IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, October 2026
Impact Factor: 8.2

Student Career Guidance System Using Machine Learning

Sultana Hasan Sayyad and Ms. Asha Mane
Sadhu Vaswani Institute of Management Studies for Girls, Pune
Savitribai Phule Pune University, Pune
sultanasayyad2095@gmail.com and ashamane90@gmail.com

Abstract: *Career selection is an important decision for students, as their academic performance, skills, and employability attributes can influence their future career opportunities. This research presents a Student Career Guidance System using Machine Learning for predicting suitable career domains from student-related information. A secondary dataset containing 50,000 student records was used for the study, including academic, skill, and employability-related attributes. The dataset was preprocessed and different Machine Learning classification techniques, namely Logistic Regression, Decision Tree, Random Forest, and Gradient Boosting, were applied and compared. The performance of the models was evaluated using accuracy, precision, recall, and F1-score. Among the evaluated models, Random Forest achieved the highest accuracy of 70.49% and was selected as the final model. The proposed system can provide data-driven career guidance by predicting suitable career domains and can support students in making informed career-related decisions based on their available information.*

Keywords: Career Domain, Career Prediction, Machine Learning, Student Career Guidance

I. INTRODUCTION

A suitable career choice is the foundation on which every student's educational and professional journey is built. A student's career choice is influenced by various factors such as their education, skills, interests, and academic record. Since students today have a wide range of career options and making right and accurate career decision has become a major challenge. To solve this problem and to give proper direction to students as per their ability an effective career guidance system is needed**.**

Machine Learning offers a data-driven and effective solution to these challenges by analyzing large datasets of students. This technology analyzes students' academic records, key skills, and employability attributes to find hidden patterns. Various advanced Machine Learning algorithms, like Logistic Regression, Decision Tree, Random Forest and Gradient Boosting can be used to accurately predict and recommend the suitable career path. Therefore it is feasible to develop an automatic and data-driven career guidance system, which can provide accurate and personalized suggestions to each student.

Recently, Machine Learning techniques have been widely used to help students choosing suitable career paths and to improve career guidance systems. The algorithms used for student career prediction are Decision Tree, SVM, AdaBoost, and Random Forest, in which the Random Forest algorithm yielded the best performance with an accuracy of about 93%. Furthermore, investigated a large-scale dataset of campus behavior using the ACCBOX and XGBoost methods to predict career choice, which achieved an accuracy of approximately 63.8%.

Reviewing the above studies, we find that many existing studies are based on small, manually collected or domain-specific datasets. In addition, some studies use a small number of student attributes and have limited validation on large and diverse student datasets. Thus, there is a need for a Machine Learning-based career guidance system able to leverage a large and diverse student dataset with academic, skill and employability attributes to provide effective and data-driven career domain recommendations.

Copyright to IJARSCT
www.ijarsct.co.in
DOI: 10.48175/IJARSCT-38843

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, October 2026
Impact Factor: 8.2

To address the limitations of previous studies, this study proposes a Machine Learning-based Student Career Guidance System by using a comprehensive secondary dataset obtained from Kaggle, which contains 50,000 student records. The dataset includes important attributes related to students' academic performance, skills and employability which are used to train and compare different Machine Learning algorithms. The Random Forest model was the best performing model among the evaluated models with an accuracy of 70.49%. The proposed system can assist students in identifying suitable career domains and make informed career decisions based on their academic performance, skills and employability attributes.

OBJECTIVES OF THE STUDY

The main objectives of this study are:

1. To develop a Machine Learning-based Student Career Guidance System that can predict appropriate career domains for students.
2. To use students' academic performance, skills, and employability-related attributes as input variables for career domain prediction.
3. To train and evaluate four machine learning algorithms, specifically Logistic Regression, Decision Tree, Random Forest, and Gradient Boosting.
4. To measure the performance of each developed model using evaluation metrics including accuracy, precision, recall, and F1-score.
5. To determine and select the best-performing machine learning model for incorporation into the proposed Student Career Guidance System.

