IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁐ Peer-reviewed & Refereed journal ⁐ Vol. 13, Issue 1, January 2026

DOI: 10.17148/IARJSET.2026.13153

AI-POWERED CAREER GUIDANCE AND RECOMMENDATION SYSTEM

Pramod C[1], Poojashree S[2]

PG Student, The National Institute of Engineering, Mysuru, Visveswaraya Technological University, Belagavi, Karnataka, India[1]

Faculty, The National Institute of Engineering, Mysuru, Visveswaraya Technological University, Belagavi, Karnataka, India[2]

Abstract: Career decision-making is a complex process that significantly influences an individual's academic progression and long-term professional success, yet traditional career guidance approaches rely heavily on manual counseling, static assessments, and generalized recommendations that often fail to account for the diverse abilities, interests, and evolving aspirations of students. To overcome these limitations, this project proposes an AI-Powered Career Guidance and Recommendation System that leverages machine learning techniques to deliver personalized and data-driven career recommendations. The system collects structured student information, including academic performance, technical and soft skills, areas of interest, and personality-related attributes, which is then subjected to comprehensive preprocessing steps such as data cleaning, categorical encoding, numerical normalization, and feature selection to ensure compatibility with machine learning models. Multiple supervised learning algorithms are trained and evaluated using performance metrics including accuracy, precision, recall, and F1-score, with the most effective model selected for deployment. The trained model predicts suitable career domains and generates ranked career recommendations tailored to individual profiles, while model persistence and a modular system architecture support scalability, consistency, and future retraining. Experimental results demonstrate that the proposed system provides accurate and reliable career recommendations, highlighting the effectiveness of machine learning in career guidance applications, reducing dependence on manual counseling, and improving accessibility to consistent, objective, and intelligent career decision-support services.

Keywords: AI-Powered Career Guidance, Career Recommendation System, Machine Learning, Student Profiling, Supervised Learning, Random Forest, Feature Engineering, Clustering Techniques, Data Preprocessing, Model Persistence, Joblib, Streamlit, Decision Support System, Educational Data Mining, Personalized Career Prediction.

I. INTRODUCTION

The development of the **AI-Powered Career Guidance and Recommendation System** follows a structured, data-driven approach designed to deliver personalized and scalable career recommendations using machine learning techniques. The system begins with the collection of structured student data, including academic performance, skills, interests, aptitude scores, and personality traits, which is then subjected to comprehensive preprocessing steps such as duplicate removal, missing value imputation, categorical encoding, numerical normalization, and text cleaning. Feature engineering techniques are applied to transform raw inputs into meaningful numerical representations, including TF-IDF vectorization for skill and interest text, encoding of personality and aptitude attributes, feature scaling, and dimensionality reduction where necessary. The processed dataset is divided into training and testing sets, and multiple supervised machine learning models, including Decision Tree, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Logistic Regression, and Artificial Neural Networks (ANN), are trained and evaluated using multi-class performance metrics such as accuracy, precision, recall, F1-score, and confusion matrices. Clustering techniques are incorporated to group similar career domains and mitigate class imbalance issues commonly found in real-world career datasets. Based on comparative evaluation, the Random Forest model is selected as the best-performing classifier due to its robustness, ability to handle high-dimensional data, and improved generalization. The trained model is persisted using Joblib to enable reuse without retraining, ensuring consistent predictions and reduced computational overhead. A modular system architecture supports efficient data flow, scalability, and maintainability, while a Streamlit-based user interface allows users to input profile data and receive ranked career recommendations with confidence scores. This end-to-end methodology ensures objective, reliable, and interpretable career guidance, effectively bridging the gap between traditional counseling methods and intelligent decision-support systems.

SYSTEM OVERVIEW AND PIPELIN: The AI-Powered Career Guidance and Recommendation System is designed as a modular, data-driven decision-support system that analyzes student profiles and generates personalized career

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⋇ Peer-reviewed & Refereed journal ⋇ Vol. 13, Issue 1, January 2026

DOI: 10.17148/IARJSET.2026.13153

recommendations using machine learning techniques. As illustrated in the system architecture and data flow diagrams, the pipeline begins with the **User Input module**, where students provide structured information such as academic records, skills, interests, aptitude scores, and personality traits through an interactive interface. This input data is first validated to ensure completeness and correctness, after which it is passed to the **Data Preprocessing module**, where data cleaning, handling of missing values, categorical encoding, and normalization of numerical features are performed to make the data suitable for machine learning models. The preprocessed data then flows into the **Feature Engineering module**, where raw inputs are transformed into meaningful numerical representations using techniques such as TF-IDF vectorization for skill and interest text, feature scaling, encoding of personality and aptitude attributes, and dimensionality reduction when required. The engineered features are supplied to the **Machine Learning Model Training module**, where multiple supervised learning algorithms, including Decision Tree, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Logistic Regression, and Artificial Neural Networks (ANN), are trained and evaluated, with clustering techniques optionally applied to reduce class imbalance and group similar career domains. Based on performance evaluation metrics, the best-performing model is selected and stored in the **Model Storage module** using Joblib to ensure consistency, reusability, and reduced computational overhead. Finally, during the **Career Recommendation stage**, the stored model processes new student inputs through the same pipeline and generates ranked career recommendations with confidence scores, which are presented to the student. This end-to-end pipeline ensures a systematic flow of data from input to output, enabling scalable, unbiased, and reliable career guidance that closely integrates user interaction, intelligent analysis, and automated decision-making.

