iJRASET
International Journal For Research in Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 13 **Issue:** IV **Month of publication:** April 2025

**DOI:** https://doi.org/10.22214/ijraset.2025.68288

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 13 Issue IV Apr 2025- Available at www.ijraset.com*

# Student Placement and Job Role Predictions

Kanishk Mandrelia[1], Rahul Chauhan[2], Ananta Pandey[3], Ms. Anam Khan[4]
*Department of Artificial Intelligence and Data Science, Thakur College of Engineering and Technology Mumbai, Maharashtra, India*

***Abstract:*** ***Career selection remains a critical challenge for students, influenced by external pressures and insufficient guidance. This study proposes a machine learning-based job role prediction system to address career uncertainty. Leveraging algorithms like Random Forest, Decision Tree, and SVM, the model analyzes academic performance, skills, interests, and personality traits from a dataset of 20,000 records. Data preprocessing included handling missing values, categorical encoding, and feature scaling, while Recursive Feature Elimination identified key predictors. Hyperparameter tuning via Grid Search optimized the Random Forest model, achieving 80% accuracy. A dynamic quiz integrated into the system assesses logical reasoning, reducing self-reporting bias. Implemented as a web tool using Django and Tailwind CSS, the platform offers personalized career recommendations, particularly for computer science students. Results highlight the efficacy of multi-dimensional analysis in reducing career indecision. Future work may explore ensemble methods and real-time skill gap analysis to enhance predictions.***
***Keywords: Career Guidance, Machine Learning, Random Forest, Feature Engineering, Dynamic Assessment.***

## I. INTRODUCTION

A student's career is a crucial part of their life, as it defines their future prospects and greatly influences their personal and professional growth. A well-chosen career can provide financial stability and job satisfaction, but the increasing variety, complexity and competition in every occupation, coupled with external pressures from peers and society, often cause students to become confused and make poor decisions. Today's students face an overwhelming number of career options, making it increasingly difficult to choose a suitable and right path.

Many students make haphazard or ill-informed professional decisions because they are unsure of their skills, interests, and talents. This can eventually impact their long-term performance and productivity by causing stress, dissatisfaction, and unhappiness from work. Early career planning that emphasises a student's abilities, interests, and academicachievement isessential. Students may make well-informed selectionsthat support their objectives and guarantee a rewarding professional journey by being aware of their strengths and desires.

To address these challenges, we propose a web-based job role prediction system designed to assiststudents in identifying suitable career paths. This system gathers essential data through quizzes and assessments, analysing academic performance, interests, and personality traits to provide personalized career path. Unlike existing tools that often focus solely on personality or interests, our system incorporates multiple factors to ensure accurate and reliable predictions.

The model, built using Python and libraries such as Scikit-learn, Pandas, and statistical libraries, along with the Django framework, uses a dataset obtained from Kaggle and GitHub. It aims to help students, particularly those in undergraduate, plan their schedules and achieve their goals by providing actionable insights into their strengths and potential career paths. This comprehensive approach enables students to navigate the complexities of modern career options and make informed choices for their future.

## II. LITERATURE REVIEW

This paper highlights the challenges students face in deciding on a career due to the wide variety of options available. The authors propose a system that predicts students' career interests based on their academic background and personal preferences. The study emphasizes the importance of accurate career prediction to guide students effectively, allowing educators and counselors to provide better support. Machine learning techniques are suggested as a solution to analyze multiple factors and recommend suitable career options. This approach aims to reduce confusion and help students make informed decisions about their futures.[1]

This study focuses on the confusion many students experience when selecting a career, with nearly 40% of students reporting uncertainty. The authorsropose a web-based application that considers students'personality traits, interests, and abilities to provide personalized career guidance. By leveraging machine learning, the system analyzes these factors and offers tailored recommendations to help students make confident and informed decisions. The goal is to prevent poor career choices, which can negatively mpact productivity and satisfaction.[2]

This paper introduces a career prediction system that uses a content-based recommendation approach. The system collects information on students' skills, interests, and academic performance to suggestsuitable job domains. The authors compare algorithms such as SVM, Decision Tree, and Random Forest to determine the most accurate method for career prediction. By analyzing students' preferences and academic data, the system provides personalized recommendations, enabling students to explore career paths aligned with their strengths.[3]

