iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** VI **Month of publication:** June 2026

DOI: https://doi.org/10.22214/ijraset.2026.83686

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 14 Issue VI June 2026- Available at www.ijraset.com

Smart Career Guidance Platform Using AI for Career Mapping and Job Role Prediction

Dr. K. SubbaRao[1], Dabbakuti John Victor[2], Murugula Shanmuk[3], Jidugu Bhargava Adithya[4], Bodipaga Siva Nagaraju[5]
[1]*Professor & HOD,* [2,3,4,5]*Student, Department of CSE – Data Science, St. Ann's College of Engineering & Technology, Chirala, India*

Abstract: ***Choosing an appropriate career path is among the most consequential decisions in a student's academic life, yet it remains a challenging endeavour owing to the lack of structured guidance, rapidly evolving job market dynamics, and limited awareness of requisite skills across domains. Conventional career counseling approaches often fail to deliver personalized recommendations, leading to suboptimal career choices and academic disengagement. This paper presents a Smart Career Guidance Platform built upon Artificial Intelligence, designed to assist students in identifying suitable career paths by analyzing their skills, interests, and behavioral preferences. The system employs a two-stage machine learning pipeline: a Random Forest classifier predicts the most appropriate career domain from user survey responses, while a LightGBM classifier recommends the top three relevant job roles based on the user's technical skill profile. Feature selection techniques including Correlation Analysis, Principal Component Analysis (PCA), and SHAP (SHapley Additive exPlanations) are employed to enhance prediction accuracy and model interpretability. The platform additionally generates a personalized career roadmap providing step-by-step guidance for skill development and career growth. Implemented as a Streamlit-based web application, the system achieves a prediction accuracy of approximately 90% for career field classification, demonstrating the effectiveness of AI-driven decision-support systems in the domain of career guidance.***

Keywords: Career Guidance; Machine Learning; Random Forest; LightGBM; Career Prediction; Job Role Recommendation; Streamlit; Feature Selection; Personalized Roadmap; Artificial Intelligence

I. INTRODUCTION

The rapid advancement of technology and the growing diversity of career opportunities have rendered career selection an increasingly complex challenge for students and early-career professionals. Traditional career guidance methodologies, which rely predominantly on manual counseling, general aptitude assessments, and static informational resources, often fail to account for the multidimensional nature of individual skills, interests, and aspirations. Moreover, these conventional approaches lack the adaptability required to keep pace with the continuously evolving demands of the modern job market, where new roles and skill requirements emerge at an unprecedented rate.

In response to these limitations, intelligent career guidance systems leveraging Artificial Intelligence (AI) and Machine Learning (ML) have garnered significant scholarly and practical attention. Abisoye Opeyemi et al. [1] highlighted the importance of web-based career guidance platforms in providing structured and accessible recommendations to students, while El-Sofany and El-Seoud [2] demonstrated how cloud-based solutions can enhance the scalability and remote accessibility of such systems. García et al. [3] further established that machine learning techniques can be effectively employed to assess student skills and recommend appropriate career paths. Similarly, Gokarn et al. [4] and Kanathur et al. [5] showed that AI-driven systems can analyze diverse user data to deliver personalized career recommendations with significantly greater accuracy than traditional approaches.

Despite these advances, several research gaps remain. Many existing systems rely on single-modality inputs such as academic performance or aptitude scores, without integrating behavioral, interest-based, and skill-level features comprehensively. Furthermore, most systems do not provide a structured post-prediction roadmap to guide users toward achieving their career goals, and the absence of interactive interfaces limits practical adoption.

This paper presents a Smart Career Guidance Platform employing a two-stage machine learning pipeline to predict suitable career domains and recommend relevant job roles. The platform integrates Random Forest for career field classification and LightGBM for job role recommendation, delivering a comprehensive, data-driven career guidance experience through an interactive Streamlit web application. In addition to career predictions, the system generates a personalized career roadmap with structured learning phases, relevant topics, and recommended certifications.

