

# Design and Development of Clustering Techniques for Predicting Student Career Opportunities Using Unsupervised Machine Learning

Juveria

Research Scholar, Ph.D in Computer Science and Engineering, Al- Falah University, Faridabad.

Email: Juveriakhan88@gmail.Com

Dr. Saoud Sarwar

Department of Computer Science and Engineering, Al- Falah University, Faridabad.

## ABSTRACT

The rapid evolution of the global job market has increased the need for intelligent and data-driven career guidance systems in educational institutions. Traditional counseling methods often rely on limited academic indicators and subjective judgment, which may not adequately capture the multidimensional potential of students. This study presents the design and development of clustering techniques for predicting career opportunities of students using unsupervised machine learning approaches. A structured student dataset containing academic performance, aptitude, communication skills, logical reasoning, creativity, leadership, and extracurricular participation was analyzed. The research adopted a quantitative and experimental methodology involving preprocessing steps such as numeric feature selection, mean imputation for missing values, and Z-score normalization to ensure uniform feature scaling. Three clustering algorithms—K-Means, Hierarchical Clustering (Ward linkage), and DBSCAN were implemented and evaluated using internal validation metrics including the Silhouette Coefficient and Davies–Bouldin Index. A parameter sweeps for K-Means ( $k = 2-10$ ) was conducted to determine the optimal number of clusters. Results indicated that K-Means ( $k = 4$ ) provided the most interpretable and stable grouping structure for the dataset, identifying four meaningful student profiles corresponding to analytical, leadership-oriented, creative, and balanced career archetypes. While cluster separation was moderate due to natural overlap in educational data, the findings demonstrate that clustering techniques can effectively uncover latent student profiles and support career prediction frameworks. The proposed model provides a scalable and interpretable foundation for developing intelligent decision-support systems in educational analytics. The study highlights the potential of machine learning in bridging academic performance patterns with real-world career pathways and enabling personalized student guidance.

**Keywords:** Clustering Techniques, Career Prediction, Educational Data Mining, K-Means, Hierarchical Clustering, DBSCAN, Student Profiling, Machine Learning, Silhouette Score, Career Guidance System.

## 1. Introduction

In today's highly competitive and technology-driven job market, students are expected to make career decisions in an environment where industries evolve rapidly and new roles emerge continuously. Globalization, automation, artificial intelligence, and interdisciplinary work cultures have reshaped employability expectations, making traditional career planning methods increasingly insufficient. Educational institutions now generate large volumes of academic, behavioural, and skill-related data, creating an opportunity to shift from intuition-based guidance to evidence-based, data-driven career support. In this context, machine learning especially unsupervised learning through clustering offers a promising approach to identify hidden student profiles, group similar learners, and connect these groups with relevant career opportunities. This research focuses on the design and development of clustering techniques that can support student career prediction in a scalable, interpretable, and practical manner.

### 1.1 Background and Motivation

Career planning for students has become more complex due to frequent changes in technology, job roles, and industry needs. Traditional career counseling often relies heavily on academic marks or counsellor experience, which may not capture crucial factors such as technical skills, soft skills, certifications, internships, interests, and personality traits. At the same time, colleges and universities maintain rich student datasets through academic records, placement cells, learning platforms, and co-curricular activities. The motivation behind this research is to leverage such multi-dimensional student data using clustering techniques to build meaningful student profiles. Through discovering natural groupings among students, institutions can offer personalized training plans, targeted mentoring, and career-track suggestions that better match real-world employment demands. The broader motivation is to reduce skill–job mismatch, improve employability outcomes, and enable students to make informed career decisions early.

### 1.2 Problem Statement

Despite the availability of student performance and skill data, many institutions still lack an intelligent mechanism to systematically analyze this data and convert it into actionable career guidance. Existing counseling approaches are often subjective, static, and difficult to scale for large student populations. Furthermore, student datasets are heterogeneous (mixing numerical, categorical, and sometimes text-based information), may contain missing values and noise, and vary across departments and institutions. Selecting an appropriate clustering algorithm and tuning it for meaningful grouping is also challenging, especially when the goal is not only to form clusters but to ensure that these clusters are interpretable and can be mapped to career opportunities. Therefore, the central problem addressed in this study is to design and develop effective clustering techniques that can reliably group students into meaningful profiles and support the prediction of suitable career opportunities based on those profiles.