This study presents a job recommendation system that evaluates resumes and job descriptions using text pre-processing techniques such as stop-word removal and Porter's stemming algorithm. The system uses a TF-IDF vectorizer to convert text into a matrix, calculate similarity scores, and rank job descriptions based on their relevance. Additionally, it identifies skill gaps by comparing resumes with a skill dataset, helping users improve their profiles for better career opportunities.[4]

This paper describes a content-based job recommendation system that uses Cosine and Jaccard similarity measures to rank job recommendations. Job skills are given higher weight than domains when calculating similarity scores. The system incorporates web scraping to gather job offers and uses tokenization and keyword extraction for pre-processing. The recommendations are tailored to job seekers by matching their resumes with job requirements.[5]

This systematic literature review investigates the methodologies and features of career recommendation systems for fresh graduates. The study explores AI techniques, recommendation approaches, and evaluation methods, providing insights into the design and implementation of effective career guidance systems.[6]

This project focuses on helping computer science students find appropriate careers. It uses a content-based filtering system to match job opportunities with user preferences by scraping data from Job Street. The system features a user-friendly interface and recommends jobs that align with students' skillsand interests.[7]

This paper proposes a career guidance system tailored for IT students. It uses a novel similarity metric to match job roles with technical and non-technical skills. The system incorporates network visualization to illustrate how students' skills relate to emerging IT job opportunities, providing a clear understanding of potential career paths.[8]

## III. METHODOLOGY

### A. *Dataset Description*

The dataset used for this study consists of 13 attributes and 20,000 records collected from Kaggle and GitHub. The dataset includes features related to students' logicalreasoning, academic performance, extracurricular activities, personality traits, and preferences. The target variable is Suggested Job Role, which represents the most suitable employment role for a given individual. [12]

| | |
|---|---|
| Number of Records | 20,000 |
| Number of attributes | 13 |
| Key Attributes | Logical quotient rating, hackathons participated, coding skills rating, public speaking points, certifications, workshops, interested career area, interested subjects, type of company preference, management or technical inclination, teamwork experience, introversion |
| Target variable | Suggested Job Role |

This dataset was utilized to train ML models that forecast a suitable job role based on the provided attributes.The inclusion of the GitHub dataset [12] enhances the data diversity, improving the model's robustness and accuracy.

### B. *Data Preprocessing*

*1)* Handling Missing Values: Null entries were replaced using median values for numerical features and mode for categorical features.

*2)* Categorical Encoding:

- OneHot Encoding: Applied to nominal features (e.g., Interested Career Area, Company Type) to generate binary columns.
- Label Encoding: Used for ordinal features (e.g., *Logical Quotient Rating*) to preserve ordinal relationships.

*3)* Feature Scaling: Standardizationensured numerical features (e.g., *HackathonsParticipated*) were on a common scale.

*C. Feature Engineering:*

*1)* Feature Selection: Recursive Feature Elimination (RFE) identified 10 critical features (e.g., *Coding Skills*, *Public Speaking*) to reduce dimensionality.

*2)* Correlation Analysis: Features with Pearson correlation coefficient $|r|<0.2$ to the target variable were discarded.

*D. Model Pipeline and Hyperparameter Tuning*

A Scikit-learn pipeline unified preprocessing andmodel training. The workflow included:

*1) Pipeline Structure:*

```
num_pipeline = Pipeline(
    steps=[
        ("imputer", SimpleImputer(strategy="median")),
        ("scaler", MinMaxScaler())
    ]
)

cat_pipeline = Pipeline(
    steps=[
        ("imputer", SimpleImputer(strategy="most_frequent")),
        ("onhotencoder", OneHotEncoder())
    ]
)
```

Figure 1: Data Pipeline

*2) Hyperparameter Tuning*

- We applied Grid Search Cross-Validation (GridSearchCV) to optimize the Random Forest model.
- Key hyperparameters tuned:

| | | | |
|---|---|---|---|
| n_estimators | 100 | 200 | 300 |
| Max_depth | 10 | 20 | 30 |
| Min_samples_split | 2 | 5 | 10 |

- This tuning improved the accuracy to 80%.