II. LITERATURE REVIEW

Machine Learning has become an important approach for student career prediction and career guidance. Previous studies have used different Machine Learning algorithms to predict suitable career choices for students based on their academic performance, skills, interests, and other relevant attributes. Researchers have used various techniques such as Random Forest, Support Vector Machine, Decision Tree, K-Nearest Neighbors, and XGBoost for career prediction and recommendation. The reviewed studies lay a foundation for developing an effective, data-driven Student Career Guidance System.

VidyaShreeram and Muthukumaravel (2021) investigated the use of Machine Learning techniques for student career prediction . The study evaluated algorithms such as Decision Tree, Random Forest, Support Vector Machine, and AdaBoost for predicting suitable career choices. Among the evaluated models, Random Forest achieved the highest performance with an accuracy of approximately 93%. The study demonstrated that Machine Learning can be effectively used to support student career prediction.[2]

Nie et al. (2020) proposed a career choice prediction approach based on large-scale campus behavioral data. The study used the ACCBOX approach and compared its performance with algorithms such as Random Forest, Support Vector Machine, and Decision Tree. The proposed approach achieved an accuracy of approximately 63.8%, demonstrating the potential of student behavioral data for career choice prediction. The study also highlighted the importance of analyzing large-scale student data for understanding career-related behavior.[1]

Wang et al. (2022) developed an enhanced Support Vector Machine framework for predicting college students' career decisions. The proposed CBBOA-SVM-FS framework incorporated feature selection to improve the prediction performance of the model. The study reported an accuracy of 94.20%, showing the effectiveness of the enhanced SVM approach. The findings indicate that feature selection and optimization techniques can improve Machine Learning-based career prediction.[3]

El-Keiey et al. (2025) studied career recommendation for undergraduate students using different Machine Learning techniques. The researchers evaluated algorithms including KNN, Naive Bayes, SVM, Decision Tree, Random Forest, and Gradient Boosting. Random Forest achieved strong performance, with an accuracy of 90% on the evaluated student

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, October 2026
Impact Factor: 8.2

dataset. The study also showed that feature selection can significantly improve the performance of career recommendation models.[5]

Al-Dossari et al. (2020) developed CareerRec, a Machine Learning approach for predicting career paths of Information Technology graduates. The study evaluated algorithms including KNN, Decision Tree, Bagging, Gradient Boosting, and XGBoost using technical and soft-skill-related attributes. XGBoost achieved the best performance with an accuracy of 70.47%. The study demonstrated that both technical and soft skills can be useful for predicting suitable career paths.[6]

Dahanke et al. (2022) proposed an intelligent career guidance system using Machine Learning techniques. The study used KNN classification and K-Means clustering to provide career guidance based on students' skills and abilities. The proposed system reported an accuracy of more than 90%, demonstrating the potential of Machine Learning for career guidance. However, the study was based on a relatively small manually created dataset with limited student attributes.[4]

III. RESEARCH GAP

Existing research has explored different Machine Learning algorithms for student career prediction and career recommendation. Various studies have used algorithms such as Random Forest, SVM, Decision Tree, KNN, and XGBoost.

However, some studies have used small datasets, limited student information, or specific groups of students. There is a need to study different Machine Learning models using a larger dataset with academic, skill, and employability-related information.

The proposed research addresses this gap by using 50,000 student records and 23 input features to compare Logistic Regression, Decision Tree, Random Forest, and Gradient Boosting for predicting suitable career domains.

IV. RESEARCH METHODOLOGY

1. Research Design: The study will adopt a descriptive and analytical research design.This is because the research paper will involve analyzing the student records, academic traits, and skill attributes to unveil the model's performance

2. Sources of Data: The study will adopt secondary data source.This is because the researcher will obtain the data from an open-source data platform known as Kaggle.

3. Sources of Dataset: There are approx 50,000 student records in this dataset.This dataset contains relevant information on student academics, skills set and employability attributes.

4. Method of analysis: Initially, the data will go through a preliminary analysis process to indicate the number of missing values, duplicates records and irrelevant features. The data will then be partitioned into training and testing sets in a ratio of 80:20. The training set contains 40000 records while the testing set contains 10000 records.

The data will then go through standardization process by using StandardScaler() to fit the training set and transform the test set. At this point, the numerical features will be converted to categorical features by using One-Hot Encoding technique.

The final step will involve developing four different classification models; Logistic Regression, Decision Tree, Random Forest and Gradient Boosting. The four algorithms will be used to predict the most probable career domain for the students. The prediction performance of each model will be compared and contrasted using Accuracy, Precision, Recall and F1-score. A classification report will also be generated to unveil the prediction performance of each class of the predicted career domains.

Finally, the study will compare the overall accuracy of the four classification algorithms and the best algorithm will be used to develop the student career guidance system. From the analysis, it was evident that Random Forest model was the best algorithm with a prediction accuracy of 70.49%.

V. ROLE OF MACHINE LEARNING IN STUDENT CAREER GUIDANCE

Machine Learning can be helpful in student career guidance due to its ability to analyze numerous characteristics of a student's profile. For this research, the student information, such as academic performance, skills, projects,

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, October 2026
Impact Factor: 8.2

certifications, internships, communication, problem-solving ability, and employability-related attributes, is considered for career domain prediction. The Machine Learning model goes into the analysis of these attributes and learns the connections between the student's profile and potential career domains based on the given information.

In this research, the classification algorithm that performs best on the given data was selected. Algorithms, such as Logistic Regression, Decision Tree, Random Forest, and Gradient Boosting, were tested and compared to one another using accuracy, precision, recall, and F1-score. Out of the tested models, the Random Forest algorithm showed a prediction accuracy of 70.49% and was chosen for this task. The trained model can then predict a suitable career option based on the information provided in the system.

The result of the model can provide the student with a solid basis for their potential career path based on their academic performance, skills set, and employability-related attributes. Thus, it can be concluded that Machine Learning can be a helpful tool in student career guidance. However, it is important to note that the predicted result is to be considered as a guideline or a recommendation towards one's career path, but not a definitive answer.

VI. CHALLENGES OF MACHINE LEARNING-BASED CAREER GUIDANCE

Machine learning can be utilized for student career guidance by predicting appropriate career domains based on student-related information. In this study, 23 input features consist of 15 numerical features and 8 categorical features were used to predict student Career Domain. Information such as age, attendance, study hours, CGPA, programming skills, projects, certifications, internships, communication skills, problem-solving ability, academic performance, university year, major, etc., are important features to predict a student's career domain.

The collected student data was preprocessed before applying the Machine Learning models to ensure that both numerical and categorical information were appropriately used for prediction purposes. In this study, four classification algorithms namely Logistic Regression, Decision Tree, Random Forest, and Gradient Boosting were applied to predict the Career Domain of students. The performance of the classification models was compared based on accuracy, precision, recall, and F1-score to determine the best-fit model for the given dataset.

Based on the experimental results, the Random Forest model predicted the career domain with the highest accuracy (70.49%) compared to the other models. As a result, Random Forest was selected as the best classification algorithm for the proposed Student Career Guidance System. The suggested model can utilize the information provided by the student to predict the most appropriate career domain, which can provide the student with an initial career guidance.

In this way, Machine Learning can be leveraged for career guidance by predicting the appropriate career of a student based on his/her information. However, it should be noted that the predicted career domain should be used for guidance purposes only.

VII. DATA ANALYSIS

The study findings proved that all the four standard ML models could predict the career domains of students. But for each model the prediction accuracy was different. Logistic Regression and Decision Tree attained an accuracy of 70.14% and 62.88% respectively. Gradient Boosting accuracy 69.65 percent . The Random Forest algorithm achieved the best result with an accuracy of 70.49% in turn.

