International Journal of Creative and Open Research in Engineering and Management
ISSN: **3108-1754 (Online)**
Volume 02 Issue 05 May-2026 | Impact Factor: 3.5

AI-Based Career Path Recommendation System

HARSHINI E[1], PRIYAJANANI R[2], KAVIYA D[3], MR .SUGUMARAN V R[4]

1, 2, 3 Members - 5th Semester B.E Students, Department of Computer Science and Engineering, E.G.S.Pillay Engineering College, Nagapattinam, Tamilnadu, India
4 Professor, Department of Computer Science and Engineering, E.G.S.Pillay Engineering College, Nagapattinam, Tamilnadu, India

How to Cite this Article:

E, H., R, P. & D, K. (2026). AI-Based Career Path Recommendation System. International Journal of Creative and Open Research in Engineering and Management, <i>02</i>(05). https://doi.org/10.55041/ijcope.v2i5.338

https://doi.org/10.55041/ijcope.v2i5.338

Abstract — Career selection is one of the most pivotal decisions in a student's academic journey, yet a significant proportion of engineering graduates remain uncertain about the most suitable career path for their profile. Traditional career guidance approaches rely predominantly on academic performance metrics such as cumulative grade point average (CGPA), thereby neglecting critical factors including technical skills, certifications, extracurricular achievements, and professional networking activities. This paper presents an AI-Based Career Path Recommendation System designed to address these shortcomings by leveraging machine learning algorithms to provide personalized, data-driven career guidance to engineering students. The proposed system collects multidimensional student profile data encompassing CGPA, technical skill sets, certification records, and LinkedIn profile attributes. Following rigorous data preprocessing and feature selection, the system applies supervised classification models, specifically Decision Tree and Random Forest classifiers, to identify latent patterns between student attributes and industry job roles. Experimental evaluation conducted on diverse student datasets demonstrates a prediction accuracy in the range of 90–95%, substantiating the effectiveness of the proposed approach. The system significantly outperforms conventional methods by incorporating a broader attribute space and producing contextually relevant career recommendations. The findings suggest that AI-driven recommendation frameworks have considerable potential for integration into academic career counselling platforms to support informed decision-making among engineering students.

Keywords — Machine Learning, Loan Prediction, Classification, Banking System, Data Analysis

I. INTRODUCTION

The rapid evolution of the technology industry has introduced a diverse and continuously expanding spectrum of career opportunities for engineering graduates. Disciplines such as artificial intelligence, cloud computing, cybersecurity, data science, and full-stack development have emerged as dominant career domains, each requiring a distinct combination of technical competencies and soft skills. Despite this wealth of opportunities, a substantial number of students complete their undergraduate programmes without a clear understanding of which career path aligns best with their individual profile, resulting in suboptimal career choices, extended periods of unemployment, and occupational dissatisfaction.

Career guidance in academic institutions has traditionally been conducted through periodic counselling sessions, aptitude tests, and alumni interactions. While these mechanisms provide a degree of direction, they are inherently limited by their manual nature, lack of scalability, and inability to process the multidimensional attributes that define a student's suitability for a specific role. Furthermore, such methods are often reactive rather than proactive, addressing career confusion only after it has manifested.

The proliferation of machine learning (ML) and artificial intelligence (AI) technologies has opened new avenues for the automation of complex pattern recognition and classification tasks. Recommendation systems built upon ML frameworks have demonstrated remarkable success in domains such as e-commerce, content streaming, and healthcare. The application of similar principles to academic career guidance presents a compelling opportunity to deliver scalable, personalised, and objective career recommendations.

This paper proposes an AI-Based Career Path Recommendation System that integrates multiple student attributes—including CGPA, technical skill sets, industry certifications, and LinkedIn profile information—to predict the most suitable engineering career paths. By employing Decision Tree and Random Forest classifiers, the system learns associations between student profiles and career outcomes, enabling it to generate accurate recommendations for new students. The remainder of this paper is structured as follows: Section II defines the problem statement; Section III enumerates the system objectives; Section IV presents the literature review; Section V describes the proposed system; Section VI details the system architecture; Section VII elaborates on the methodology; Section VIII discusses the results; Section IX addresses limitations and future scope; and Section X concludes the paper.

II. PROBLEM STATEMENT

The central problem addressed in this research is the inadequacy of existing career guidance mechanisms available to engineering students. A considerable number of graduating students experience significant confusion and uncertainty when evaluating career options, primarily due to limited exposure to industry roles and a lack of awareness of how their individual skill sets align with job market demands. This confusion is further compounded by the absence of a structured, data-driven framework that can systematically analyse a student's academic and extracurricular profile to generate targeted career recommendations.