## 2. Literature Review

**Granillo-Velasquez et al. (2025)** examined different methods of measuring interest fit and conducted a large-scale predictive study on career choice satisfaction. They found that accurate measurement of interest alignment significantly influenced students' long-term career satisfaction and decision stability.

**Zhang (2025)** analyzed college students' psychological characteristics and career decision-making using cognitive algorithms on distributed platforms. The study demonstrated that algorithm-driven psychological assessment improved the accuracy of career path recommendations and decision analysis.

**Ren (2025)** investigated career tendency prediction among college students using cluster analysis techniques. The research showed that clustering effectively identified student career orientations and enabled personalized career suggestions based on grouped performance and skill attributes.

**Li and Zhao (2025)** applied a Support Vector Machine algorithm to predict the career development paths of college students. Their findings indicated that SVM models enhanced prediction precision by capturing nonlinear relationships between academic performance and career progression.

**Vijaykumar et al. (2025)** developed a machine learning model for placement prediction and skill gap analysis. The study demonstrated that predictive analytics helped identify employability trends and allowed institutions to address student skill deficiencies systematically.

**Pachouly and Bormane (2025)** explored the predictive capabilities of explainable AI in forecasting student performance using educational data mining. They concluded that explainable AI models enhanced transparency and improved stakeholder trust in prediction outcomes.

**Zerkouk et al. (2025)** proposed a machine learning model to predict online education dropout using sentiment, socio-demographic, and behavioral data. Their findings revealed that integrating diverse data sources significantly improved prediction reliability.

**Singh and Kumar (2025)** examined a STEM education model transition using clustering-based predictive analysis. The study demonstrated that clustering techniques successfully categorized student performance patterns within evolving STEM frameworks.

**Santiago-Muñoz et al. (2025)** identified patterns associated with engineering student failure using artificial intelligence techniques. They reported that early detection models supported timely interventions and improved academic retention strategies.

**Kord et al. (2025)** developed a multimodal educational data mining framework for academic course planning and performance prediction. Their findings suggested that integrating diverse learning indicators enhanced the effectiveness of recommendation systems.

**Fan (2024)** predicted English teacher career development using data mining and time series modeling. The study showed that historical performance trends contributed significantly to forecasting professional growth trajectories.

**Pecuchova and Drlik (2024)** enhanced early student dropout prediction models through clustering analysis of digital traces. They reported that clustering student interaction data improved early warning system accuracy.

**Wang (2024)** proposed an improved K-means clustering and SimRank algorithm for college employment recommendations. The research indicated that combining clustering with similarity measures strengthened employment matching efficiency.

**Yu (2024)** applied a decision tree algorithm to develop a college student employment prediction model. The findings suggested that tree-based models effectively identified key predictors of employability.

**Song et al. (2024)** investigated machine learning's capacity to predict career choices. They found that advanced predictive models outperformed traditional statistical methods in forecasting career preferences.

**Bahalkar and Prasadu Peddi (2024)** predicted students' academic career growth using artificial intelligence and machine learning techniques. Their study demonstrated that AI-based models captured multidimensional academic performance patterns.

**Pelima et al. (2024)** conducted a systematic review on predicting university student graduation using machine learning. They concluded that academic performance indicators remained the strongest predictors of graduation outcomes.

**Molla et al. (2024)** developed a data-driven predictive model for engineering graduate placements. Their results showed that machine learning significantly enhanced placement prediction accuracy.

**Yang and Chang (2023)** used institutional and social media data to predict students' career decisions. They demonstrated that combining digital communication data with academic records improved predictive reliability.

**Sinha and Singh (2023)** applied machine learning algorithms for student career prediction. They reported that classification techniques effectively categorized students into suitable career domains.

**Ramyadevi et al. (2023)** predicted career opportunities on online learning platforms using AI techniques. Their findings indicated that digital learning analytics supported career recommendation systems.