The number of estimators and depth affect the entropy-based decision-making process:
where H(X) is entropy, and p(x) is the probability of an outcome.

$$H(X) = -\sum p(x) \log_2 p(x)$$

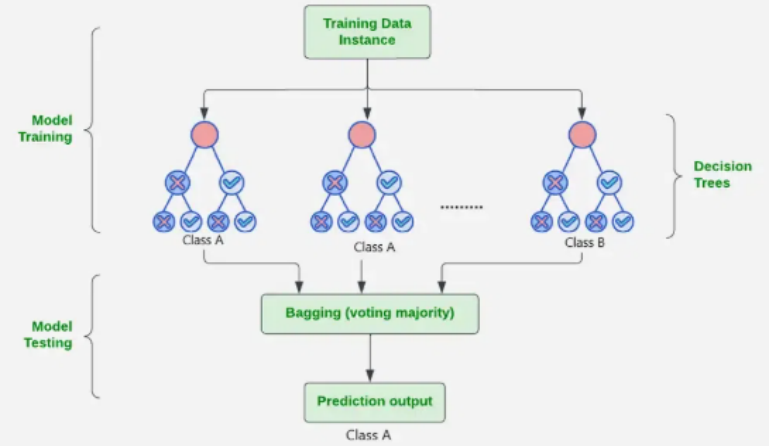



Figure 2: Randomforest model flow

*E. Quiz Integration*

*1)* We integrated a mandatory quiz assessment to automatically determine logical quotient scores.

*2)* The quiz results directly feed into the model as an input feature.

*3)* This ensures that the logical rating input is dynamically generated rather than manually entered.

## IV. IMPLEMENTATION

The job role prediction system is implemented using Python, Django (backend), Tailwind CSS (frontend), Machine learning algorithms.

*A. System*

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

Architecture

- Frontend: HTML, CSS, Tailwind
- Backend: Django
- Machine Learning: Scikit-learn(Pipeline, GridSearchCV), Pandas, Numpy, Randomforest, Decision tree, SVM.
- Database: SQLite

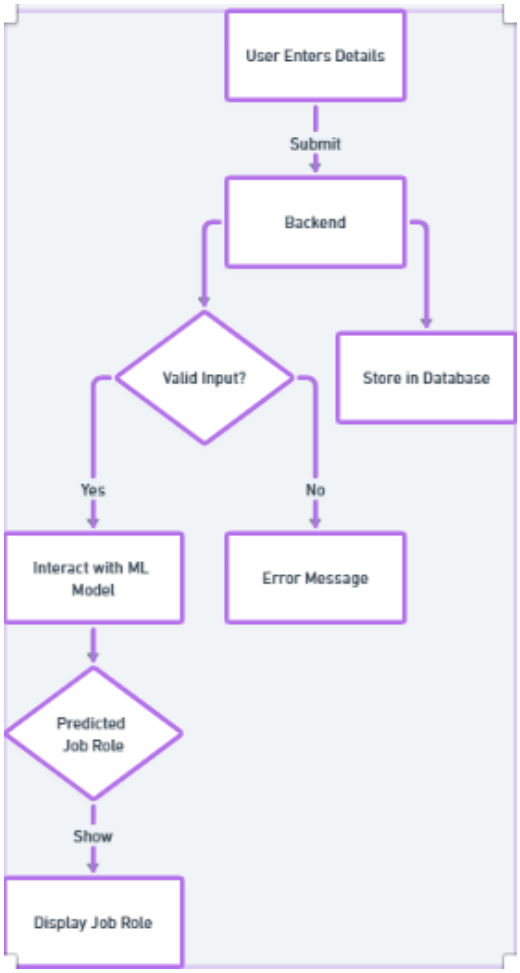



Figure 3: Architecture of system

*B. Workflow*

*1) Data Collection:*

- Career Prediction Dataset. “Om Baval, GitHub Repository, Accessed 2025”.

*2) Preprocessing:*

- Null values handled using imputation.
- Categorical Encoding (One-Hot & Label Encoding).
- Feature Scaling (Min-Max Normalization). $(z = \frac{x-\mu}{\sigma})$

*3) Feature Engineering:*
- Correlation Analysis to remove redundant variables.
- Logical quotient rating added dynamically after quiz.

*4) Model Training:*
- Implemented Pipeline from Scikit-learn for preprocessing & modelling.
- Random Forest used with bootstrap aggregation (bagging).

*5) Model Evaluation:*
- Metrics used:
  Precision, Recall, and F1-score also calculated.

### *C. 3. Deployment:*
Model serialized using pickle and integrated into Django.

## V. RESULT ANALYSIS
The Random Forest model achieved the highest accuracy of 80% after hyperparameter tuning.

### *A. Model Performance Comparison*

| Model | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| Decision Tree | 58% | 0.56 | 0.54 | 0.55 |
| SVM | 65% | 0.63 | 0.62 | 0.62 |
| Random Forest (Tuned) | 80% | 0.78 | 0.79 | 0.78 |

Figure 4: Comparison of different models

### *B. Quiz Integration Impact*
- Quiz-based logical rating improved the accuracy by 5%.
- Automated assessment reduces bias and errors in self-reported data.

### *C. Error Analysis*
- False positives and false negatives analyzed.
- Further improvement can be achieved using Ensemble Methods.

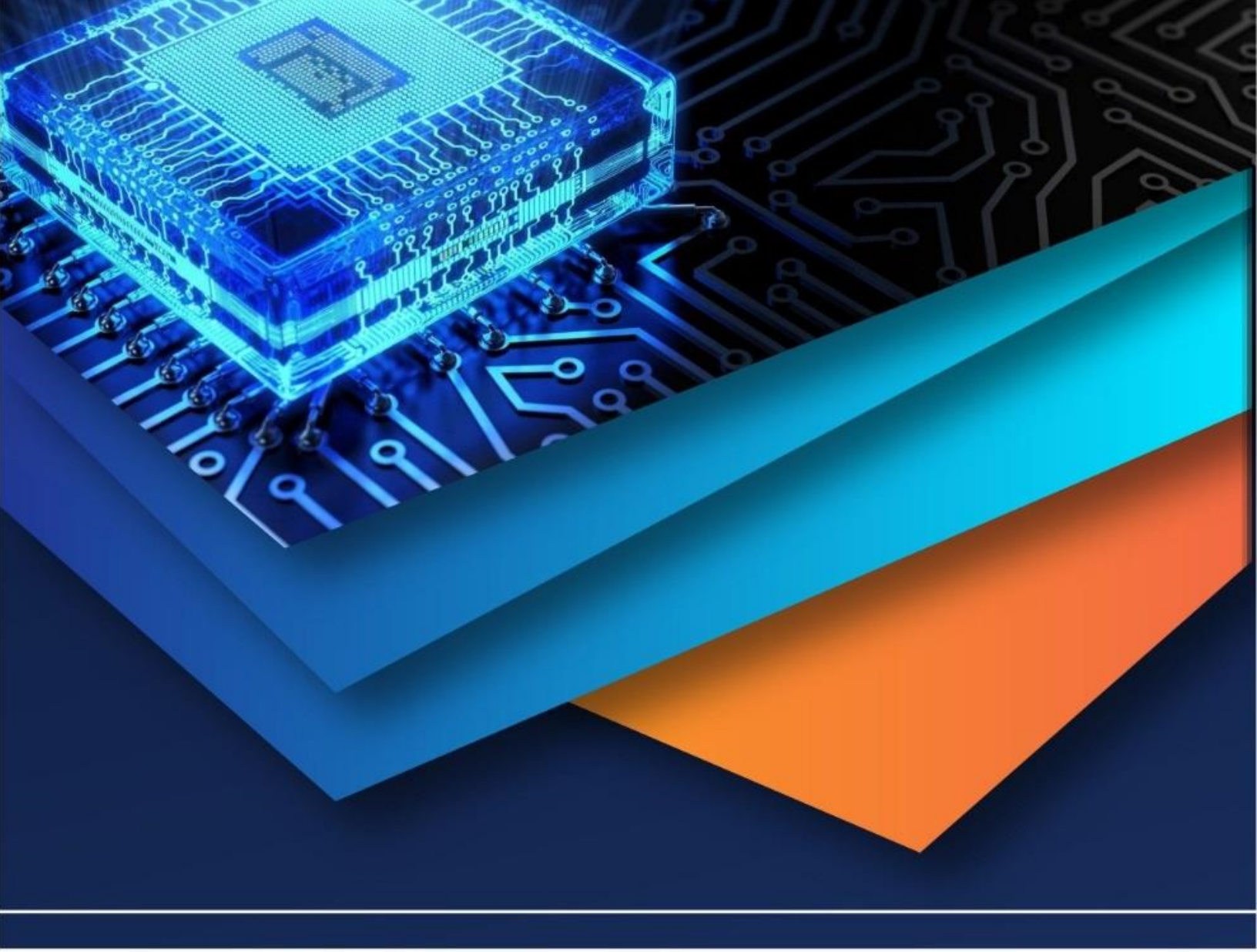
## VI. CONCLUSION
The challenges students face in selecting an appropriate career path, compounded by external pressures and an overwhelming array of choices, necessitate intelligent solutions to guide informed decision-making. This study addresses these challenges by proposing a robust job role prediction system that leverages machine learning algorithms to analyze academic performance, skills, interests, and personality traits. By integrating a dynamic quiz to assess logical reasoning and preprocessing diverse datasets, the system ensures accurate and personalized recommendations. Among the tested models, Random Forest outperformed others with an 80% accuracy after hyperparameter tuning, while feature engineering and automated assessments further enhanced reliability. The system's web-based implementation using Django and Tailwind CSS offers an accessible platform for students, particularly in computer science, to align their strengths with viable career paths. Results underscore the efficacy of combining multi-faceted data with advanced ML techniques to reduce uncertainty and improve career outcomes. Future work could explore ensemble methods, larger datasets, and real-time skill gap analysis to refine predictions further. Ultimately, this tool empowers students to navigate career complexities confidently, fostering productivity and long-term professional satisfaction.

