(X)}$$

Where

$P(C|X)$ = Posterior probability

$P(X|C)$ = Likelihood

$P(C)$ = Prior probability

$P(X)$ = Evidence

The model selects the class with the highest probability as the predicted output

Random Forest

Random Forest is an ensemble learning algorithm that combines multiple decision trees to improve prediction accuracy and reduce overfitting.

Prediction Formula

$$y = \frac{1}{T} \sum_{t=1}^T h_t(x)$$

Where,

T = Number of trees

$h_t(x)$ = Prediction from each decision tree

y = Final predicted output

The final classification result is obtained through majority voting of all decision trees.

Model Performance Evaluation

To evaluate the performance of the implemented machine learning models, the accuracy metric is used. Accuracy measures the proportion of correctly predicted instances out of the total number of predictions made by the model.

Accuracy Formula

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

Where:

- **TP (True Positive)** – Correctly predicted positive instances
- **TN (True Negative)** – Correctly predicted negative instances
- **FP (False Positive)** – Incorrectly predicted positive instances
- **FN (False Negative)** – Incorrectly predicted negative instances

Accuracy helps determine how effectively the model predicts suitable career paths based on the user's skills, education, and resume features. Higher accuracy indicates better model performance in identifying correct career recommendations.

VII. RESULTS

The performance of the proposed Career Advisor system was evaluated using multiple machine learning algorithms including Decision Tree, Naïve Bayes, and Random Forest. The models were trained using the processed dataset obtained after resume analysis and feature extraction. The dataset was divided into training and testing sets to evaluate the prediction performance of each algorithm. The models analyze the user’s skills, education, and resume information to recommend suitable career paths and identify potential skill gaps. The experimental evaluation shows that all three machine learning models were able to classify user profiles into appropriate career categories with reasonable accuracy. The Decision Tree model achieved an accuracy of 84%, providing interpretable classification results based on skill attributes. The Naïve Bayes classifier achieved an accuracy of 81%, demonstrating efficient probabilistic classification with faster computation time. Among the evaluated models, the Random Forest algorithm achieved the highest accuracy of 90%, due to its ensemble learning approach which combines multiple decision trees to improve prediction performance and reduce overfitting.

The results indicate that integrating machine learning techniques with resume analysis can effectively support users in identifying suitable career opportunities. The system successfully analyzes the user’s skill profile and recommends relevant career paths while also identifying missing skills required for specific job roles. This helps users improve their skill set and make informed career decisions

OUTPUT SCREENS

The output screens of the *Personalized Career Advisor System* display the results generated based on user inputs such as skills, education, and resume data. These screens present predicted career options, recommended skills, and relevant suggestions in a clear and user-friendly format. The interface is designed to help users easily understand their results and navigate through different features like recommendations and chatbot support. Overall, the output screens ensure effective visualization of the system’s functionality and enhance user experience.