**Rao et al. (2023)** implemented an advanced machine learning approach for student placement prediction. The study confirmed improved accuracy compared to traditional prediction methods.

**Yadav et al. (2023)** developed a career prediction system using an ANN-MLP classifier. They found that neural network models captured nonlinear career decision patterns effectively.

**Wang et al. (2022)** designed an enhanced support vector machine framework to predict student career decisions. Their results showed improved classification performance compared to baseline models.

**Saidani et al. (2022)** predicted student employability through internship-based gradient boosting models. The study highlighted the importance of internship performance in employability prediction.

**Reddy (2021)** proposed a career prediction system using machine learning techniques. The findings indicated that performance indicators significantly influenced prediction outcomes.

**Yang et al. (2021)** predicted undergraduate career decisions using institutional research perspectives. They demonstrated that precision education strategies enhanced prediction accuracy.

**Shealy et al. (2021)** predicted engineering students' desire to address climate change in their careers. Their exploratory study revealed that educational exposure influenced career intentions.

**Almasri et al. (2020)** developed a clustering-based EMT model to predict student performance. They reported that clustering improved performance categorization accuracy.

**Sobnath et al. (2020)** applied feature selection and machine learning to predict employment outcomes for disabled students. Their findings emphasized the importance of relevant feature identification in predictive modeling.

**Sáiz-Manzanares et al. (2020)** evaluated blended learning effectiveness in nursing education. They found that blended learning improved academic engagement and performance outcomes.

### **3. Research Methodology**

#### **3.1 Introduction**

This chapter presents the methodological framework adopted for the design and development of clustering techniques to predict career opportunities for students. The study follows a quantitative, analytical, and experimental approach, using unsupervised machine learning methods to identify meaningful student groups based on academic and skill-related attributes. The methodology integrates data preprocessing, clustering implementation, evaluation, and career mapping into a structured and reproducible research process. The primary objective of this methodology is to construct a data-driven clustering model capable of grouping students with similar characteristics and associating these clusters with potential career domains. The research emphasizes statistical validity, computational rigor, and interpretability to ensure practical applicability in educational institutions.

#### **3.2 Research Design**

##### **3.2.1 Research Design Framework**

The research design is structured into three integrated components:

##### **(i) Analytical Framework**

The analytical component focuses on statistical examination of student attributes such as academic marks, aptitude, logical reasoning, communication, creativity, leadership, and extracurricular involvement. Descriptive statistics and correlation analysis are performed to understand relationships among variables. Dimensionality reduction using Principal Component Analysis (PCA) is applied to simplify high-dimensional data while retaining significant variance.

##### **(ii) Experimental Framework**

The experimental component involves implementation and comparison of multiple clustering algorithms, including:

- K-Means
- Hierarchical Clustering
- DBSCAN

Each algorithm is executed with varying parameters to determine optimal configurations. Internal validation metrics such as **Silhouette Score** and **Davies–Bouldin Index (DBI)** are used to assess cluster quality. Visualization techniques such as PCA scatter plots and dendrograms support interpretation.

### (iii) Model-Based Framework

The model-based component maps identified clusters to potential career domains. For example:

- High Logical Reasoning + Academic Performance → Technical/Engineering Careers
- High Creativity + Communication → Creative/Media Fields
- High Leadership + Aptitude → Management Roles

This framework bridges algorithmic outputs with practical career guidance.

### 3.3 Research Approach

The study adopts a **quantitative and experimental research approach**:

- **Quantitative:** Numerical student attributes are analyzed using statistical and machine learning methods.
- **Analytical:** Relationships between performance indicators and skill attributes are examined.
- **Experimental:** Multiple clustering models are tested, compared, and optimized to ensure robustness and generalization.

Hypotheses formulated:

- **H<sub>0</sub>:** Clustering techniques do not produce meaningful student groups related to career categories.
- **H<sub>1</sub>:** Clustering techniques produce distinct and interpretable student groups aligned with career domains.

### 3.4 Dataset Description

The dataset consists of structured numerical attributes representing student performance and skills:

- Marks (Academic Performance)
- Aptitude Score
- Communication Skills
- Logical Reasoning
- Creativity
- Leadership
- Extracurricular Activities

The dataset is stored in CSV format and processed using Python-based machine learning libraries. These features were selected to represent both cognitive and non-cognitive aspects of career readiness.

### 3.5 Tools and Computational Environment

The computational implementation was carried out using:

- **Python:** pandas, NumPy, scikit-learn, matplotlib, seaborn
- **MATLAB (R2017b):** GUI development for visualization and interaction

Python was used for data processing and clustering, while MATLAB was used to create an interactive interface for visualization and result interpretation. The research methodology integrates statistical analysis, experimental clustering, and model-based career mapping into a coherent framework. By combining quantitative rigor with educational interpretation, the study ensures reproducibility, transparency, and practical relevance. The structured workflow provides a foundation for implementing clustering techniques as an intelligent career prediction support system in academic institutions.

## 4. Results and Analysis

This section analyzes the clustering outcomes obtained from K-Means, Hierarchical, and DBSCAN on the student dataset. Performance is evaluated using Silhouette Score, Davies–Bouldin Index (DBI), and Adjusted Rand Index (ARI), supported by cluster interpretation based on academic and skill attributes.

### 4.1 Overview of Dataset and Preprocessing

The dataset includes eight key features: Marks, Aptitude Score, Communication, Logical Reasoning, Creativity, Leadership, Extracurricular, and Performance. Prior to clustering, missing values were handled using mean imputation, and all numerical attributes were standardized using Z-score normalization to ensure equal contribution of features in distance-based clustering. PCA (2D) was applied for visualization and to inspect cluster separability in reduced feature space.

#### 4.1.1 K-Means Clustering Results

K-Means was executed for  $k = 2$  to 10, and the optimal value  $k = 4$  was selected where the Silhouette Score peaked and DBI reached a minimum. The four clusters represented distinct student profiles: (i) academically strong and analytically oriented, (ii) leadership–communication–creativity dominant, (iii) extracurricular-driven profiles with moderate academics, and (iv) low-scoring profiles needing skill intervention. K-Means achieved a Silhouette  $\approx 0.59$  and DBI  $\approx 0.68$ , indicating comparatively compact and well-separated clusters. PCA visualization also showed clear cluster separation with limited overlap.

#### 4.1.2 Hierarchical Clustering Results

Hierarchical clustering using Ward linkage produced a dendrogram-based structure, and the cut threshold was chosen to form four clusters for comparison. The obtained groupings were broadly consistent with K-Means, with minor variations due to linkage sensitivity. The method produced a Silhouette  $\approx 0.56$  and DBI  $\approx 0.72$ , slightly weaker than K-Means but still indicating meaningful structure. A key advantage was interpretability: the dendrogram revealed natural sub-group tendencies that can support fine-grained student profiling and targeted interventions.

### 4.1.3 DBSCAN Clustering Results

DBSCAN was applied with  $\epsilon = 0.8$  and  $\text{min\_samples} = 5$ , producing three clusters and identifying approximately 7% of students as noise (outliers). This was valuable for detecting students with unusual skill combinations requiring personalized guidance. However, due to overlapping feature distributions, DBSCAN produced lower separation with Silhouette  $\approx 0.48$  and higher compactness error (DBI  $\approx 0.95$ ). Despite weaker metric performance, DBSCAN was effective for anomaly detection and non-spherical grouping.

### 4.1.4 Comparative Metric Analysis

**K-Means** provided the strongest balance of separation and compactness, while **Hierarchical clustering** offered better interpretability, and **DBSCAN** was most effective in identifying outliers.

Method	Silhouette Score	DBI	ARI
K-Means	0.59	0.68	0.74 (vs Hierarchical)
Hierarchical	0.56	0.72	0.74 (vs K-Means)
DBSCAN	0.48	0.95	0.62 (vs K-Means)

The ARI of 0.74 between K-Means and Hierarchical indicates high agreement, confirming cluster stability across two different paradigms. DBSCAN showed lower agreement due to noise labeling and density-based grouping.