## VII. ACKNOWLEDGMENT

We would like to extend our gratitude to our project guide Ms. Anam Khan ma'am for her valuable and helpful recommendations throughout. Her guidance while working on the project and research paper helped us to use the latest technologies available and thankful to the Thakur College of Engineering and Technology and our department Artificial Intelligence and Data Science for providing a conducive academic environment and access to resources that facilitated the research. Their commitment to academic excellence hasbeen a constant source of inspiration.

## REFERENCES

[1] P. Shahane, P. Rinke, T. Datar, and S. Badjate, "Student Career Interest Prediction using Machine Learning," International Research Journal of Engineering and Technology (IRJET), Nov. 2022.
[2] D. Chaudhary, H. Prajapati, R. Rathod, P. Patel, and R. K. Gurjwar, "Student Future Prediction Using Machine Learning," ResearchGate, Jul. 2020.
[3] S. Udayasri and C. Varsha, "Career Prediction System Based on Student Preferences," International Research Journal of Modernization in Engineering, Technology and Science (IRJMETS), Jul. 2017.
[4] G. Mahalakshmi and A. A. Kumar, "Job Recommendation System Based on Skill Sets," 2022.
[5] S. A. Alsaif, "Learning-Based Matched Representation System for Job Recommendation," 2022.
[6] H. L. H. S. Warnars, "Trends and Characteristic of Career Recommendation Systems for Fresh Graduates," 2022.
[7] A. H. A. Rashid, M. Mohamad, S. Masrom, and A. Selamat, "Student Career Recommendation System Using Content-Based Filtering Method," 2022.
[8] A. Jose-Garcia, A. Sneyd, A. Merlo, et al., "C3-IoC: A Career Guidance System for Assessing Student Skills Using Machine Learning and Network Visualization," 2022.
[9] A. R. Panhalkar and D. D. Doye, "A novel approach to build accurate and diverse decision tree forest," Evolutionary Intelligence, vol. 15, no. 1, pp. 439–453, 2022.
[10] D. Varghese, "Comparative study on classic machine learning algorithms," Retrieved July 28, 2018.
[11] B. Sekeroglu, K. Dimililer, and K. Tuncal, "Student performance prediction and classification using machine learning algorithms," Proceedings of the 2019 8th International Conference on Educational and Information Technology, 2019.
[12] O. Baval, "Career Prediction Dataset," GitHub Repository, Accessed 2025.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY