



Automated Career Counselling Using Random Forest Algorithm

S. Nafeesa Thehseen
CSE Department

KLMCEW
Kadapa, India.

K.Uma Maheswari
CSE Department

KLMCEW
Kadapa, India.

Abstract— Choosing the right career is a pivotal milestone in a student's life and often shapes their future professional trajectory. Traditional career counselling, though valuable, can be influenced by subjective opinions and limited by human bias, accessibility, or lack of data-driven insights. To address this challenge, this paper introduces an Automated Career Counselling System that leverages machine learning to provide personalized career recommendations. By analyzing a student's academic background, practical experience, and soft skills, the system predicts optimal career paths using a trained Random Forest Classifier. Key inputs include GPA, internships, completed projects, communication skills, and other personal competencies.

The model was built using Python, with the scikit-learn library for machine learning and Flask for developing a lightweight and responsive web application. The user interface is designed to be intuitive and accessible, allowing students to interact easily with the system and receive actionable insights. The training data consists of a labelled dataset that maps various student profiles to corresponding suitable careers. After training and evaluation, the model demonstrated promising accuracy, confirming its reliability in assisting career decisions.

Unlike conventional guidance methods, this system offers a consistent and objective evaluation of students' potential career options. By removing human subjectivity, it ensures that each recommendation is solely based on data and measurable factors, making it suitable for large-scale deployment across educational institutions. Counsellors and administrators can use this system to support a growing student base with consistent quality and efficiency.

Looking ahead, the system has strong potential for further enhancement. Integrating real-time labour market data and trends could allow the platform to recommend careers based not only on student profiles but also on current industry demands. Additionally, incorporating automated resume analysis and natural language processing for evaluating written inputs or essays could significantly improve the personalization of suggestions.

Keywords—GPA, Support Vector Machines (SVM), decision tree classifiers, communication skills, personal competencies.

I. INTRODUCTION

In the evolving landscape of education and employment, students often face uncertainty in choosing the right career path. Traditional career counselling methods, which rely on personal judgment or static aptitude tests, frequently fail to capture the complexities of an individual's skill set, academic history, and professional interests. The increasing availability of student data and the rise of intelligent systems have made it possible to develop smarter, automated tools that can assist in making informed career decisions.

The rapid advancements in machine learning have opened up new avenues in educational technology. Among these, career prediction systems that analyse student profiles and suggest suitable professions have gained substantial attention. These systems utilize data-driven algorithms that learn patterns from historical data to provide accurate, unbiased predictions. By doing so, they empower students to understand their strengths and align them with career opportunities that are most relevant to their skill sets.

This paper introduces an intelligent career counselling system that employs the Random Forest Algorithm, a powerful ensemble learning method known for its high accuracy and robustness in classification tasks. The model takes various parameters into account, such as GPA, number of internships, number of projects, coding skill level, and communication ability. These factors are known to influence job performance and are thus used to predict the most probable career field for a student. What sets this system apart from traditional counselling is its scalability and accessibility. Students can access career suggestions through a simple web interface, where they enter their details and receive predictions in real time. The system is developed using Python, leveraging popular libraries like scikit-learn for machine learning and Flask for web deployment. This makes the tool not only efficient but also portable and user-friendly.

II. LITERATURE SURVEY

The concept of career prediction using data-driven methods has garnered increased attention over the past decade due to the rise of artificial intelligence and machine learning technologies. Numerous research efforts have focused on using classification algorithms to analyse educational data and recommend optimal career paths for students based on their profiles.

One of the earliest influential works in this space was the use of decision tree classifiers, which offered a clear visual structure for decision-making processes. These were favoured for their interpretability and ability to handle categorical and numerical data. However, decision trees are prone to overfitting, especially with smaller datasets.

Support Vector Machines (SVMs) have also been explored for career prediction due to their effectiveness in high-dimensional spaces. SVMs provide a good margin of separation between classes but require careful tuning of hyperparameters and do not inherently provide probabilistic output, which limits their practical use in recommendation systems.

Random Forests mitigate the shortcomings of individual decision trees by constructing a multitude of trees during training and outputting the class that is the mode of the classes of the individual trees. This approach improves accuracy, reduces overfitting, and works well even with datasets containing noisy or missing data.