A. Objectives of the System

The primary objective of this work is to develop an intelligent, AI-based career guidance system that assists students in selecting suitable career paths based on their individual profiles. The specific objectives include: (i) developing an AI-based system for accurate career field prediction; (ii) analyzing user skills, interests, and preferences using machine learning techniques; (iii) recommending the top three suitable job roles for each user; (iv) generating a structured and personalized career development roadmap; and (v) delivering all capabilities through a user-friendly Streamlit web interface.

B. Role of AI in Career Guidance

Artificial Intelligence has emerged as a powerful solution for modern career guidance by learning patterns from user skills, interests, and historical data; handling high-dimensional and complex datasets efficiently; adapting to changing job market trends and emerging roles; and providing personalized recommendations while reducing human bias in career decision-making. Machine learning models identify complex relationships between skills, career domains, and job roles that would be infeasible to derive through manual analysis, enabling data-driven, objective career recommendations at scale.

II. LITERATURE SURVEY

Career guidance systems have evolved substantially from rule-based approaches to sophisticated AI-driven platforms. This section reviews existing methodologies, input types, feature selection methods, and machine learning models applied to career prediction, and identifies research gaps addressed by the proposed system.

A. Existing Career Guidance Approaches

Early career guidance systems employed rule-based logic providing recommendations based on predefined conditions. While straightforward to implement, such systems lacked the flexibility and personalization necessary for individual-level guidance. Holland [8] introduced the Theory of Career Choice classifying individuals and environments into six broad categories; however, this framework lacks adaptability for dynamic contemporary job markets. Super [9] proposed a life-span, life-space approach and Lent et al. [10] contributed Social Cognitive Career Theory, providing foundational theoretical frameworks not designed for computational implementation.

K-Nearest Neighbors (KNN) has been explored for career prediction through similarity-based matching, but becomes computationally expensive with large datasets. Naive Bayes classifiers offer probabilistic classification, though the assumption of feature independence does not hold when skills and interests are correlated. Gunwant et al. [7] conducted a systematic literature review of career guidance expert systems, highlighting persistent limitations of traditional approaches. Dahanke et al. [6] proposed an intelligent career guidance system using machine learning, further demonstrating the need for more robust and personalized prediction frameworks.

B. Types of Inputs Used in Career Prediction

Career prediction systems rely on multiple input types to improve accuracy and personalization. Skill-based inputs encompass technical skills such as programming and data analysis as well as domain-specific knowledge. Interest-based inputs capture user preferences for specific domains such as IT, Business, and Design. Behavioral and personality inputs reflect problem-solving ability, analytical thinking, and communication skills, which vary across career roles. Educational and background data including academic qualifications and certifications provide context about user knowledge levels. User response data collected through structured questionnaires and self-assessment surveys converts preferences into structured formats consumable by machine learning models.

C. Feature Selection Methods

Feature selection is a crucial preprocessing step that identifies the most relevant input features while removing redundant data, improving model performance and reducing overfitting. Correlation Analysis removes highly correlated features to eliminate redundancy. Principal Component Analysis (PCA) transforms correlated features into orthogonal components for effective dimensionality reduction. SHAP (SHapley Additive exPlanations) quantifies each feature's contribution to model predictions, providing quantitative importance scores and improved model interpretability.

D. Machine Learning Models for Career Prediction

Random Forest, an ensemble of multiple decision trees, offers high accuracy and reduces overfitting through bagging, though it may require substantial memory for very large datasets. LightGBM provides faster training speeds and efficient processing of large structured datasets through its leaf-wise growth strategy, though it can be sensitive to hyperparameter settings. Kaur and Sharma [12] demonstrated the superiority of ensemble methods over individual classifiers for career guidance. Singh and Kumar [13] proposed an AI-based career recommendation system showing improved personalization. Patel and Shah [14] surveyed career prediction using machine learning, and Gupta and Verma [15] proposed an AI-based counseling system for students, collectively affirming the practical value of machine learning in this domain.

E. Research Gaps

The literature review reveals several persistent gaps: lack of personalization in systems that provide generalized suggestions without considering individual skills and interests; limited use of comprehensive multi-modal data with many systems relying on static or narrow datasets; insufficient integration of behavioral and interest-based inputs alongside academic features; absence of structured career roadmaps guiding users beyond the prediction stage; and lack of interactive, user-friendly interfaces. The proposed system addresses these gaps through multi-modal inputs, advanced ensemble models, personalized roadmap generation, and an accessible Streamlit web application.

III. PROPOSED METHODOLOGY

The proposed Smart Career Guidance Platform follows a systematic methodology encompassing data collection, preprocessing, feature selection, two-stage model training, and output generation. The methodology is designed to transform raw user inputs into structured, accurate, and personalized career recommendations.

A. System Overview

The system implements a two-stage machine learning pipeline. In Stage 1, the Random Forest algorithm predicts the most suitable career domain based on user survey responses capturing interests, preferences, and behavioral traits. In Stage 2, the LightGBM algorithm recommends the top three job roles based on the user's selected technical skill set and the Stage 1 predicted career domain. The system additionally generates a personalized career roadmap comprising learning phases, core topics, and recommended certifications.

B. Data Collection and Preprocessing

The system utilizes structured datasets comprising information on skills, career domains, and job roles. Data sources include career datasets in CSV format, user input collected through structured questionnaires, educational background data, and job role datasets mapping required skills to career opportunities. Table I presents the data sources used in the system.

Table I
Data Sources Used In The System

Source	Data Type	Purpose
Career Dataset (CSV)	Skill sets, career domains, job roles	Core data for model training
User Input (Questionnaire)	Skills, interests, preferences	Personalized career prediction
Educational Data	Academic performance, subjects	User background analysis
Job Role Dataset	Skills required for specific jobs	Skill-to-career mapping

Data preprocessing encompasses: handling missing values through default value assignment and mean imputation; standardizing categorical skill representations using predefined encoding schemes; normalizing numerical features using Min-Max Scaling to a range of 0 to 1; and detecting and treating outliers using Z-Score analysis and the Interquartile Range (IQR) method.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VI June 2026- Available at www.ijraset.com

C. Data Splitting and Transformation

The preprocessed dataset is divided into training and testing subsets using an 80:20 split with stratified sampling to ensure proportional class representation. Cross-validation techniques are applied to enhance model robustness across varying data distributions. Categorical features are transformed using LabelEncoder for the career field target variable and MultiLabelBinarizer for encoding skill sets into binary indicator matrices.

D. Feature Selection Techniques

The system employs three complementary feature selection techniques. Correlation Analysis removes redundant or highly related features. PCA reduces dimensionality in high-dimensional feature spaces by projecting data onto orthogonal principal components. SHAP analysis quantifies and ranks each feature's contribution to model predictions. Table II presents the key input features used in the system.

Table II
Key Features Used In The System

Feature	Type	Description
Technical Skills	Skill-Based	Programming, data analysis, domain-specific knowledge
Soft Skills	Behavioral	Communication, teamwork, problem-solving ability
User Interests	Preference-Based	Areas of interest: IT, Business, Design, etc.
Academic Performance	Educational	Subject strengths and academic background
Career Preferences	Sentiment	Preferred job roles or career domains

E. Model Selection and Hyperparameter Optimization

Random Forest is selected for Stage 1 for its ability to handle large, high-dimensional datasets effectively; its ensemble approach that reduces overfitting by aggregating majority votes from multiple decision trees; and its compatibility with both categorical and numerical features. LightGBM is selected for Stage 2 for its leaf-wise tree growth strategy achieving superior accuracy; its significantly faster training speed compared to traditional gradient boosting; and its efficient memory utilization for large structured datasets. Table III presents the hyperparameters tuned for both models.

TABLE III
Hyperparameters Used In Model Optimization

Parameter	Description	Model Affected
Number of Trees	Controls the number of decision trees in the ensemble	Random Forest
Max Depth	Maximum depth of each decision tree	Random Forest, LightGBM
Learning Rate	Rate at which the model adapts during gradient descent	LightGBM
Number of Leaves	Determines the complexity of the LightGBM tree	LightGBM
Min Samples Split	Minimum samples required to split an internal node	Random Forest