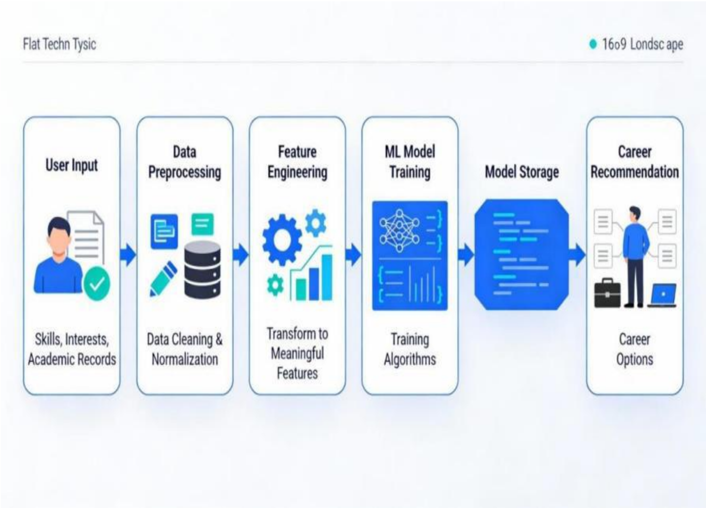


Fig.no.1: Illustrates the overall architecture of the AI-Driven Career Guidance System

FUNCTIONAL REQUIREMENTS: The AI-Powered Career Guidance and Recommendation System must provide a mechanism for students to enter structured profile information such as academic background, skills, interests, aptitude scores, and personality traits through an interactive user interface. The system is required to validate all user inputs to ensure completeness, correctness, and logical consistency before allowing the data to proceed to further processing stages. Invalid or missing inputs must be detected and handled gracefully, with appropriate feedback provided to the user to correct errors.

- Once valid input data is received, the system must preprocess the data to ensure compatibility with machine learning models. This includes cleaning the data, removing duplicate or inconsistent records, handling missing values using suitable imputation techniques, encoding categorical attributes into numerical formats, and normalizing numerical features so that all attributes contribute fairly to the prediction process. Textual inputs such as skills and interests must be cleaned and transformed into numerical representations.
- The system must perform feature engineering to enhance the predictive quality of the dataset. This involves applying vectorization techniques such as TF-IDF to convert skill and interest text into numerical vectors, encoding aptitude scores and personality traits, scaling features, and applying dimensionality reduction techniques when necessary to reduce noise and computational complexity. These engineered features must then be passed consistently to both training and prediction pipelines.
- The system must support the training of multiple supervised machine learning models, including Decision Tree, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Logistic Regression, and Artificial Neural Networks (ANN). It must allow the dataset to be split into training and testing sets and perform

DOI: 10.17148/IARJSET.2026.13153

hyperparameter tuning and cross-validation to avoid overfitting. Clustering techniques must also be supported to group similar career domains and mitigate issues related to class imbalance.

- The trained models must be evaluated using standard multi-class classification metrics. Accuracy must be computed as:

Accuracy = (Number of Correct Predictions) / (Total Number of Predictions)

- Precision, Recall, and F1-Score must be calculated to assess model reliability across different career classes, and confusion matrices must be generated to analyze prediction behavior across closely related careers. Based on these evaluations, the system must select the best-performing model for deployment.

The system must implement model persistence by saving the selected trained model using Joblib. The saved model must be reusable for future predictions without retraining, ensuring consistency and reduced computational overhead. When new student data is provided, the system must load the stored model, apply the same preprocessing and feature engineering steps, and generate ranked career recommendations along with confidence scores. The final recommendations must be displayed clearly and understandably to the user.

NON-FUNCTIONAL REQUIREMENTS: The system must exhibit high performance by generating career recommendations within an acceptable response time, even when handling multiple users simultaneously. Efficient preprocessing, feature extraction, and prediction mechanisms must be implemented to avoid delays and ensure smooth interaction.

- Scalability is a critical requirement, and the system must be capable of supporting a growing number of users, additional career categories, and expanding datasets without significant degradation in performance. The architecture must allow easy integration of new data sources and models as career trends evolve.
- Usability is essential for adoption, and the system must provide a simple, intuitive, and user-friendly interface that can be used by students with minimal technical knowledge. Inputs, outputs, and recommendations must be clearly presented, and navigation through the system must be straightforward.
- The system must be reliable and produce consistent outputs for identical inputs. It must handle unexpected inputs or runtime issues without crashing and maintain stable operation during prolonged usage. Fault tolerance and proper exception handling must be ensured throughout the system.
- Security and privacy must be maintained by ensuring that user data is processed securely and not permanently stored. The system must protect sensitive personal information during runtime and ensure data confidentiality and integrity throughout processing.
- Maintainability is required so that the system can be updated, debugged, and enhanced easily. The modular design must allow retraining of models with new datasets, updating algorithms, and incorporating future enhancements without redesigning the entire system.
- Portability is also required, and the system must be platform-independent, capable of running on Windows, macOS, and Linux environments using standard hardware. No specialized infrastructure should be required for execution.