The evaluation of the Random Forest model also revealed that its performance varied across career domains. In particular, the algorithm demonstrated a high level of precision (0.86), recall (0.95), and F1-score (0.90) for the Cybersecurity domain. For the AI & Data domain, the algorithm demonstrated a precision of 0.84, recall of 0.9, and F1-score of 0.87.

In turn, the Software Development and Embedded & Automation domains had a moderate level of F1-scores (0.79 and 0.77, respectively). In contrast, the Business & Analytics, Cloud & DevOps, IT Support, and Not Placed domains had a low level of performance. One reason for this outcome may be the imbalance in the distribution of student records across career domains.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, October 2026
Impact Factor: 8.2

Thus, approximately 70% of students in the dataset were labeled as "Not Placed." Altogether, the evaluation results demonstrated that the Random Forest model outperformed other algorithms in terms of accuracy. Therefore, it was selected as the final model for the Student Career Guidance System.

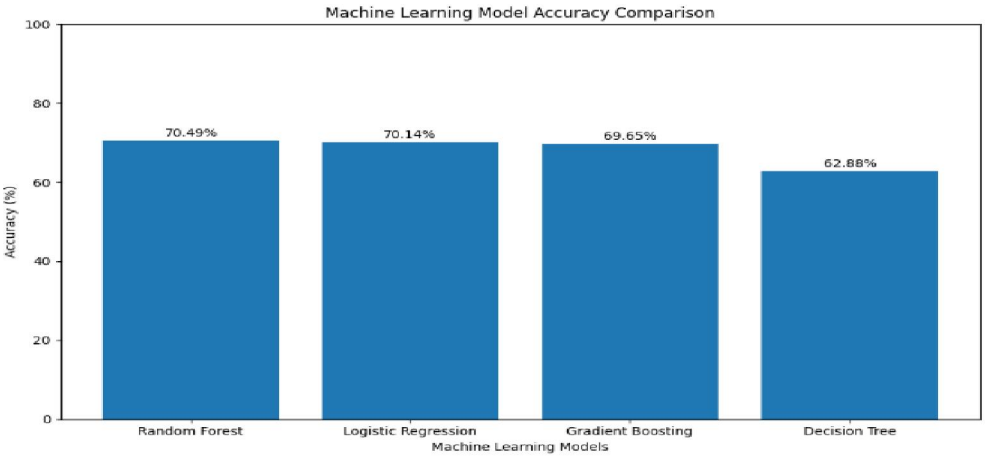


Fig 1. Machine Learning Model Accuracy Comparison

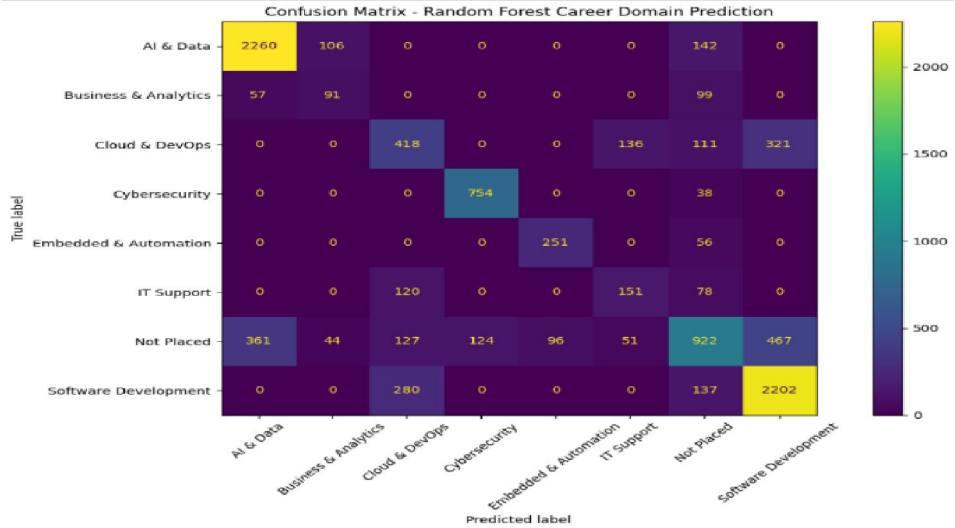


Fig 2.Confusion Matrix of Random Forest for Career Domain Prediction

VIII. CONCLUSION

The application of Machine Learning can contribute to a student's career guidance by transforming students' related data into a career domain recommender system. It can assist students by interpreting the most appropriate career fields according to their academic, skill, employability factors, and other attributes. The proposed system can support students in identifying and selecting the most suitable career options based on existing data.

The review of the existing research revealed that various machine learning algorithms can be implemented for career prediction and recommendation purposes. The considered studies utilized Random Forest, Support Vector Machine, Decision Tree, K-Nearest Neighbor, Gradient Boosting, and XGBoost algorithms for predicting the most relevant

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, October 2026
Impact Factor: 8.2

career domains. The results of the current research demonstrated that different algorithms yielded various accuracy rates in the selected dataset.

Nevertheless, it is essential to note that the application of Machine Learning cannot guarantee the best career recommendation for every student. The effectiveness of the career prediction model relies on the accuracy of the dataset, the student's attributes considered in the system, the selected algorithm and its performance, and the unique characteristics of the students.

In summary, it can be stated that the proposed Student Career Guidance System can be an essential solution for students in selecting the most appropriate career based on existing data. The highest accuracy rate (70.49%) was recorded for the Random Forest algorithm and was utilized as the final career prediction model.

LIMITATIONS OF THE STUDY

The study uses secondary data from Kaggle, and there is no data collected from actual students, which means the study cannot statistically verify how the system would perform when real students make career decisions. The model's accuracy can also differ for students who are studying in different colleges and regions

FUTURE SCOPE

Future work can be directed to improve the proposed Student Career Guidance System by using a larger and more diverse student database. Exploratory studies involving primary surveys, student records, and independent sets could assess the provided solution's accuracy and performance in actual conditions. Moreover, future research could explore other attributes related to students' interests, inclinations, and additional characteristics. A variety of other Machine Learning algorithms and methods for feature selection and parameter optimization could be considered to improve the prediction model's efficacy. Finally, the system could be designed as a web or mobile application to offer user-friendly access to the career guidance portal.

REFERENCES

[1] M. Nie, Z. Hui, J. Jiang, R. Zhang, W. Deng, and W. Yang, "Career Choice Prediction Based on Campus Big Data—Mining the Potential Behavior of College Students," *Applied Sciences*, vol. 10, no. 8, Art. no. 2841, 2020, doi: 10.3390/app10082841.

[2] N. VidyaShreeram and A. Muthukumaravel, "Student Career Prediction Using Machine Learning Approaches," *Proceedings of the First International Conference on Computing, Communication and Control System (I3CAC)*, 2021, doi: 10.4108/eai.7-6-2021.2308642.

[3] Z. Wang, G. Liang, and H. Chen, "Tool for Predicting College Student Career Decisions: An Enhanced Support Vector Machine Framework," *Applied Sciences*, vol. 12, no. 9, Art. no. 4776, 2022, doi: 10.3390/app12094776.

[4] D. Ajay, N. Shinde, A. Dhagate, and H. Shaikh, "An Intelligent Career Guidance System using Machine Learning," *International Research Journal of Engineering and Technology (IRJET)*, vol. 9, no. 3, pp. 1684–1691, 2022.

[5] S. El-Keiey, D. ElMenshawy, and E. Hassanein, "Career Recommendation Based on Feature Selection for Undergraduate Students Using Machine Learning Techniques," *International Journal of Advanced Computer Science and Applications*, vol. 16, no. 3, 2025.

[6] H. Al-Dossari, F. A. Nughaymish, Z. Al-Qahtani, M. Alkahlifah, and A. Alqahtani, "CareerRec: A Machine Learning Approach to Career Path Choice for Information Technology Graduates," *Engineering, Technology & Applied Science Research*, vol. 10, no. 6, pp. 6589–6596, 2020.