A. EXISTING SYSTEM:

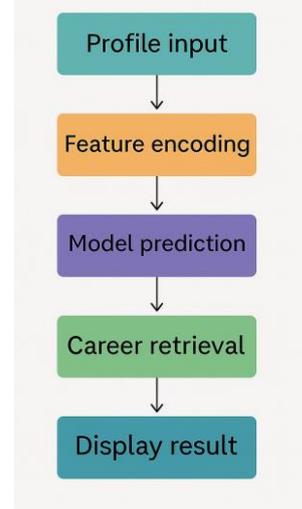
In many institutions, career guidance sessions are conducted in batches, offering little opportunity for personalized advice. Furthermore, the tools and aptitude tests used are often outdated and not aligned with industry trends. This creates a mismatch between the capabilities of the student and the demands of the job market. The lack of data-centric systems also means historical student success data isn't leveraged to guide new batches. As a result, students may end up choosing career paths that don't reflect their potential or interests, leading to dissatisfaction and underperformance in the long term.

B. PROPOSED SYSTEM:

Unlike traditional counselling sessions that require prior scheduling and are prone to variability in guidance quality, this system is standardized and based on data analytics. It empowers students to receive insights at any time during their academic journey and helps them align their projects, internships, and certifications with a targeted career path. Moreover, institutions can also use anonymized predictions to identify trends and update curriculum or placement support services accordingly.

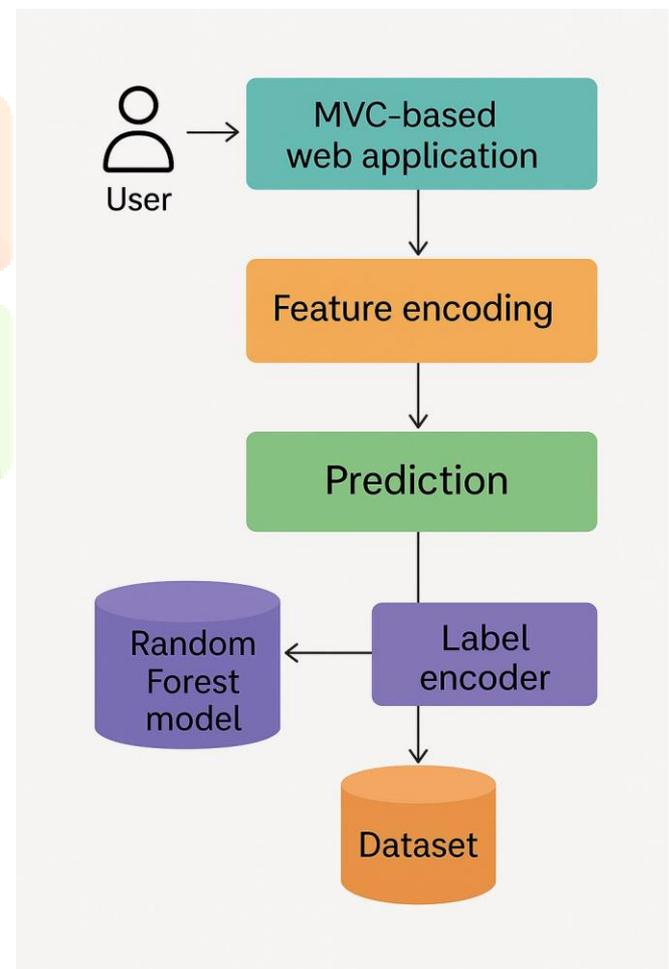
The proposed system is designed to assist students in identifying the most suitable career paths based on key academic and experiential factors. At its core, the system employs a Random Forest classifier, a robust and widely used machine learning algorithm known for its accuracy and resistance to overfitting. The classifier processes a combination of inputs provided by students, such as GPA, internship experience, project involvement, and skill ratings, to generate a list of career recommendations tailored to the individual's strengths and interests.

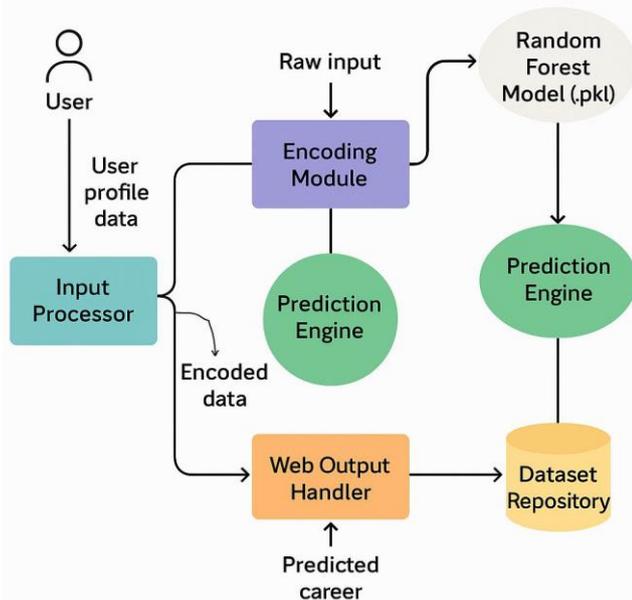
Flow chart for proposed system:



To ensure ease of access and wide usability, the system is deployed through a user-friendly web interface. This interface allows students to seamlessly input their data and receive real-time predictions without needing technical expertise.

III. ARCHITECTURE:





IV. SYSTEM DESIGN

A. Input Design:

The input is collected through dropdowns, number fields, and text boxes, ensuring that user interaction is seamless and error-free. The system requires the following inputs:

GPA – Entered as a decimal value (e.g., 3.5)

Number of Internships – Integer input reflecting work exposure

Number of Projects – Integer input indicating technical engagement

Coding Skills – Selected from a dropdown (e.g., Beginner, Intermediate, Advanced)

Communication Skills – Dropdown input (e.g., Poor, Average, Good, Excellent)

Leadership Experience – Dropdown input (e.g., None, Some, Significant)

Extracurricular Activities – Integer input representing involvement beyond academics

Preferred Work Environment – Dropdown input (e.g., Corporate, Research, Startup)

Once validated, the inputs are pre-processed and encoded using Label Encoder before being fed to the Random Forest classifier, ensuring compatibility with the model's training format. The system design ensures a balance between **ease of use**, **accuracy**, and **robustness**, forming the foundation for reliable career prediction.

B. OUTPUT:

After processing the input data through the Random Forest model, the system generates a career prediction, which is then dynamically rendered on the result page of the web interface.

Key Features of the Output Design:

Prediction Display:

The predicted career (e.g., Data Scientist, AI Specialist, Software Engineer) is shown prominently in a styled output box using HTML and CSS.

Dynamic Rendering:

Flask with Jinja2 templating is used to inject the model's output into the predict.html page, ensuring real-time and seamless result delivery.

User Guidance:

A "Try Again" button is provided, allowing users to go back to the input form and submit different profiles for analysis, enhancing interactivity.

Styling and Layout:

- Results are shown in a centered container with modern CSS design including borders, shadows, and padding.
- The predicted career is styled using a larger font size and highlighted color (#007bff) to draw immediate attention.

Responsiveness:

- Output pages are fully responsive, adapting to various screen sizes including desktops, laptops, tablets, and mobile devices using Bootstrap.

Example Output Behavior:

- If a career is predicted:
- "Your predicted career is: AI Specialist".

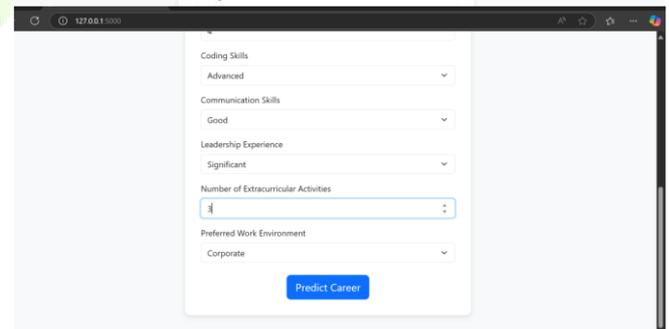
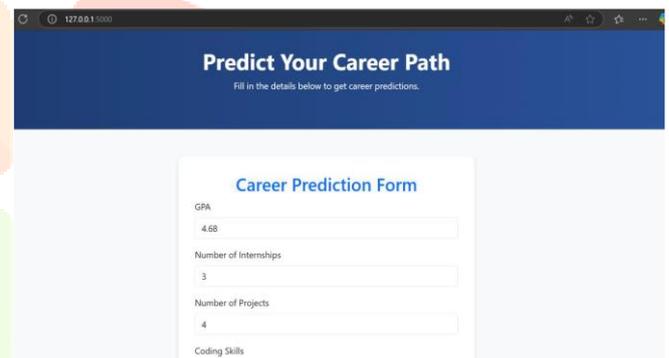
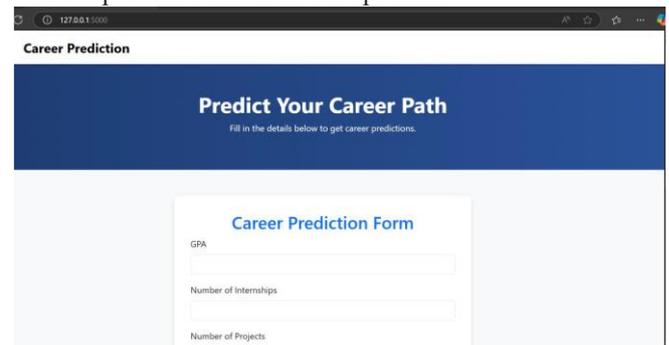


fig: Entering Inputs

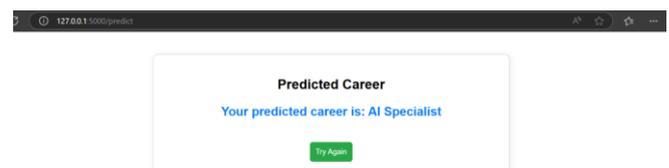


Fig: predicted output

INPUT: STUDENT SKILLS RATINGS OUT OF 5					CAREER OPTED
A	INTERN SHIPS	SKILLS	PROJECTS	CODING SKILLS	
5	2	GOOD	5	ADVANCED	DATA SCIENTIST
5	1	AVERAGE	3	EXCELLENT	DEVELOPER
7	4	GOOD	6	GOOD	CYBER SECURITY ANALYST
	3	GOOD	5	GOOD	WEB DEVELOPER
	3	EXCELLENT	3	GOOD	NETWORK ENGINEER
8	2	GOOD	2	AVERAGE	AI SPECIALIST

V. FUTURE ENHANCEMENT

While the current system effectively predicts suitable career paths based on user input using a Random Forest Classifier, there are several opportunities for future enhancement to increase its accuracy, usability, and scalability. One major area of improvement is dataset enrichment. Incorporating more diverse and detailed input features such as academic transcripts, standardized test scores, technical certifications, aptitude test results, and even personality assessments could significantly refine the prediction capability. Additionally, user behaviour tracking, such as learning history or internship feedback, can be factored into future models to personalize recommendations.

Another key upgrade could be the inclusion of a dashboard or analytics module where users can track their skill development, explore multiple career options, and receive continuous guidance over time. Finally, deploying the system on cloud infrastructure can ensure better scalability, performance, and data security, making it feasible for institutional or large-scale academic use. These enhancements can transform the project from a prediction tool into a comprehensive, intelligent career counselling platform.

V. CONCLUSION

The project “Automated Career Counselling Using Random Forest Algorithm” successfully demonstrates how machine learning can be leveraged to provide intelligent, data-driven guidance for students and job seekers in choosing their ideal career path. By analysing key attributes such as academic performance, technical skills, leadership experience, and work environment preferences, the system offers a personalized career recommendation that reduces the dependency on traditional, manual counselling methods. The use of the Random Forest algorithm ensures high accuracy, robustness, and the ability to handle complex, non-linear data relationships.

In conclusion, this project not only bridges the gap between technology and career guidance but also sets a precedent for integrating AI in educational and professional development systems.

VI. REFERENCES

Scikit-learn Documentation – Provided comprehensive guidance for implementing the Random Forest algorithm, preprocessing techniques, and model evaluation methods.
Available at: <https://scikit-learn.org>

Python Official Documentation – Served as a reference for syntax, functions, and modules used in building both backend logic and data handling.
Available at: <https://docs.python.org>

Flask Framework Documentation – Used for building the web application architecture and managing routes, forms, and data handling with ease.
Available at: <https://flask.palletsprojects.com>

Bootstrap 5 Documentation – Assisted in designing a responsive and modern user interface for the input and output pages of the web application.

Available at: <https://getbootstrap.com>

Kaggle Dataset Repository – Offered inspiration for structuring and refining the career dataset used in training the machine learning model.

Available at: <https://www.kaggle.com>

W3Schools Tutorials – Referred for quick examples and explanations of HTML, CSS, and basic web development practices.

Available at: <https://www.w3schools.com>

GitHub Projects – Explored existing open-source career prediction and ML-based counselling systems to understand architecture and implementation.

Available at: <https://github.com>

Jupyter Notebooks – Used during the model development phase for testing code snippets, preprocessing data, and visualizing accuracy metrics.

Available at: <https://jupyter.org>

Research Papers on Career Guidance Systems – Studied several IEEE and Springer papers for understanding the real-world applications and challenges in AI-driven career counselling.

Sources: IEEE Xplore and SpringerLink academic libraries.



Medium Articles and Blogs – Followed several expert-written blogs on machine learning best practices, model optimization, and deployment strategies.
Available at: <https://medium.com>