IV. SYSTEM DESIGN

The system architecture organizes the Smart Career Guidance Platform into three principal layers. The Input Layer collects user survey responses and skill selections through the Streamlit web interface. The Processing Layer performs data preprocessing, feature encoding, and two-stage machine learning inference using the trained Random Forest and LightGBM models. The Output Layer presents the predicted career field, top three job roles, and a personalized career roadmap.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VI June 2026- Available at www.ijraset.com

A. Architectural and Logical Design

The architectural design defines the overall structure and component interactions of the system. The system workflow begins with user input collection (skills and preferences), proceeds through data preprocessing and feature encoding, invokes the Random Forest model for career field prediction, and then passes the combined feature representation to LightGBM models for job role recommendation, finally generating and displaying the career roadmap. The logical design employs LabelEncoder for the career field target variable and MultiLabelBinarizer for skill set encoding, transforming raw categorical data into structured numerical matrices suitable for machine learning models.

B. UML Design Documentation

The system design is documented through a comprehensive set of UML diagrams. The Class Diagram captures the static structure depicting five principal classes: User (storing skills, interests, and preferences), Career (containing domain information), JobRole (representing roles linked to career domains), Model (handling ML-based prediction), and Dataset (storing training data). The State Diagram illustrates dynamic state transitions from user input collection through preprocessing, model inference, and result presentation. The Use Case Diagram identifies interactions between the student user and system functionalities including survey completion, skill selection, prediction, and roadmap generation. The Sequence Diagram details the temporal ordering of messages exchanged between system components during a prediction session. The Block Diagram provides an end-to-end overview of the data flow from user input through preprocessing, encoding, Random Forest prediction, LightGBM recommendation, and final output display.

V. IMPLEMENTATION

A. Software and Hardware Requirements

The system is implemented using Python 3.10 with the following core dependencies: Streamlit 1.41.1 for the web application interface; Pandas 2.2.3 and NumPy 1.23.5 for data manipulation and numerical computation; Scikit-learn 1.6.1 for preprocessing utilities and model evaluation; LightGBM 4.5.0 for gradient boosting classification; Matplotlib 3.10.1, Seaborn 0.13.2, and Plotly 5.18.0 for data visualization; Joblib 1.4.2 for model serialization; and Imbalanced-learn 0.12.0 for handling class imbalance. Development is conducted in the Anaconda environment using Jupyter Notebook for exploratory data analysis and iterative model experimentation. Minimum hardware requirements are an Intel i3 processor, 4 GB RAM, and 10 GB free disk space, with an Intel i5 processor and 8 GB RAM recommended for optimal performance.

B. Implementation Modules

The implementation is organized into five modules. The Data Preparation module loads career field and skills datasets from CSV files, performs missing value imputation, removes duplicate records, and standardizes skill category representations. The Feature Encoding module applies LabelEncoder to the career field target variable and MultiLabelBinarizer to skill strings, producing binary indicator matrices. The career field column is additionally one-hot encoded, and the skill indicators and career field encodings are horizontally stacked to form the combined feature matrix for LightGBM training.

The Model Training module implements the two-stage pipeline. A Random Forest classifier is trained on encoded survey response features with stratified 80:20 train-test splitting. LightGBM classifiers are trained on the combined feature matrix to produce job role recommendations. Trained models are serialized to disk as Pickle files using Joblib for efficient inference-time loading. The Prediction module processes incoming user inputs through the same preprocessing pipeline applied during training, generates career field predictions decoded via LabelEncoder's inverse_transform, and passes the combined representation to LightGBM for top-three job role recommendations. The Roadmap Generation module retrieves pre-defined structured learning paths based on the predicted career field and job roles, presenting sequential phases, relevant topics, and certifications.

C. Streamlit Web Application

The front-end interface is developed using Streamlit, providing an interactive platform where users complete a structured multi-section survey capturing behavioral preferences and domain interests, select technical skills from a categorized interface, and receive predicted career fields, top three job roles, and structured learning roadmaps. The Streamlit framework ensures rapid deployment, responsive design, and accessibility for users without technical expertise.

VI. TESTING AND VALIDATION

A comprehensive testing protocol was designed to validate each functional module of the system. Seven test cases were developed covering data loading, preprocessing, feature extraction, train-test splitting, user input encoding, career prediction output, and roadmap generation. Table IV presents the test scenarios, expected outcomes, actual outcomes, and pass/fail status for all test cases.