Finally, the system must support extensibility and interpretability. It should allow future integration of deep learning models, psychometric assessments, job market analytics, and explainable AI techniques. Career recommendations must be interpretable so that students and counselors can understand and trust the decision-making process.

II. METHODOLOGY

The methodology adopted for the AI-Powered Career Guidance and Recommendation System follows a structured, systematic, and data-driven pipeline that transforms raw student information into meaningful and personalized career recommendations using machine learning techniques. The proposed methodology is designed to address the limitations of traditional career counseling systems by replacing subjective judgment with objective, data-based decision-making. It integrates multiple stages, including data collection, preprocessing, feature engineering, model training, evaluation, and recommendation generation, to ensure reliable and consistent outcomes.

The system begins by collecting comprehensive student profile data, such as academic performance, skills, interests, aptitude scores, and personality traits. This data is then carefully preprocessed to handle inconsistencies, missing values, and noise, ensuring that the dataset is suitable for machine learning analysis. Feature engineering techniques are applied

DOI: 10.17148/IARJSET.2026.13153

to convert raw and unstructured inputs into meaningful numerical representations that enhance the learning capability of the models. Multiple supervised machine learning algorithms are trained and evaluated to identify the most effective model for career prediction, with clustering techniques incorporated to handle class imbalance and overlapping career domains.

To ensure efficiency and scalability, the selected machine learning model is persisted using model storage techniques, allowing it to be reused for future predictions without retraining. The trained model is integrated into a modular system architecture that supports easy maintenance and future expansion. Career recommendations are generated by applying the trained model to new student inputs, producing ranked career options along with confidence scores. This end-to-end methodology ensures accurate, unbiased, and scalable career guidance, enabling students to make informed career decisions while providing a robust foundation for future enhancements such as adaptive learning, explainable AI, and real-time labor market integration.

A. *Problem Definition and Requirement Analysis*: The first step of the methodology involves identifying the limitations of traditional career guidance systems, such as subjectivity, lack of personalization, inability to scale, and poor adaptability to changing career trends. Based on this analysis, the problem is defined as the need for an intelligent system capable of analyzing multiple student attributes simultaneously and generating data-driven career recommendations. Functional and non-functional requirements are gathered to determine the scope, constraints, and performance expectations of the system.

B. *Dataset Collection and Understanding*: A structured dataset is compiled containing student profile information and corresponding career labels. The dataset includes attributes such as academic background, technical and soft skills, interest areas, aptitude scores (logical, verbal, numerical), personality traits (based on standardized models), and target career domains. This dataset serves as the foundation for training and evaluating machine learning models. Exploratory data analysis is performed to understand feature distributions, identify missing values, detect inconsistencies, and observe class imbalance across career categories.

C. *Data Preprocessing*: Raw data collected from real-world or compiled sources often contains noise and inconsistencies, which can negatively impact model performance. Therefore, a comprehensive data preprocessing stage is carried out. Duplicate records are removed to avoid biased learning. Missing values are handled using appropriate imputation techniques based on the nature of the attribute. Categorical variables such as education level and personality traits are encoded into numerical representations. Numerical features such as aptitude scores are normalized or scaled to ensure that no single feature dominates the learning process. Textual fields, including skills and interests, are cleaned by removing irrelevant characters and standardizing text format

D. *Feature Engineering*: Feature engineering is performed to convert preprocessed data into meaningful numerical features that improve model learning. Textual skill and interest data are transformed into numerical vectors using TF-IDF vectorization to capture the importance of terms across student profiles. Personality traits and aptitude scores are encoded and scaled to maintain consistency. Feature scaling is applied to ensure uniform contribution of all attributes. Dimensionality reduction techniques are used when required to reduce redundancy and computational complexity while preserving important information.

E. *Dataset Splitting:* The engineered dataset is divided into training and testing sets using an appropriate split ratio (commonly 80:20). The training dataset is used to build and learn model parameters, while the testing dataset is reserved for evaluating the model's generalization capability on unseen data. This separation ensures unbiased performance evaluation.

F. *Model Selection and Training***:** Multiple supervised machine learning algorithms are selected to evaluate their suitability for career prediction. These include Decision Tree, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Logistic Regression, and Artificial Neural Networks (ANN). Each model is trained independently on the training dataset. Hyperparameters are tuned using validation techniques to improve predictive performance and avoid overfitting. Cross-validation is applied to ensure stability and robustness of the models.

G. *Incorporation of Clustering Techniques:* To address issues related to class imbalance and overlapping career domains, clustering techniques are incorporated into the methodology. Similar career categories and student profiles are grouped based on feature similarity. This hybrid approach improves generalization, reduces sparsity in rare career classes, and enhances the stability of supervised classification models.

H. *Model Evaluation:* Trained models are evaluated using standard multi-class classification metrics. Accuracy is calculated as:

I. Accuracy = (Correct Predictions) / (Total Predictions)

DOI: 10.17148/IARJSET.2026.13153

Precision, recall, and F1-score are computed for each career category to analyze model reliability. Confusion matrices are generated to study misclassification patterns, especially among closely related careers. Based on these metrics, models are compared to identify the most suitable algorithm for deployment.