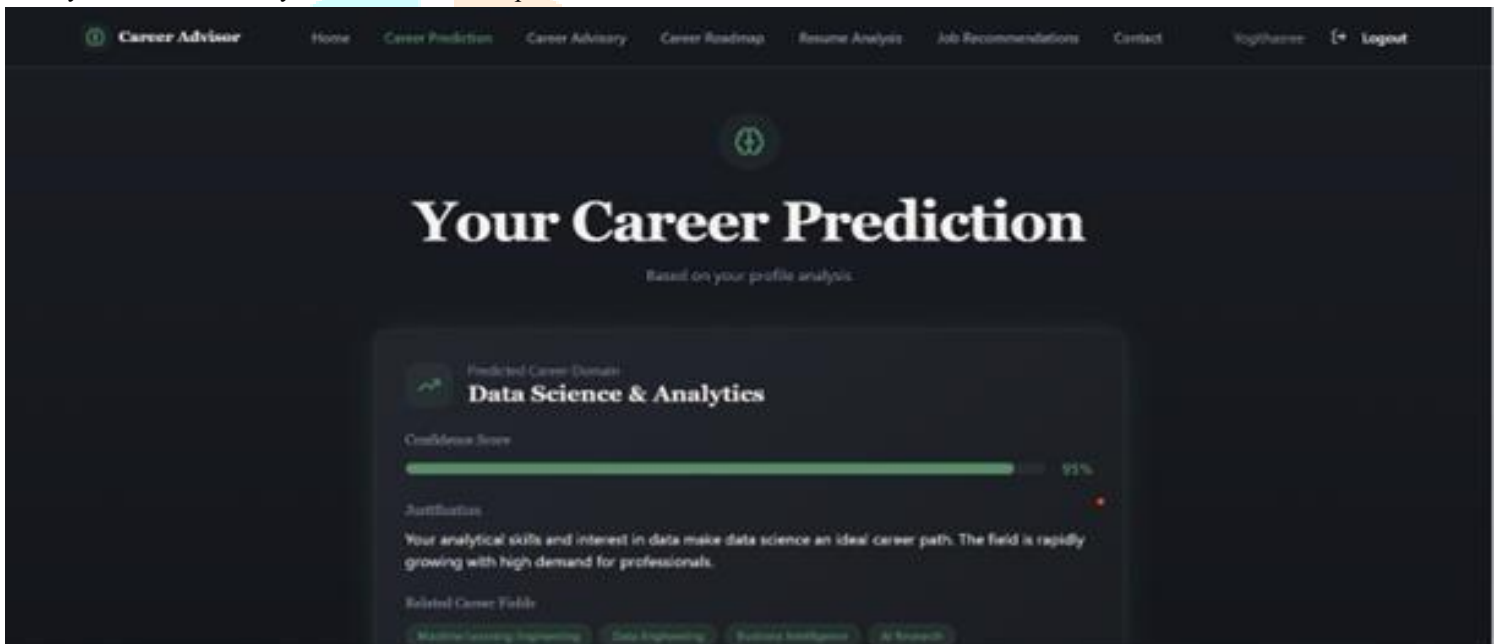


Fig 4: Career Predicted Page



Fig 5: Resume Analysis Page

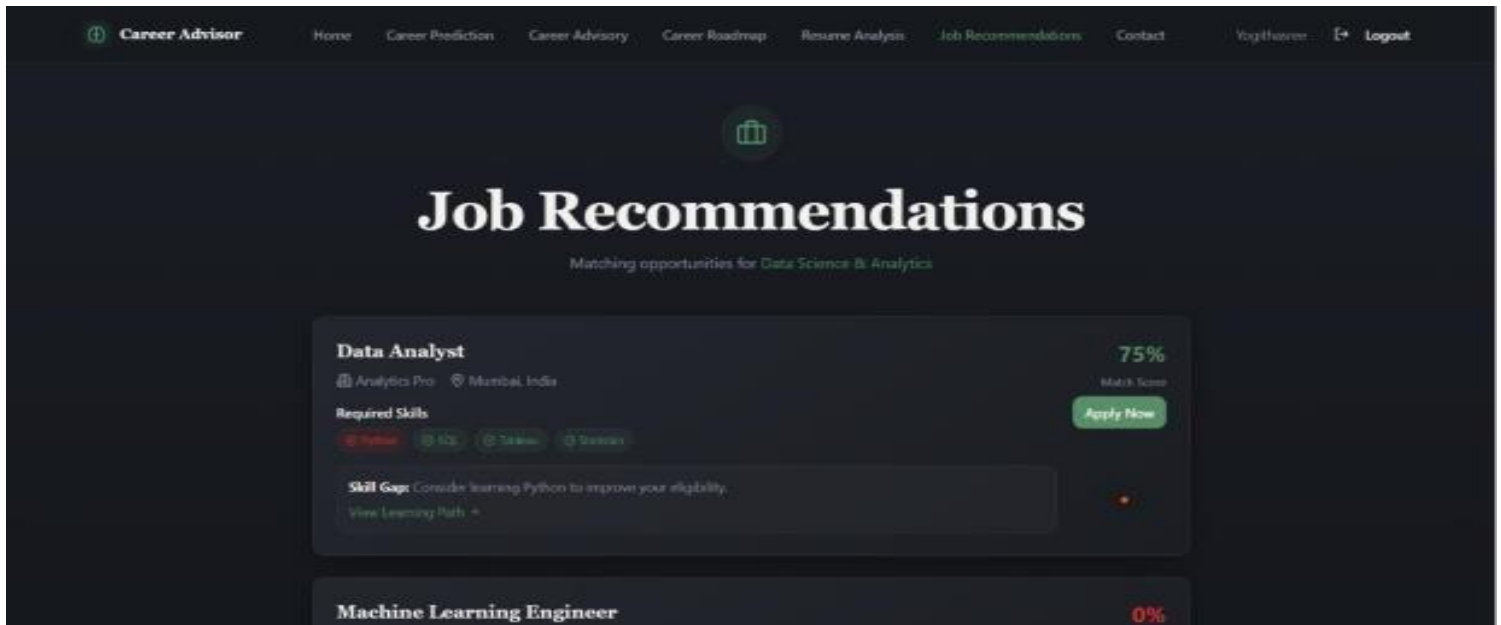


Fig 6: Job Recommendation Page

VIII. MODEL PERFORMANCE EVALUATION

The performance of the proposed Career Advisor system was evaluated using multiple machine learning algorithms including Decision Tree, Naïve Bayes, and Random Forest. These models were trained using the processed dataset obtained after resume analysis and feature extraction. The dataset was divided into training and testing subsets in order to evaluate how well each model performs on unseen data. Performance evaluation is essential to determine the effectiveness of the models in predicting suitable career paths based on the user's skill profile.

Several evaluation metrics were used to measure the performance of the implemented models. These metrics include Accuracy, Precision, Recall, and F1-Score, which provide insights into how well the model correctly classifies career categories. Accuracy measures the proportion of correct predictions made by the model, while precision evaluates the correctness of positive predictions. Recall measures the ability of the model to correctly identify relevant career categories, and the F1-score provides a balanced measure between precision and recall.

Model	Accuracy	Precision	Recall	F1-Score
Decision Tree	84%	0.83	0.82	0.82
Naïve Bayes	81%	0.80	0.79	0.79
Random Forest	90%	0.89	0.88	0.88

From the performance comparison, it can be observed that the Random Forest model achieved the highest accuracy among the evaluated algorithms. This is mainly because Random Forest combines multiple decision trees and aggregates their predictions, which improves classification performance and reduces the risk of overfitting. The Decision Tree model also performed reasonably well due to its ability to split the dataset based on important skill attributes, making the classification process interpretable. On the other hand, the Naïve Bayes classifier demonstrated slightly lower accuracy, but it provides faster computation and efficient probabilistic classification. Despite the difference in performance, all three models were able to effectively analyze user skill profiles and classify them into relevant career domains. Based on the overall evaluation results, the Random Forest model was selected as the final model for generating career recommendations in the proposed system.

The results demonstrate that integrating machine learning techniques with resume analysis and skill extraction can effectively support users in identifying suitable career paths. The system not only predicts potential career opportunities but also helps users understand the gap between their current skills and the skills required for specific job roles, enabling them to improve their qualifications and make informed career decisions.

TECHNOLOGY STACK

Programming Language: Python

Machine Learning Framework: Scikit-learn **Data Processing Libraries:** Pandas, NumPy **Text Feature Extraction:** TF-IDF Vectorization

Machine Learning Algorithms: Decision Tree, Naïve Bayes, Random Forest

NLP Techniques: Tokenization, Stop-word Removal, Lemmatization, Text Normalization

Frontend Development: React.js, HTML, CSS, JavaScript

Backend Framework / Services: Supabase REST API

Database Management System: PostgreSQL

Development Environment: Visual Studio Code/Jupyter Notebook

Version Control: Git & GitHub

IX. CONCLUSION

The proposed Career Advisor system was developed to assist users in identifying suitable career paths based on their skills, education, and resume information. The system integrates machine learning techniques and natural language processing methods to analyze user profiles and generate personalized career recommendations. By applying text preprocessing and feature extraction techniques such as TF-IDF vectorization, the system converts resume data into meaningful numerical representations that can be effectively analyzed.

The system evaluates user skill sets and compares them with career requirements to recommend relevant job roles and identify potential skill gaps. This helps users understand which skills they already possess and which areas need improvement in order to achieve their desired career goals. The system also provides guidance by suggesting career options that align with the user's abilities and interests. Overall, the developed system demonstrates how intelligent data analysis can support users in making informed career decisions. By combining resume analysis with automated career recommendations, the system provides a practical solution for students and job seekers seeking career guidance. In the future, the system can be further improved by integrating larger datasets, real-time job market information, and additional recommendation techniques to enhance prediction accuracy and provide more personalized career insights.

X. ACKNOWLEDGEMENT

We would like to express our sincere gratitude to our guide **Mrs. A. Sai Pavani**, Assistant Professor, Department of CAI, KKR & KSR Institute of Technology and Sciences, for her valuable guidance, encouragement, and continuous support throughout the development of this project. We also extend our thanks to the faculty members of the CAI Department for providing the necessary resources and academic support. Our appreciation goes to our institution for offering a conducive learning environment to carry out this work. Finally, we thank our friends and family for their constant motivation and encouragement during the completion of this project.

XI. REFERENCES

1. **S. Kumar and R. Garg**, "An Intelligent Career Guidance System Using Machine Learning," *International Journal of Advanced Research in Computer Science*, vol. 9, no. 2, pp. 45–50, 2018.
2. **S. B. Kotsiantis**, "Supervised Machine Learning: A Review of Classification Techniques," *Informatica*, vol. 31, no. 3, pp. 249–268, 2007.
3. **H. Witten, E. Frank, M. A. Hall, and C.J. Pal**, *Data Mining: Practical Machine Learning Tools and Techniques*, 4th ed., Morgan Kaufmann, 2016.
4. **Faliagka, K. Ramantas, A. Tsakalidis, and G. Tzimas**, "An Application of Machine Learning Algorithms to an Online Recruitment System," *IEEE International Conference on Tools with Artificial Intelligence*, pp. 215–220, 2012.
5. **R. Burke**, "Hybrid Recommender Systems: Survey and Experiments," *User Modeling and User-Adapted Interaction*, vol. 12, no. 4, pp. 331–370, 2002.