Table IV
System Testing And Validation Results

Test Case ID	Test Scenario	Expected Result	Actual Result	Status
TC_01	Data Loading Test	Datasets loaded with required columns	Data loaded successfully	Pass
TC_02	Data Preprocessing	Correct encoding without errors	Encoding completed successfully	Pass
TC_03	Feature Extraction	Skills converted to numerical format	Skills encoded successfully	Pass
TC_04	Train-Test Split	80/20 split, no missing values	Split successful	Pass
TC_05	User Input Encoding	Correct numerical transformation	Input encoded successfully	Pass
TC_06	Career Prediction Output	Accurate career field and top 3 roles	Predictions displayed correctly	Pass
TC_07	Roadmap Generation	Roadmap with phases and certifications	Roadmap generated successfully	Pass

All seven test cases passed successfully, confirming the functional correctness of each system module. The data loading and preprocessing tests verified encoding pipeline integrity. Feature extraction and train-test split tests confirmed correct data transformation and partitioning. Career prediction and roadmap generation tests validated the end-to-end workflow, confirming accurate, structured, and user-interpretable outputs.

VII. RESULTS AND DISCUSSION

The Smart Career Guidance Platform produces three primary outputs for each user session: the predicted career field, the top three recommended job roles, and a personalized career development roadmap.

A. User Interface

The Streamlit-based interface presents users with a structured multi-section survey capturing behavioral preferences and domain interests, followed by a categorized skill selection interface. The layout is designed for accessibility and ease of use, requiring no specialized technical background from the end user.

B. Career Field Prediction

Upon submission of survey responses, the system preprocesses the inputs and invokes the trained Random Forest model to predict the most suitable career field. The prediction result is displayed along with a descriptive explanation of the career domain. The two-stage prediction approach, where the career domain is first determined before specific roles are recommended, constrains the job role search space and improves the relevance of final recommendations.

C. Job Role Recommendation

The LightGBM-based Stage 2 processes the combination of the predicted career field and the user's selected technical skills to generate the top three job role recommendations. Each recommendation is presented with a description of the role and its alignment with the user's profile, enabling users to make an informed comparison between options.

D. Career Roadmap Generation

The roadmap generation module presents a structured learning plan comprising sequential development phases, core topics within each phase, and recommended certifications. This transforms the prediction from a static recommendation into actionable guidance, helping users understand not only which career to pursue but also how to systematically develop the competencies required to achieve it.

E. Performance

The system achieves a prediction accuracy of approximately 90% for career field classification on the held-out test set, demonstrating the effectiveness of the Random Forest ensemble approach. This accuracy level indicates reliable generalization to unseen user profiles across diverse skill and interest combinations. The use of stratified splitting and cross-validation ensures that the reported performance reflects consistent model behavior across varying data distributions rather than a specific random partition.

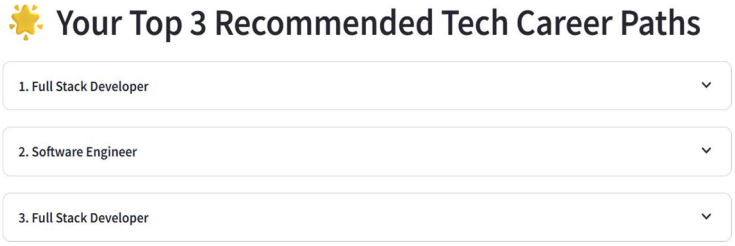


Figure 1. Output from Top Career Path Recommendations

V. CONCLUSION

This paper has presented a Smart Career Guidance Platform leveraging a two-stage machine learning pipeline comprising Random Forest and LightGBM to deliver personalized career field predictions and job role recommendations. The system achieves approximately 90% accuracy for career field classification, demonstrating the practical effectiveness of ensemble machine learning in the career guidance domain.

The platform addresses key limitations of existing systems by integrating multiple input modalities including technical skills, soft skills, behavioral preferences, and academic background; providing structured post-prediction career roadmaps with learning phases, topics, and certifications; and delivering all functionalities through an accessible Streamlit web application. The comprehensive testing protocol confirmed functional correctness and reliability across all seven test scenarios.