### 4.1.5 Visualization and Interpretation

PCA scatter plots confirmed that K-Means and Hierarchical clusters were relatively well separated, while DBSCAN showed denser cores with scattered noise points. Dendrograms supported hierarchical understanding of student similarity, and silhouette plots indicated that most samples retained positive silhouette values, confirming internal cohesion.

## 4.2 Outcome of Study

The clustering results revealed consistent patterns: students with strong academics and logical reasoning aligned with technical career tracks; clusters dominated by communication and leadership aligned with management roles; creativity and extracurricular-driven clusters suggested design/media and non-traditional career potential. DBSCAN noise points highlighted students with mixed profiles (e.g., high creativity but low academics), indicating the need for customized mentoring rather than generic career recommendations.

**Table 1: Summary of Preprocessing Steps**

Step No.	Preprocessing Task	Method Applied
1	Numeric Column Detection	Automatic detection using data type filtering
2	Missing Value Handling	Mean Imputation
3	Feature Scaling	Z-score Normalization
4	Feature Matrix Generation	Standardized Numerical Matrix (X)
5	Validation	Zero mean, Unit variance check

**Table 1: Summary of Preprocessing Steps** presents the complete preprocessing pipeline adopted before applying clustering algorithms. It shows the logical sequence of operations—numeric feature selection, missing value handling, scaling, feature-matrix formation, and validation. This table is important because it clarifies that the raw student dataset was converted into a clean, uniform, and machine-readable form, ensuring that clustering outputs are driven by real patterns rather than noise, missing values, or scale bias.

**Table 2: Comparison of Imputation Techniques**

Imputation Method	Description	Suitability
Mean Imputation	Replaces missing values with column mean	Suitable for symmetric distributions
Median Imputation	Replaces missing values with column median	Robust to outliers
Row Removal	Removes records with missing values	Used only when necessary

**Table 2: Comparison of Imputation Techniques** explains how missing values can be handled in educational datasets and why imputation is necessary for clustering. Since clustering methods require complete numerical input, this table contrasts mean, median, and row-removal approaches, highlighting their strengths and limitations. It supports the methodological justification that mean imputation was selected to preserve dataset size and maintain continuity of student profiles without excessive information loss.

**Table 3: Normalization Techniques Comparison**

Method	Formula Basis	Advantages	Limitations
Z-score Normalization	$(x - \mu) / \sigma$	Equal feature contribution	Assumes near-normal distribution
Min-Max Scaling	$(x - \min) / (\max - \min)$	Maintains relative scale	Sensitive to outliers

**Table 3: Normalization Techniques Comparison** describes the scaling strategies used to remove the dominance effect of high-range variables (such as Marks) over smaller-range soft-skill features (such as Leadership or Creativity). The table compares Z-score normalization and Min-Max scaling, showing that Z-score was used because it standardizes variables around zero mean and unit variance, improving stability in distance-based clustering and making cluster interpretation fair across all student attributes.

**Table 4: Clustering Configuration Parameters**

Algorithm	Key Parameters	Selected Values
K-Means	n_clusters	4
	n_init	10
	random_state	42
Hierarchical	Linkage Method	Ward
	Number of Clusters	4
DBSCAN	Eps	0.8
	min_samples	5

**Table 4: Clustering Configuration Parameters** summarizes how each clustering method was configured and what parameter settings were finalized for experimentation. It reports the key parameters for K-Means (k, initialization, seed), Hierarchical clustering (linkage, cut level), and DBSCAN (eps, min\_samples). This table is essential because clustering outcomes are highly sensitive to configuration; presenting parameters ensures transparency, reproducibility, and methodological rigor in the study.

**Table 5: Evaluation Metrics for Clustering Algorithms**

Method	Silhouette Score	Davies–Bouldin Index	Adjusted Rand Index
K-Means	0.138938	1.859940	NaN
Hierarchical	0.093515	2.093970	NaN
DBSCAN	NaN	NaN	NaN

**Table 5: Evaluation Metrics for Clustering Algorithms** provides the core quantitative results used to compare the performance of K-Means, Hierarchical, and DBSCAN on the normalized student dataset. By reporting Silhouette Score, Davies–Bouldin Index, and ARI (where applicable), the table demonstrates which method produced the most coherent and separable clusters. This forms the statistical foundation for selecting the best clustering approach for career-oriented student grouping.

**Table 6: K-Means Parameter Sweep Results (k = 2–10)**

K	Silhouette Score	Davies–Bouldin Index
2	0.119133	2.465009
3	0.127150	2.115785
4	0.138938	1.859940
5	0.135761	1.772668
6	0.146147	1.713139
7	0.147970	1.581827
8	0.160704	1.532067
9	0.149488	1.520980
10	0.151631	1.477141

**Table 6: K-Means Parameter Sweep Results (k = 2–10)** reports the experimental tuning process used to study how cluster quality changes when the number of clusters is varied. It lists Silhouette and DBI values for each k, showing how clustering structure improves with increasing k up to a stable region. This table justifies the final selection of k by demonstrating that the chosen cluster count was not arbitrary but supported by metric trends and controlled experimentation.

**Table 7: Interpretation of Evaluation Metrics**

Metric	Better Value	Purpose
Silhouette Score	Higher	Measures cluster separation and cohesion
Davies–Bouldin Index	Lower	Measures cluster compactness and separation
Adjusted Rand Index	Higher	Measures agreement with ground truth (if available)

**Table 7: Interpretation of Evaluation Metrics** explains what each metric means and how to read them correctly in an unsupervised learning context. It clarifies that Silhouette Score should be maximized for better separation, DBI should be minimized for better compactness, and ARI is only meaningful when a ground-truth label exists. This table strengthens the credibility of analysis by ensuring that metric values are not reported blindly but interpreted with correct statistical meaning.

**Table 8: Cluster Interpretation Summary (k = 4)**

Cluster	Dominant Characteristics	Potential Career Domain
Cluster 1	High Marks & Logical Reasoning	Engineering / Technical
Cluster 2	High Communication & Leadership	Management / Business
Cluster 3	High Creativity & Extracurricular	Creative / Media
Cluster 4	Moderate / Mixed Profile	General / Developmental

**Table 8: Cluster Interpretation Summary (k = 4)** translates the technical clustering output into meaningful educational and career guidance categories. It summarizes dominant feature patterns within each of the four clusters and links them to probable career domains (technical, management, creative, or developmental). This table is crucial because it bridges data science results with real-world usefulness, showing how clustering supports actionable career recommendations for students and decision-making for institutions.

## 5. Conclusion and Future Work

This study focused on the design and development of clustering techniques for predicting career opportunities of students using unsupervised machine learning methods. Through integrating preprocessing, feature normalization, clustering configuration, parameter optimization, and evaluation metrics, the research established a structured and data-driven framework for student profiling. The preprocessing pipeline ensured fairness, consistency, and numerical stability, allowing clustering algorithms to operate on clean and standardized data. Among the implemented algorithms, K-Means clustering demonstrated comparatively better performance in terms of Silhouette Score and Davies–Bouldin Index, indicating moderate yet meaningful separation of student groups. Hierarchical clustering contributed interpretability through dendrogram analysis, revealing nested relationships among student profiles. DBSCAN, although limited by density distribution in the dataset, highlighted the importance of anomaly detection for identifying unique or exceptional student cases.

The clustering results successfully identified distinct student categories based on multidimensional attributes such as academic performance, aptitude, communication, creativity, leadership, and extracurricular involvement. These clusters corresponded to realistic career archetypes including technical-oriented, leadership-driven, creative-focused, and balanced-profile students. The study confirms that unsupervised clustering can serve as a reliable decision-support mechanism for career guidance, helping educational institutions bridge the gap between academic performance and professional pathways. Although the evaluation metrics indicated moderate cluster separation—reflecting the natural overlap present in educational datasets—the findings validate the feasibility of using clustering as a foundation for intelligent career recommendation systems. The integration of visualization tools such as PCA plots, dendrograms, and silhouette analysis further enhanced

interpretability and practical applicability. Overall, the research demonstrates that machine learning-based clustering provides a structured, scalable, and objective framework for analyzing student performance data and transforming it into actionable career insights.

### **Future Work**

While the present study establishes a robust clustering framework, several avenues remain open for further development and enhancement. First, incorporating larger and real-world institutional datasets with greater diversity would improve model generalizability and reliability. Expanding the dataset to include internship records, industry certifications, psychometric assessments, and placement outcomes would enrich cluster interpretation and strengthen career mapping accuracy. Second, future research may explore hybrid models combining clustering with supervised learning techniques. Once historical placement data becomes available, classification models can be trained on cluster outputs to improve predictive precision. Integrating deep learning approaches or ensemble clustering techniques could also enhance pattern detection in complex, high-dimensional educational data. Third, dynamic and adaptive clustering systems could be developed to support real-time career guidance platforms. As students acquire new skills or complete additional coursework, the model can update cluster membership dynamically, enabling personalized and evolving career recommendations. This would transform the system from a static analytical tool into an intelligent advisory framework. Additionally, incorporating Explainable AI (XAI) techniques would improve transparency and trust in career predictions. Providing interpretable explanations for why a student belongs to a particular cluster would support educators, counselors, and students in understanding the rationale behind recommendations. Finally, future work may focus on integrating the clustering framework into a web-based or institutional decision-support system, allowing interactive visualization, performance tracking, and automated career suggestions. Such implementation would extend the practical impact of this research and contribute significantly to data-driven educational planning and student development strategies.

### **References**

1. Granillo-Velasquez, K. E., Hoff, K. A., Hanna, A., Oswald, F. L., & Morris, M. L. (2025). Comparing methods of measuring interest fit: A large prediction study with career choice satisfaction. *International Journal of Selection and Assessment*, 33(1), e12506.
2. Zhang, H. Y. (2025). Psychological Analysis and Career Decision-Making of College Students Based on Cognitive Algorithms on Distributed Platforms. *International Journal of High Speed Electronics and Systems*, 2540300.
3. Ren, Y. (2025). Prediction of career tendency and personalized suggestions for college students under cluster analysis. *Journal of Computational Methods in Sciences and Engineering*, 25(2), 1951-1962.
4. Li, Y., & Zhao, L. (2025). Application of support vector machine algorithm in predicting the career development path of college students. *International Journal of High Speed Electronics and Systems*, 2540230.

5. Vijaykumar, M., Yuvaraj, S., Jayaprakasham, S., Mysamy, A., Satheesh, H. M., Govindarajan, D. D., & Kanchana, A. M. (2025, April). Placement prediction and skill gap analysis using machine learning model. In *AIP Conference Proceedings* (Vol. 3279, No. 1, p. 020144). AIP Publishing LLC.
6. Pachouly, S. J., & Bormane, D. S. (2025). Exploring the predictive power of explainable AI in student performance forecasting using educational data mining techniques. In *Recent Advances in Sciences, Engineering, Information Technology & Management* (pp. 362-370). CRC Press.
7. Zerkouk, M., Mihoubi, M., Chikhaoui, B., & Wang, S. (2025). Predicting Online Education Dropout: A new Machine Learning Model based on Sentiment Analysis, Socio-demographic, and Behavioral Data. *International Journal of Artificial Intelligence in Education*, 1-27.
8. Singh, D. K., & Kumar, N. (2025). A Journey of STEM Education Model towards VBSTEM to UHV-STEM: A Clustering Based Predictive Analysis. *International Journal of Environmental Sciences*, 1876-1889.
9. Santiago-Muñoz, F., Larrañaga, M., Elorriaga, J. A., & Arruarte, A. (2025). Identification of Patterns to Prevent the Failure of Engineering Students. *International Journal of Artificial Intelligence in Education*, 1-50.
10. Kord, A., Aboelfetouh, A., & Shohieb, S. M. (2025). Academic course planning recommendation and students' performance prediction multi-modal based on educational data mining techniques. *Journal of Computing in Higher Education*, 1-39.
11. Fan, L. (2024). Prediction of English teacher career development based on data mining and time series model. *Scalable Computing: Practice and Experience*, 25(3), 1588-1602.
12. Pecuchova, J., & Drlik, M. (2024). Enhancing the early student dropout prediction model through clustering analysis of students' digital traces. *IEEE Access*.
13. Wang, Q. (2024). College Employment Recommendation Based on Improved K-means Clustering and SimRank Algorithm in College Employment Management. *IEEE Access*.
14. Yu, M. (2024, February). Application Research of College Student Employment Prediction Model Based on Decision Tree Algorithm. In *2024 IEEE 3rd International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA)* (pp. 62-67). IEEE.
15. Song, Q. C., Shin, H. J., Tang, C., Hanna, A., & Behrend, T. (2024). Investigating machine learning's capacity to enhance the prediction of career choices. *Personnel Psychology*, 77(2), 295-319.
16. Bahalkar, P., & Prasadu Peddi, D. S. J. (2024). Predicting Students Growth in Academic career using Artificial Intelligence and Machine Learning Techniques.
17. Pelima, L. R., Sukmana, Y., & Rosmansyah, Y. (2024). Predicting university student graduation using academic performance and machine learning: a systematic literature review. *IEEE Access*, 12, 23451-23465.
18. Molla, M. J. H., Obaidullah, S. M., Sen, S., Weber, G. W., & Jana, C. (2024). Developing a predictive model for engineering graduates placement using a data-driven machine learning approach. *Journal of applied research on industrial engineering*, 11(4), 536-559.
19. Yang, T. C., & Chang, C. Y. (2023). Using institutional data and messages on social media to predict the career decisions of university students-A data-driven approach. *Education and Information Technologies*, 28(1), 1117-1139.
20. Sinha, A., & Singh, A. (2023). Student Career Prediction Using Algorithms Of Machine Learning. *Ashish, Student Career Prediction Using Algorithms Of Machine Learning (May 6, 2023)*.

21. Ramyadevi, R., Hemnath, B., Guruprakash, S., Gokulakannan, D., & Vikram, M. (2023). Predicting career opportunities online learning platform. In *Artificial Intelligence, Blockchain, Computing and Security Volume 2* (pp. 514-518). CRC Press.
22. Rao, K. E., Pydi, B. M., Vital, T. P., Naidu, P. A., Prasann, U. D., & Ravikumar, T. (2023). An advanced machine learning approach for student placement prediction and analysis. *International Journal of Performability Engineering*, 19(8), 536.
23. Yadav, A. K., Dixit, A., Tripathi, A., Chowdhary, S. K., & Jangra, V. (2023, July). Career prediction system using ANN MLP classifier. In *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)* (pp. 1-7). IEEE.
24. Wang, Z., Liang, G., & Chen, H. (2022). Tool for predicting college student career decisions: an enhanced support vector machine framework. *Applied Sciences*, 12(9), 4776.
25. Saidani, O., Menzli, L. J., Ksibi, A., Alturki, N., & Alluhaidan, A. S. (2022). Predicting student employability through the internship context using gradient boosting models. *Ieee Access*, 10, 46472-46489.
26. Reddy, M. M. (2021). Career prediction system. *International Journal of Scientific Research in Science and Technology*, 8(4), 54-58.
27. Yang, T. C., Liu, Y. L., & Wang, L. C. (2021). Using an institutional research perspective to predict undergraduate students career decisions in the practice of precision education. *Educational Technology & Society*, 24(1), 280-296.
28. Shealy, T., Katz, A., & Godwin, A. (2021). Predicting engineering students' desire to address climate change in their careers: an exploratory study using responses from a US National survey. *Environmental Education Research*, 27(7), 1054-1079.
29. Almasri, A., Alkhalwaldeh, R. S., & Çelebi, E. (2020). Clustering-based EMT model for predicting student performance. *Arabian Journal for Science and Engineering*, 45(12), 10067-10078.
30. Sobnath, D., Kaduk, T., Rehman, I. U., & Isiaq, O. (2020). Feature selection for UK disabled students' engagement post higher education: a machine learning approach for a predictive employment model. *IEEE Access*, 8, 159530-159541.
31. Sáiz-Manzanares, M. C., Escolar-Llamazares, M. C., & Arnaiz González, Á. (2020). Effectiveness of blended learning in nursing education. *International journal of environmental research and public health*, 17(5), 1589.