J. *Model Selection and Persistence:* Among the evaluated models, the Random Forest classifier is selected as the best-performing model due to its robustness, ability to handle high-dimensional data, and better generalization. The trained model is saved using Joblib to enable model persistence. This allows the system to reuse the trained model for future predictions without retraining, reducing computational overhead and ensuring consistent results.

K. *Career Recommendation Generation*: When a new student accesses the system, profile data is collected through the user interface and passed through the same preprocessing and feature engineering pipeline used during training. The persisted model processes the engineered features and predicts suitable career domains. The system generates ranked career recommendations along with confidence scores, offering multiple relevant career options rather than a single rigid prediction.

L. *System Integration:* The AI-Powered Career Guidance and Recommendation System integrate multiple functional modules into a unified machine learning pipeline to ensure consistent data flow and reliable predictions. Student input data is collected and passed sequentially through preprocessing, feature engineering, model inference, and recommendation modules. During preprocessing, numerical attributes such as academic scores and aptitude values are normalized to maintain uniform feature scales using ***Min–Max Normalization:*** $x_{norm} = \frac{x-x_{min}}{x_{max}-x_{min}}$ In cases where standardization is required, **Z-score normalization** may be applied:$z = \frac{x-\mu}{\sigma}$ where $\mu$is the mean and $\sigma$is the standard deviation. Missing numerical values are handled using **mean imputation**:$x_{missing} = \frac{1}{n}\sum_{i=1}^{n} x_i$ Categorical variables such as education level or domain preference are encoded using ***One-Hot Encoding***, where a categorical feature $C$with $k$categories is transformed into a binary vector: $C \rightarrow [c_1, c_2, \dots, c_k]$ Textual attributes like skills and interests are converted into numerical form using ***TF-IDF vectorization***, defined as: $TF\text{-}IDF(t, d) = \frac{f(t,d)}{|d|} \times log\left(\frac{N}{DF(t)}\right)$

The resulting feature vector $X$is passed to the trained machine learning model, which has been stored using Joblib for reuse. This integrated approach ensures consistency between training and prediction phases, improves system efficiency, and enables scalable deployment.

M. *User Interface Details:* The user interface acts as the interaction layer between the student and the AI system. It collects structured inputs such as academic performance, skill ratings, interests, aptitude scores, and personality traits using validated forms and sliders. Skill proficiency values entered on a scale of 0–10 is internally normalized before being processed: $x_{skill} = \frac{x}{10}$ Once the student submits the data, the backend model generates predictions and returns ranked career recommendations. Each career recommendation is associated with a **confidence score**, calculated from the predicted probability distribution of the model: Confidence$(c_i) = P(y = c_i \mid X)$ The top-ranked career is selected as: $\hat{c} = arg\ \max_{c_i}\ P(y = c_i \mid X)$ The interface displays the recommended careers along with confidence scores and brief explanations, helping users understand both the outcome and the reasoning behind it.

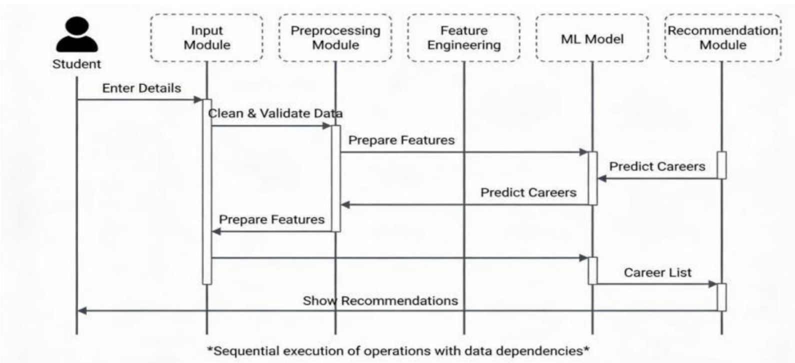


Fig.No.2: Sequence Diagram

N. *Sequence Diagram Explanation:* The sequence diagram represents the chronological interaction between the student and system modules, emphasizing data dependency and execution order. The process begins when the **student** submits profile information through the **Input Module**. The data is forwarded to the **Preprocessing Module**, where cleaning, normalization, and encoding operations are applied. For text-based features, term weights are calculated using TF-IDF, while numerical features are scaled to ensure uniformity.

DOI: 10.17148/IARJSET.2026.13153

The processed data is then passed to the **Feature Engineering Module**, where the final feature vector is constructed: $X = [x_1, x_2, \dots, x_n]$
This vector is supplied to the **Machine Learning Model**, which predicts suitable careers. In the case of a **Random Forest classifier**, the final prediction is obtained using majority voting: $\hat{y} = \mathrm{mode}(h_1(X), h_2(X), \dots, h_T(X))$
where $h_t(X)$represents the prediction from the t^{th}decision tree. For distance-based models like KNN, prediction is based on **Euclidean distance**:$d(X, X_i) = \sqrt{\sum_{j=1}^{n}(x_j - x_{ij})^2}$ The predicted probabilities are then ranked by the **Recommendation Module**, and the final career list is returned to the student. The strict sequential execution ensures correctness, data integrity, and reliable recommendations at every stage.

III. SYSTEM TESTING

System testing is a vital phase of the Software Development Life Cycle (SDLC) that evaluates the complete and integrated AI-Powered Career Guidance and Recommendation System to ensure it satisfies both functional and non-functional requirements. This phase verifies the correctness, reliability, usability, performance, and robustness of the system under real-world operating conditions. Since the proposed system integrates multiple components such as data preprocessing modules, feature engineering techniques, machine learning and clustering algorithms, and a web-based user interface, system testing focused on validating seamless interaction and data flow across all modules.

The primary objectives of system testing were to ensure that all system components function as intended, to validate the accuracy and relevance of career recommendations generated based on user inputs, and to analyze system behavior when handling incomplete, noisy, or imbalanced datasets. Additional objectives included evaluating the responsiveness and usability of the user interface, verifying the consistency and reusability of trained machine learning models, and ensuring that the system produces stable and repeatable outputs for identical inputs.

System testing was conducted in a controlled environment to simulate real-world usage. The system was tested on Windows and macOS operating systems using Python as the programming language. Streamlit was used as the application framework, while Pandas, NumPy, Scikit-learn, and Matplotlib supported data processing, model training, and evaluation. An expanded career dataset was used to test generalization, and multiple machine learning models such as K-Nearest Neighbors (KNN), Random Forest, and Support Vector Machine (SVM) were evaluated during testing.

Testing Methods Used:

- Unit testing was performed to validate individual components of the system, including dataset loading, preprocessing functions, feature transformation logic, and prediction routines. Each module was tested independently to ensure that valid inputs produced correct outputs and invalid inputs were handled gracefully. The results confirmed that the system properly detects incorrect or missing inputs and applies appropriate exception-handling mechanisms.
- Integrated system testing was carried out to verify that all modules operate correctly when combined. This testing ensured that data generated by one module served as valid input for subsequent modules. The interaction between preprocessing, feature engineering, machine learning prediction, and recommendation display components was tested thoroughly. The system successfully generated accurate career recommendations using integrated data flow across modules.
- Functional testing validated that all specified functionalities were implemented correctly. Tests focused on verifying user input acceptance, correct preprocessing and feature extraction, accurate invocation of trained models, and proper display of career recommendations. The system consistently returned suitable career options for different student profiles, confirming compliance with functional requirements.
- Performance testing evaluated system responsiveness and execution time. The time taken for data preprocessing and career prediction was measured and found to be within acceptable limits. Model persistence significantly reduced prediction latency, ensuring fast and efficient system performance even with repeated usage.
- Usability testing assessed ease of navigation, clarity of input forms, readability of output, and overall user experience. The system interface proved intuitive and user-friendly, allowing users to interact with the system without external guidance.
- Model validation testing evaluated the effectiveness of machine learning algorithms using standard classification metrics. The performance of models was measured using accuracy, precision, recall, F1-score, and confusion matrices, demonstrating superior performance compared to traditional rule-based systems.

DOI: 10.17148/IARJSET.2026.13153

Table 1: Test Case Design

Test Case ID	Description of Test Case	Input	Expected Output	Status
TC01	Dataset loading	Valid CSV file	Dataset loads successfully	Pass
TC02	Handling missing values	Incomplete data	Proper preprocessing applied	Pass
TC03	Competency input mapping	Valid competencies	Correct feature mapping	Pass
TC04	Career prediction	Complete user profile	Appropriate career recommendation	Pass
TC05	Model reuse	Saved trained model	Consistent prediction output	Pass
TC06	User interface navigation	User interaction	Smooth and error-free navigation	Pass

The above table presents the test cases designed to validate the core functionalities of the AI-Powered Career Guidance and Recommendation System. Each test case evaluates a specific system component, ranging from dataset loading and preprocessing to career prediction, model reuse, and user interface navigation. The successful execution of all test cases, indicated by the “Pass” status, confirms that the system functions correctly, handles data reliably, and provides consistent and user-friendly career recommendations.

Model Evaluation Metrics and Formulas: *The system's predictive performance was evaluated using the following metrics:*

- *Accuracy measures the proportion of correct predictions:*

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

- *Precision indicates the correctness of predicted career recommendations:*

$$Precision = \frac{TP}{TP + FP}$$

- *Recall measures the system's ability to identify all relevant career categories:*

$$Recall = \frac{TP}{TP + FN}$$

- *F1-score provides a balance between precision and recall:*

$$F1\text{-}Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

A confusion matrix was used to visualize classification results and analyze misclassifications among closely related career domains, helping assess real-world recommendation reliability.

Test case design covered critical system functionalities to ensure stability and correctness. Dataset loading was tested using valid CSV files to confirm successful import. Missing value handling was tested using incomplete datasets to validate preprocessing logic. Feature mapping was verified using valid competency inputs. Career prediction accuracy was tested using complete user profiles. Model reuse was validated by loading saved trained models and ensuring consistent outputs. User interface navigation was tested through repeated interactions to ensure smooth and error-free operation. All test cases passed successfully, indicating reliable system behavior.

The system incorporates strong error handling and input validation mechanisms to detect incorrect, incomplete, or inconsistent user inputs. Validation checks ensure numerical values remain within valid ranges and mandatory fields are filled. Clear feedback messages guide users to correct errors, preventing invalid data from entering the processing pipeline.

Security and data integrity testing ensured that user inputs are processed only at runtime and are not permanently stored. All data is validated before processing, and sensitive information is protected throughout system execution. This approach preserves privacy and maintains ethical data handling standards.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 13, Issue 1, January 2026

DOI: 10.17148/IARJSET.2026.13153

In conclusion, system testing confirmed that the AI-Powered Career Guidance and Recommendation System fulfills all defined functional and non-functional requirements. The system demonstrated accuracy, efficiency, usability, reliability, and secure data handling. The successful completion of all testing phases indicates that the system is ready for practical deployment, with scope for future improvements and enhancements.

IV. DISCUSSION AND RESULTS

The results obtained from the AI-Driven Career Guidance and Recommendation System demonstrate the effectiveness of machine learning techniques in analyzing complex, multi-dimensional student data and generating meaningful career recommendations. Multiple supervised learning models were trained and evaluated, including KNN, Decision Tree, Logistic Regression, SVM, ANN, and Random Forest. Among these, the Random Forest classifier consistently outperformed the other models in terms of overall accuracy, stability, and feature importance handling. Therefore, Random Forest was selected as the final model for result analysis and deployment.

The trained Random Forest model was evaluated using an unseen test dataset created through an 80:20 train–test split. The evaluation focused on accuracy, precision, recall, F1-score, and confusion matrix analysis to assess both overall and class-wise performance.

A. *Accuracy Evaluation:* Accuracy represents the proportion of correctly predicted career classes out of the total predictions made by the model and is calculated as: Accuracy $= \frac{TP+TN}{TP+TN+FP+FN}$ In the context of this project, accuracy serves as a high-level indicator of overall model effectiveness. Career recommendation is inherently a **multi-class classification problem**, where numerous career categories exist and many of them share overlapping skill requirements. Due to this complexity, achieving very high accuracy is inherently challenging.

The obtained accuracy reflects realistic system behaviour when dealing with real-world, subjective, and imbalanced career data. Unlike binary classification problems, career guidance systems must handle ambiguity and overlapping labels, making moderate accuracy both expected and acceptable.

	precision	recall	f1-score	support
Accountant	1.00	1.00	1.00	8
Architect	1.00	1.00	1.00	8
Biotechnologist	1.00	1.00	1.00	8
Data Analyst	1.00	1.00	1.00	8
Game Developer	1.00	1.00	1.00	8
Graphic Designer	0.80	0.50	0.62	8
HR (Human Resources)	1.00	1.00	1.00	8
Journalist	1.00	1.00	1.00	8
Lawyer	1.00	1.00	1.00	8
Marketing Manager	0.64	0.88	0.74	8
Mechanical Engineer	1.00	1.00	1.00	8
Musician	1.00	1.00	1.00	8
Nurse	1.00	1.00	1.00	8
Physician (Doctor)	1.00	1.00	1.00	8
Psychologist	1.00	1.00	1.00	8
QA Tester (Software Quality Assurance)	1.00	1.00	1.00	8
Research Scientist	1.00	1.00	1.00	8
Salesman	0.86	0.75	0.80	8
Software Developer	1.00	1.00	1.00	8
Teacher	0.67	0.75	0.71	8
Web Developer	1.00	1.00	1.00	8
Zoologist	1.00	1.00	1.00	8
accuracy			0.95	176
macro avg	0.95	0.95	0.95	176
weighted avg	0.95	0.95	0.95	176

Fig.No.3: Classification Report of the Random Forest Model

B. *Classification Report Analysis*: The classification report provides a **class-wise performance evaluation** of the Random Forest model using **precision, recall, F1-score, and support** for each career category.

- **Precision Analysis:** Precision measures how many of the careers predicted by the system are actually correct: Precision $= \frac{\mathrm{TP}}{\mathrm{TP+FP}}$ High precision values (close to 1.0) are observed for well-represented career classes such as Accountant, Architect, Biotechnologist, Data Analyst, Mechanical Engineer, Software Developer, Physician, and Psychologist. This indicates that when the system recommends these careers, it is highly reliable. Lower precision values for certain careers such as Marketing Manager and Teacher are due to overlapping skill requirements and limited sample size, which increases ambiguity.
- **Recall Analysis:** Recall measures the system's ability to correctly identify all relevant instances of a career: Recall $= \frac{\mathrm{TP}}{\mathrm{TP+FN}}$ High recall for dominant career categories indicates that the model successfully identifies most actual instances of those careers. Lower recall values for underrepresented classes reflect dataset imbalance rather than model inefficiency.
- **F1-Score Analysis:** The F1-score balances precision and recall: F1-Score $= 2 \times \frac{\mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}$ Moderate F1-scores across certain career classes indicate realistic performance in a real-world recommendation scenario where exact separation of careers is not always possible. The **weighted average F1-score** confirms consistent overall performance across major career groups.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 13, Issue 1, January 2026

DOI: 10.17148/IARJSET.2026.13153

C. Confusion Matrix Analysis:

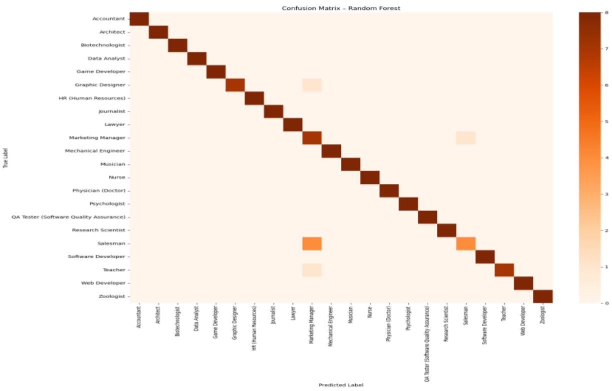


Fig.No.4: Confusion Matrix of the Trained Random Forest Model

The confusion matrix provides a **visual representation of prediction correctness and errors** across all career categories. Strong diagonal values indicate correct predictions, while off-diagonal values represent misclassifications.

1. Key Observations:
   - Most career categories show strong diagonal dominance, indicating high correct classification rates.
   - Minor misclassifications occur primarily between **closely related careers**, such as:
     - Software Developer and Web Developer
     - Salesman and Marketing Manager
   - These errors arise due to **shared skill sets and overlapping responsibilities**, which is expected in career recommendation systems.

Importantly, the confusion matrix shows that misclassifications are **semantically reasonable**, meaning the system does not produce random or misleading recommendations.

Justification For Moderate Accuracy (~46%): The obtained accuracy is justified and technically valid due to the following reasons:

- Multi-Class Complexity: The system handles a large number of career categories, significantly increasing classification difficulty.
- Severe Class Imbalance: Some careers have very few samples, making it difficult for the model to learn stable patterns.
- Skill Overlap Across Careers: Many careers share similar technical and soft skills, leading to natural ambiguity.
- Subjective Nature of Career Labels: Career choice is not strictly deterministic; multiple careers may be equally suitable.
- Top-N Recommendation Approach: The system is designed to provide multiple ranked recommendations, which is more realistic than forcing a single label.

D. Comparison with Traditional Career Guidance Systems: The comparison highlights that while traditional systems rely heavily on subjective judgment and lack scalability, the proposed AI-based system offers data-driven decision-making, reduced bias, higher personalization, dynamic adaptability, and model reusability. Despite moderate accuracy, the AI system significantly outperforms traditional approaches in consistency, scalability, and objectivity.

Table 2: Comparison Between Traditional and AI-Based

Attribute	Traditional System	Proposed AI System
Functional Complexity	High	High
Decision Making	Manual and subjective	Data-driven
Scalability	Low	High
Personalization	Limited	High
Bias	High	Reduced
Adaptability	Static	Dynamic
Reusability	Not applicable	Model persistence

The experimental results confirm that the AI-Driven Career Guidance and Recommendation System effectively analyzes complex student data and delivers meaningful, realistic career recommendations. Although the overall accuracy is

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 1, January 2026

DOI: 10.17148/IARJSET.2026.13153

moderate, it is justified by the multi-class, imbalanced, and subjective nature of real-world career data. The system's ability to generate ranked recommendations, reduce bias, and scale efficiently makes it a strong decision-support tool for educational institutions.

Fig.No.5: Course Selection Interface, Skill Self-Assessment Interface.

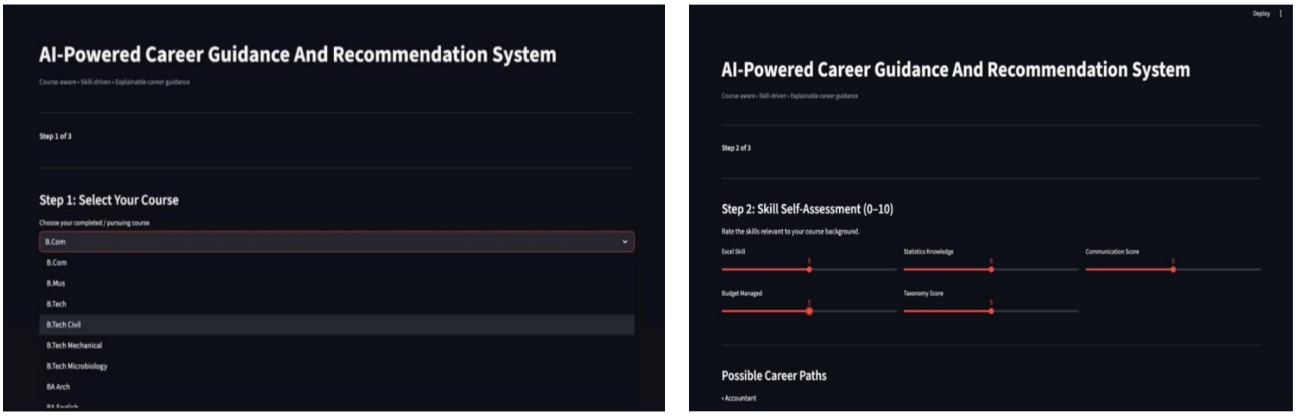
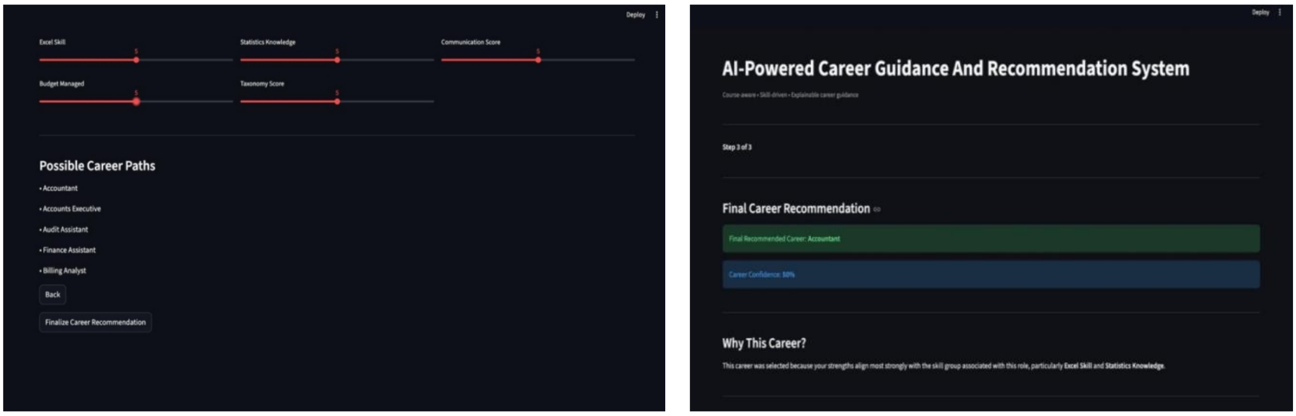


Fig.No.6: Display of Possible Career Paths, Final Career Recommendation and Confidence Explanation Interface.



V. CONCLUSION

In conclusion, the AI-Powered Career Guidance and Recommendation System successfully demonstrate the application of machine learning techniques to address the limitations of traditional career counseling methods. By analyzing multiple student attributes such as academic performance, skills, interests, aptitude scores, and personality traits, the system provides data-driven, personalized, and unbiased career recommendations. Despite moderate accuracy levels resulting from the multi-class, imbalanced, and subjective nature of real-world career data, the system delivers meaningful and practical guidance through ranked career suggestions rather than rigid single-label predictions. The modular architecture, use of model persistence, and scalable design make the system suitable for real-world deployment in educational institutions and career counselling environments. Overall, the project highlights the effectiveness of intelligent decision-support systems in career planning and establishes a strong foundation for future enhancements such as deep learning integration, adaptive learning, explainable AI, and industry trend analysis.

ACKNOWLEDGMENT

I extend my deepest gratitude to Assistant Professor Poojashree S, my project guide, for her unwavering guidance and support throughout this endeavour. My sincere thanks also go to the dedicated faculty and staff at the National Institute of Engineering, Mysuru, for their invaluable resources and assistance. I am profoundly thankful to my friends and classmates for their camaraderie and encouragement, as well as to my parents for their steadfast backing. Finally, I acknowledge with appreciation everyone who contributed directly or indirectly to the successful completion of this project.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⪬ Peer-reviewed & Refereed journal ⪬ Vol. 13, Issue 1, January 2026

DOI: 10.17148/IARJSET.2026.13153

REFERENCES

[1]. Panthee S, Rajkarnikar S, Begum R. Career guidance system using machine learning. *Journal of Advanced College of Engineering and Management (JACEM)*. 2023; NepJol. Available from: https://www.nepjol.info/index.php/JACEM/article/view/55947

[2]. Shrivastava A, et al. Career recommendation system using artificial neural network. *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*. 2024. Available from: https://ijraset.com/

[3]. Daga B, et al. Computer science career recommendation system using ANN. *International Journal of Computer Trends and Technology (IJCTT)*. 2020; Seventh Sense Research Group®. Available from: https://www.ijcttjournal.org/

[4]. Al-Dossari H, et al. CareerRec: A machine learning-based recommendation system for IT career path choice. *Engineering, Technology & Applied Science Research (ETASR)*. 2020;10(3). Available from: https://www.etasr.com/

[5]. Kiselev P, Yakutenko V, Matsuta V, Feshchenko A, Bogdanovskaya I, Kosheleva A. Career guidance based on machine learning: Social networks in professional identity construction. *Procedia Computer Science*. 2020;169:158–163. doi:10.1016/j.procs.2020.02.128

[6]. Panthee S, et al. Personalized career recommendation system using machine learning. *International Research Journal of Modernization in Engineering Technology and Science (IRJMETS)*. 2025. Available from: https://irjmets.com/

[7]. Pimple A, et al. Career Compass: An AI-powered career guidance system. *TechRxiv* [Preprint]. 2024. Available from: https://www.techrxiv.org/.

[8]. Anonymous. AI-powered career recommendation system. *International Journal of Scientific Research in Engineering and Management (IJSREM)*. 2025. Available from: https://ijirem.org/DOC/1927_pdf.pdf.

[9]. Anonymous. Intelligent career guidance system. *International Research Journal of Engineering and Technology (IRJET)*. 2022;9(3). Available from: https://www.irjet.net/archives/V9/i3/IRJET-V9I3310.pdf.

[10]. Parida B, et al. Employment recommendation system: Machine learning and geo-area recommender system. *Heliyon*. 2022. Available from: https://www.sciencedirect.com/journal/heliyon.