The proposed system demonstrates that AI-driven, data-centric career guidance tools have significant potential to assist students and early-career professionals in making informed, evidence-based career decisions, reducing uncertainty, and improving alignment between individual competencies and industry requirements.

VI. FUTURE SCOPE

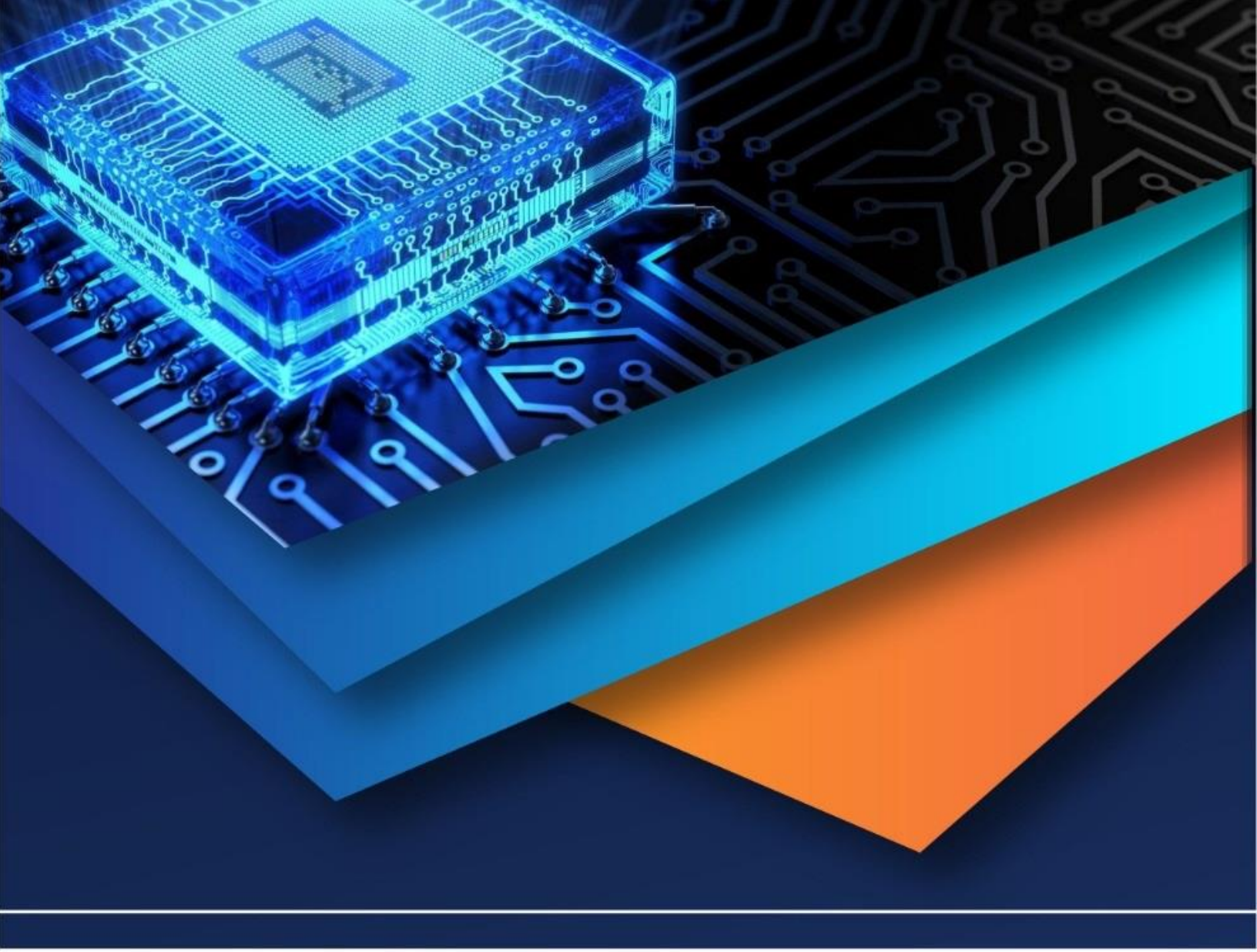
Several enhancement directions have been identified. The integration of deep learning architectures such as Artificial Neural Networks and Transformer-based models may capture more complex non-linear relationships in user data and further improve prediction accuracy for large and diverse user populations. Connecting the platform to real-time job market data through APIs from employment platforms such as LinkedIn and Indeed would enable dynamic, up-to-date job opportunity recommendations aligned with current industry hiring trends. The incorporation of Natural Language Processing techniques would allow users to describe their skills and interests in free-text format, enhancing interaction flexibility. Integration with online learning platforms such as Coursera, Udemy, and edX would enable the system to recommend specific courses and certifications tailored to identified skill gaps. Expanding the dataset to encompass global career roles and international job market profiles would broaden the system's applicability to a wider and more diverse user audience.