Existing career recommendation systems predominantly focus on academic performance metrics, particularly CGPA, as the primary determinant of career suitability. This narrow focus fails to account for a range of equally critical factors, including technical certifications obtained through platforms such as Coursera, NPTEL, and Microsoft Learn; programming language proficiencies; participation in competitive events and hackathons; and professional visibility through platforms such as LinkedIn. Consequently, these systems frequently produce generic recommendations that do not reflect the unique strengths and aspirations of individual students.

Moreover, current systems lack the capacity to provide real-time recommendations and are not designed to scale across large student populations. They also fail to capture dynamic changes in student profiles, such as newly acquired certifications or evolving skill sets, resulting in outdated guidance. There is, therefore, a pressing need for an intelligent, multi-attribute, and scalable career recommendation system that can leverage contemporary machine learning techniques to bridge the gap between student profiles and industry expectations.

III. OBJECTIVES

The primary objectives of the proposed AI-Based Career Path Recommendation System are as follows:

1. To collect, aggregate, and systematically analyse multidimensional student profile data, including CGPA, technical skill sets, industry certifications, and LinkedIn profile attributes.
2. To apply appropriate machine learning algorithms, specifically Decision Tree and Random Forest classifiers, for career role prediction based on student profile features.
3. To identify and model latent patterns and correlations between student profile attributes and corresponding engineering career roles.
4. To develop a reliable prediction module capable of accurately forecasting suitable engineering career paths for new and unseen student profiles.
5. To deliver personalised and contextually relevant career recommendations that go beyond academic metrics to encompass a holistic student profile.
6. To empower students with data-driven insights that facilitate well-informed, confident career decisions aligned with both personal strengths and market demand.

IV. LITERATURE REVIEW (SUMMARY)

The application of machine learning and data mining techniques to career guidance and job recommendation has been an active area of research over the past decade. Numerous studies have investigated the effectiveness of various algorithmic approaches in predicting suitable career paths based on student and professional profiles.

Early work in this domain primarily leveraged classification algorithms such as Naïve Bayes, k-Nearest Neighbour (k-NN), and Support Vector Machines (SVM) applied to academic performance datasets. These studies demonstrated that CGPA and academic scores could serve as moderate predictors of career suitability; however, the predictive accuracy was constrained by the limited feature space considered. For instance, studies focusing exclusively on academic performance typically reported accuracy values in the range of 70–80%, highlighting the insufficiency of single-attribute models.

Subsequent research incorporated interest-based profiling and psychometric assessments into career recommendation frameworks. Systems developed using collaborative filtering and content-based filtering techniques demonstrated improved personalisation by capturing student preferences and aligning them with career profiles. However, these approaches were often limited by the cold-start problem, wherein insufficient user interaction data degraded recommendation quality for new students.

More recent investigations have explored ensemble learning methods, particularly Random Forest and Gradient Boosting classifiers, for multi-attribute career prediction. These studies consistently reported superior accuracy compared to single-classifier baselines, attributable to the ensemble's capacity to reduce variance and handle high-dimensional feature spaces. Research published in IEEE and Springer proceedings between 2022 and 2025 has further demonstrated the utility of neural network architectures, including multilayer perceptrons and long short-term memory (LSTM) networks, for sequential career path modelling.

A critical gap identified across the reviewed literature is the absence of systems that integrate professional networking data, such as LinkedIn profiles, alongside academic and skills-based attributes. Furthermore, most existing systems do not support real-time updates to student profiles, limiting their applicability in dynamic educational environments. The proposed system addresses these gaps by incorporating a comprehensive, multi-attribute student profile and leveraging robust ensemble classification to deliver accurate, personalised career recommendations.

V. PROPOSED SYSTEM

The proposed AI-Based Career Path Recommendation System is designed to overcome the limitations of existing career guidance tools by adopting a multi-attribute, machine learning-driven approach. The system accepts a comprehensive student profile as input, encompassing academic performance (CGPA), technical skill competencies, industry-recognised certifications, and professional attributes extracted from LinkedIn profiles. These diverse data sources collectively provide a holistic representation of a student's capabilities and professional orientation.

The core of the proposed system is a machine learning pipeline that processes the collected data through a series of well-defined stages: data preprocessing, feature selection, model training, and prediction. Two supervised classification algorithms—the Decision Tree classifier and the Random Forest classifier—are employed to learn the mapping between student profile attributes and corresponding engineering career roles. The Decision Tree model provides interpretable decision pathways, enabling students to understand the reasoning behind specific recommendations. The Random Forest model, being an ensemble of multiple decision trees, offers enhanced predictive accuracy and robustness against overfitting.

The output of the system is a ranked list of personalised career recommendations tailored to each student's unique profile. Unlike traditional systems that produce a single career suggestion, the proposed system generates multiple ranked options, enabling students to explore a spectrum of suitable career paths. The system is designed with scalability in mind and can be integrated with web or mobile platforms to serve large student populations. Its modular architecture further facilitates future enhancements, including the incorporation of deep learning models and real-time job market data feeds.

VI. SYSTEM ARCHITECTURE

The system architecture of the AI-Based Career Path Recommendation System is organised into six functionally distinct layers, each responsible for a specific stage of the data processing and recommendation pipeline. The architectural design ensures a logical, sequential flow of information from raw data ingestion to the delivery of personalised career recommendations.

A. Input Layer

The Input Layer serves as the primary data ingestion point. Student profile information is collected through a structured interface that captures the following attributes: CGPA (normalised on a 10-point scale), technical skill sets (e.g., Python, Java, Machine Learning, Web Development, Data Analysis), industry certifications (e.g., AWS Certified, Google Cloud Professional, Microsoft Azure Fundamentals), and LinkedIn profile data including connection count, endorsements, and listed skills. The input layer is designed to accommodate both manual data entry and automated data import through API integrations.

B. Preprocessing Module

The Preprocessing Module performs essential data quality operations to ensure the integrity and consistency of the input data. This includes handling missing values through mean or mode imputation, removing duplicate records, and normalising continuous numerical features such as CGPA to a standardised scale. Categorical features, including skill sets and certifications, are encoded using one-hot encoding or label encoding techniques to render them suitable for consumption by ML algorithms.

C. Feature Selection Module

The Feature Selection Module identifies and retains the most informative attributes from the preprocessed dataset, thereby reducing dimensionality and mitigating the risk of model overfitting. Techniques such as correlation analysis, information gain calculation, and feature importance scores derived from the Random Forest model are employed to rank and select the optimal feature subset. This module ensures that only the most discriminative features are forwarded to the model training stage.

D. Model Training Module

The Model Training Module constitutes the computational core of the system. The preprocessed and feature-selected dataset is partitioned into training and testing subsets using an 80:20 split ratio. The Decision Tree classifier is trained using the CART (Classification and Regression Trees) algorithm, with pruning applied to prevent overfitting. The Random Forest classifier is trained as an ensemble of 100 decision trees, with bootstrap aggregation (bagging) employed to enhance generalisation. Both models are evaluated using accuracy, precision, recall, and F1-score metrics.

E. Prediction Module

Upon receiving a new student profile as input, the Prediction Module applies the trained ML models to generate career role predictions. The Random Forest classifier is used as the primary prediction model due to its superior accuracy, while the Decision Tree model is retained for explainability purposes. The prediction output comprises a ranked list of suitable career roles, each associated with a confidence score that reflects the model's certainty regarding the recommendation.

F. Output Layer

The Output Layer presents the career recommendations to the student in a user-friendly format. Each recommended career role is accompanied by a brief description of the role, the key skills required, and the confidence score assigned by the prediction module. The output interface is designed to be intuitive and accessible, with provisions for future integration into web and mobile application environments.

VII. METHODOLOGY

The methodology adopted in this study follows a systematic, step-by-step approach encompassing data collection, preprocessing, feature engineering, model development, and evaluation. The workflow is described below.

Step 1: Data Collection

Student profile data is collected from engineering students through a structured questionnaire. The dataset includes demographic information, academic performance records (CGPA), self-reported technical skill sets, certifications obtained, and LinkedIn profile attributes. The dataset used for model training and evaluation comprises profiles from a diverse cohort of students across multiple engineering disciplines.

Step 2: Data Preprocessing

The collected raw data is subjected to a comprehensive preprocessing pipeline. Missing values are identified and imputed using appropriate statistical measures. Outliers in continuous variables such as CGPA are detected and treated using the interquartile range (IQR) method. Categorical variables are encoded using label encoding, and numerical features are normalised using min-max scaling to ensure uniform feature ranges.

Step 3: Feature Selection

Feature selection is performed using a combination of filter and embedded methods. The Pearson correlation coefficient is computed to identify and remove highly correlated features that contribute redundant information. Feature importance scores are derived from an initial Random Forest model to rank features by their discriminative power. The top-ranked features are retained for model training.

Step 4: Model Training and Validation

The curated dataset is split into training (80%) and testing (20%) subsets using stratified sampling to preserve class distribution. The Decision Tree classifier is trained with a maximum depth constraint and minimum samples per leaf to prevent overfitting. The Random Forest classifier is trained with an ensemble of 100 estimators, utilising Gini impurity as the splitting criterion. Five-fold cross-validation is applied to both models to obtain robust performance estimates.

Step 5: Career Role Prediction

For a given new student profile, the trained Random Forest model generates a probability distribution over the set of possible career roles. The career role with the highest probability is designated as the primary recommendation, while roles with probabilities above a predefined threshold are included as secondary recommendations. This multi-output approach ensures that students receive a comprehensive set of career options rather than a single prescriptive suggestion.

Step 6: Recommendation Output and Display

The prediction results are formatted and presented to the student through the output interface. The recommended career roles are displayed in descending order of confidence score, accompanied by role descriptions and skill alignment summaries. This output assists students in understanding not only which career is recommended but also the rationale behind the recommendation.

VIII. RESULTS AND DISCUSSION

The proposed AI-Based Career Path Recommendation System was evaluated on a dataset comprising student profiles collected from engineering students enrolled across multiple disciplines. The dataset was partitioned into training and testing subsets in an 80:20 ratio, and model performance was assessed using standard classification metrics including accuracy, precision, recall, and F1-score.

The Random Forest classifier achieved a prediction accuracy of approximately 90–95% on the test dataset, demonstrating the model's strong capacity to generalise across unseen student profiles. The Decision Tree classifier yielded a comparatively lower accuracy of approximately 82–88%, consistent with its known susceptibility to overfitting on high-dimensional datasets when compared to ensemble methods. The performance differential between the two models underscores the advantage of ensemble learning in handling the diverse and multi-attribute nature of student profile data.

Analysis of the prediction results revealed that students with a CGPA above 8.0 combined with certifications in relevant technical domains were predominantly recommended for roles in software development, machine learning engineering, and data science. Students with strong skills in networking and cloud computing, irrespective of CGPA, were consistently recommended for roles in cloud architecture and DevOps engineering, illustrating the system's capacity to leverage skill-based attributes as primary determinants of career suitability.

The system also demonstrated that LinkedIn profile attributes, including the number of endorsed skills and professional connections, served as valuable supplementary features that improved recommendation precision, particularly for students with similar CGPA and certification profiles. This finding reinforces the importance of incorporating professional networking data in career recommendation frameworks.

Comparative evaluation against a baseline system that used CGPA as the sole predictor confirmed that the proposed multi-attribute approach yields a statistically significant improvement in prediction accuracy. The proposed system outperformed the baseline by approximately 15–20 percentage points, validating the hypothesis that a holistic, multi-attribute model is superior to single-metric approaches for career path prediction.

IX. ADVANTAGES

- Holistic Profile Analysis: The system considers multiple student attributes including CGPA, skills, certifications, and LinkedIn data, providing a comprehensive basis for career recommendations.
- High Predictive Accuracy: The Random Forest ensemble classifier achieves 90–95% prediction accuracy, significantly outperforming traditional single-metric approaches.
- Personalised Recommendations: Career suggestions are tailored to individual student profiles, ensuring relevance and specificity in the guidance provided.
- Scalability: The modular architecture of the system supports deployment across large student populations and can be readily integrated into institutional information systems.
- Explainability: The inclusion of a Decision Tree model alongside the Random Forest classifier provides interpretable decision pathways, enabling students to understand the rationale behind their recommendations.
- Data-Driven Decision Support: The system empowers students to make evidence-based career decisions, reducing reliance on subjective counsellor opinions and anecdotal guidance.

X . LIMITATIONS

Despite the promising results, the proposed system is subject to several limitations that warrant acknowledgment. First, the system's performance is contingent upon the quality and completeness of the input data. Student profiles with missing or inaccurate information may yield suboptimal recommendations. Robust data validation mechanisms are, therefore, essential for reliable system operation.

Second, the current implementation does not support real-time updates to student profiles. As students acquire new skills and certifications over the course of their studies, the system does not automatically incorporate these changes, potentially resulting in outdated recommendations.

Third, the dataset employed for model training is limited in scope, having been collected from a specific institutional context. Consequently, the generalisability of the trained models to students from different institutions, regions, or educational systems requires further validation.

Fourth, the system does not currently account for dynamic changes in job market demand. Career role relevance and industry requirements evolve continuously, and the absence of a real-time job market data feed may limit the system's responsiveness to emerging career trends.

XI. FUTURE SCOPE

Several avenues for future enhancement of the proposed system have been identified. First, the integration of advanced machine learning models, including deep neural networks, convolutional neural networks (CNNs) for feature extraction from unstructured profile data, and transformer-based models for natural language processing of LinkedIn summaries and resumes, is expected to yield further improvements in prediction accuracy.

Second, the development of a fully functional web and mobile application will significantly enhance the accessibility and usability of the system for students and institutional counsellors. The application is envisioned to support real-time profile updates, enabling the system to generate dynamic recommendations that reflect the student's evolving academic and professional profile.

Third, the incorporation of real-time job market data through APIs from platforms such as LinkedIn, Naukri, and Indeed will enable the system to align career recommendations with current industry demand, thereby improving the practical relevance of the guidance provided. Fourth, expanding the career role taxonomy to encompass emerging interdisciplinary domains such as quantum computing, biotechnology engineering, and digital humanities will broaden the system's applicability across a wider range of student interests and competencies. Finally, the implementation of user authentication, personalised dashboards, and longitudinal tracking of student career outcomes will transform the system into a comprehensive career development platform.

XII. CONCLUSION

This paper has presented an AI-Based Career Path Recommendation System that employs machine learning techniques to provide personalised, data-driven career guidance for engineering students. By integrating multidimensional student profile attributes—encompassing CGPA, technical skills, industry certifications, and LinkedIn profile data—the system addresses the critical shortcomings of conventional career guidance methods that rely predominantly on academic performance metrics.

The application of Decision Tree and Random Forest classifiers within a structured machine learning pipeline demonstrated prediction accuracies of 90–95%, representing a substantial improvement over traditional single-attribute approaches. The results confirm that ensemble learning methods are particularly well-suited to the multi-attribute, high-dimensional nature of student profile data.

The proposed system offers several significant contributions to the domain of educational technology and career guidance: it introduces a holistic, multi-attribute modelling framework; it delivers personalised and ranked career recommendations with associated confidence scores; and it provides an interpretable decision pathway through the Decision Tree component. These features collectively position the system as a robust and practical tool for academic career counselling.

The research establishes a strong foundation for the development of AI-driven career guidance platforms in academic institutions. Future work will focus on expanding the dataset, incorporating real-time job market data, and deploying the system as a scalable web and mobile application. The authors believe that the intelligent integration of AI and machine learning in academic career counselling has transformative potential to empower students to make confident, informed, and strategically sound career decisions.

REFERENCES

[1] A. Bharambe, S. Jain, and P. Kulkarni, "Career Recommendation System Using Machine Learning Techniques," in Proc. IEEE International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2022, pp. 1–6.

[2] R. Sharma, A. Gupta, and M. Singh, "A Survey on Career Prediction Systems Using Data Mining and Machine Learning," IEEE Access, vol. 10, pp. 45231–45245, 2022.

[3] P. Verma and S. Agarwal, "Random Forest-Based Job Role Prediction for Engineering Graduates," in Proc. International Conference on Computing, Communication and Security (ICCCS), 2023, pp. 1–7.

[4] K. Ramesh, T. Priya, and V. Suresh, "Multi-Attribute Student Career Counselling System Using Ensemble Learning," Journal of Engineering Education Transformations, vol. 36, no. 2, pp. 88–97, 2023.

[5] S. Das and B. Chakraborty, "Intelligent Career Guidance Using Decision Tree and Random Forest Classifiers," in Proc. IEEE Region 10 Symposium (TENSYMP), 2023, pp. 1–6.

[6] M. Ali and H. Khan, "LinkedIn-Enhanced Career Recommendation Using Hybrid Machine Learning Models," Expert Systems with Applications, vol. 215, pp. 119–128, 2024.

[7] N. Jayaraman and P. Krishnan, "Feature Selection Techniques for Student Performance Prediction: A Comparative Study," International Journal of Educational Technology in Higher Education, vol. 20, no. 1, pp. 1–18, 2024.

[8] A. Kumar and D. Mehta, "Deep Learning Approaches for Career Path Prediction in Higher Education," in Proc. International Joint Conference on Neural Networks (IJCNN), 2024, pp. 1–8.

[9] G. Prasad, S. Nair, and L. Menon, "Scalable Career Recommendation Architectures for Academic Institutions Using Cloud-Based ML Pipelines," IEEE Transactions on Learning Technologies, vol. 17, pp. 512–524, 2024.

[10] R. Anand, T. Selvi, and K. Mohan, "Real-Time Career Advisory Systems: Integrating Job Market APIs with ML-Based Student Profiling," in Proc. International Conference on Artificial Intelligence and Smart Systems (ICAIS), 2025, pp. 1–7.