VII. ACKNOWLEDGMENT

The authors express their sincere gratitude to Dr. K. Subbarao, Guide and Head of the Department of CSE – Data Science, St. Ann's College of Engineering and Technology, Chirala, for his invaluable guidance and continuous support throughout this project. The authors also thank Dr. K. Jagadeesh Babu, Principal, and the management of St. Ann's College of Engineering and Technology for providing the necessary facilities, laboratory resources, and a stimulating research environment. The cooperation of all teaching and non-teaching staff of the Department of CSE – Data Science is gratefully acknowledged.

REFERENCES

[1] A. O. Abisoye, I. O. Alabi, S. O. Ganiyu, B. O. Abisoye, and J. Omokore, "A web-based career guidance information system for pre-tertiary institution students in Nigeria," Int. J. Sci. Res. Sci. Eng. Technol. (IJSRSET), Jun. 2015.
[2] H. F. El-Sofany and M. S. A. El-Seoud, "The implementation of career and educational guidance system (CEGS) as a cloud service," Int. J. Emerg. Technol. Learn. (iJET), vol. 15, no. 20, Oct. 2020.
[3] A. J. García, A. Sneyd, A. Melro, and A. Ollagnier, "C3 IoC: A career guidance system for assessing student skills using machine learning and network visualization," Int. J. Artif. Intell. Educ., Dec. 2022.
[4] S. Gokarn, S. Taware, and R. Vartak, "Smart career guidance system using machine learning," Int. J. Sci. Res. (IJSR), Jun. 2024.
[5] R. Kanathur, R. Akhilesh, A. Soumya, and M. K. Huang, "Intelligent career guidance systems," Int. J. Innov. Sci. Res. Technol., Aug. 2023.
[6] A. Dahanke, N. Shinde, A. Dhagate, and H. Shaikh, "An intelligent career guidance system using machine learning," Int. Res. J. Eng. Technol. (IRJET), vol. 9, no. 3, Mar. 2022.
[7] S. Gunwant, J. Pande, and R. Bisht, "A systematic study of the literature on career guidance expert systems for students: Implications for ODL," J. Learn. Dev., Nov. 2022.
[8] J. L. Holland, Making Vocational Choices: A Theory of Vocational Personalities and Work Environments. Odessa, FL: Psychol. Assess. Resour., 1997.
[9] D. E. Super, "A life-span, life-space approach to career development," J. Vocat. Behav., 1980.
[10] R. W. Lent, S. D. Brown, and G. Hackett, "Social cognitive career theory," J. Vocat. Behav., 1994.
[11] M. L. Savickas, "Career construction theory and practice," in Career Development and Counseling, 2005.
[12] P. Kaur and M. Sharma, "Career guidance system using machine learning techniques," Int. J. Comput. Appl., 2021.
[13] A. Singh and R. Kumar, "Career recommendation system using artificial intelligence," Int. J. Eng. Res. Technol. (IJERT), 2020.
[14] D. Patel and P. Shah, "A survey on career prediction using machine learning," Int. J. Adv. Res. Comput. Sci., 2022.
[15] S. Gupta and A. Verma, "AI-based career counseling system for students," Int. J. Comput. Sci. Inf. Technol., 2023.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY