INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD)
© 2026 IJNRD | Volume 11, Issue 7, July 2026 | ISSN: 2456-4184 | IJNRD.ORG

PREDICTING STUDENT PLACEMENT USING MACHINE LEARNING TECHNIQUES

A Comparative Analysis of Classification Algorithms

[1]Ms. Anushree, [2]Mr. Rajesh Naik, Ms. Sharanya M , [3]Ms. Ananya Kedila

[1]Assistant Professor, [2]Assistant Professor, [3]Assistant Professor , [4]Assistant Professor
[1]Department of Master of Computer Applications

,

[1]Srinivas Institute of Technology, Mangalore, Karnataka, India

Abstract : Student placement is an important indicator of academic success and employability. This study presents a machine learning-based approach to predict student placement using factors such as CGPA, aptitude score, communication skills, programming skills, and attendance. Classification algorithms including Decision Tree, Random Forest, Logistic Regression, Support Vector Machine (SVM), and XGBoost are compared to identify the most effective model. The results can help educational institutions identify students who need additional support and improve placement outcomes through data-driven decisions**.**

IndexTerms **– Student Placement Prediction, Machine Learning, Random Forest, Educational Data Mining, Predictive Analytics.**

I. INTRODUCTION

1.INTRODUCTION

Campus placement is an important milestone in a student's academic journey and plays a significant role in determining employability. Educational institutions continuously strive to improve placement outcomes by identifying students who may require additional academic and skill development. However, predicting placement success based on traditional methods is often challenging due to the influence of multiple academic and personal factors.

Machine Learning (ML) provides intelligent techniques for analyzing educational data and predicting future outcomes with improved accuracy. By utilizing student-related attributes such as CGPA, aptitude score, communication skills, programming skills, and attendance, ML models can identify patterns that influence placement success.

This study proposes a machine learning-based approach for predicting student placement using various classification algorithms, including Decision Tree, Random Forest, Logistic Regression, Support Vector Machine (SVM), and XGBoost. The performance of these algorithms is evaluated to determine the most effective model for placement prediction. The proposed system aims to assist educational institutions in identifying students who require additional training and guidance, thereby enhancing their employability and improving overall placement performance

II. NEED OF THE STUDY.

In today's competitive job market, accurately predicting student placement has become essential for educational institutions. Early identification of students who are at risk of not being placed enables colleges to provide focused training in technical, aptitude, and communication skills. Traditional assessment methods often fail to analyze multiple factors simultaneously, leading to less accurate predictions. Machine Learning techniques offer an efficient solution by analyzing academic and skill-based data to identify patterns that influence placement outcomes. This study aims to develop a reliable prediction model that assists institutions in improving placement strategies, enhancing student employability, and increasing overall campus placement success.

2.1 Early Identification of Placement Readiness

Campus placement is one of the most important milestones in a student's academic journey, as it reflects both individual employability and the effectiveness of an educational institution. However, many students face difficulties during placement drives due to inadequate technical knowledge, poor communication skills, low aptitude performance, or insufficient practical exposure. Identifying such students only at the final stage often limits the opportunity for improvement.

An effective prediction system is therefore essential to identify students who may require additional support well before campus recruitment begins. By analyzing academic performance and skill-based attributes such as CGPA, internships, projects, aptitude scores, soft skills, and placement training, educational institutions can recognize students who are at risk of not being placed. Early prediction enables faculty members to provide personalized mentoring, targeted training programs, mock interviews, and skill enhancement activities. This proactive approach helps improve students' confidence, employability, and overall placement success while enabling institutions to make informed, data-driven decisions for better placement outcomes.

2.2 Improving Placement Decision-Making

Traditional methods of evaluating students are often based only on academic performance and may not accurately predict placement outcomes. Machine learning techniques analyze multiple factors such as CGPA, aptitude score, internships, projects, certifications, soft skills, and placement training to provide more accurate predictions and support data-driven decision-making..

2.3 Enhancing Student Employability

Predicting placement outcomes enables institutions to design personalized training programs for students based on their strengths and weaknesses. By identifying key factors influencing placement success, colleges can improve technical skills, communication abilities, and overall employability, thereby increasing campus placement rates.

III. RESEARCH METHODOLOGY

The research methodology outlines the procedures followed to develop and evaluate the student placement prediction model. It includes the population and sample, data sources, and the theoretical framework used for the study.

3.1 Population and Sample

The population of this study consists of students eligible for campus placements. A publicly available student placement dataset containing **10,000 student records** is used for analysis. Each record includes academic performance, technical skills, training, and placement-related information. The dataset serves as the sample for developing and evaluating the machine learning models..

.

3.2 Data and Sources of Data

This study uses a secondary dataset obtained from Kaggle. The dataset contains student information such as CGPA, Internships, Projects, Workshops/Certifications, Aptitude Test Score, Soft Skills Rating, Extracurricular Activities, Placement Training, SSC Marks, HSC Marks, and Placement Status. The collected data are preprocessed by removing duplicate records, handling missing values, encoding categorical variables, and normalizing numerical features before model development.

3.3 Theoretical framework

The proposed framework is based on supervised machine learning techniques for predicting student placement outcomes. The study aims to develop a predictive model that analyzes academic performance, technical skills, and extracurricular activities to determine whether a student is likely to be placed.

The dependent variable in this study is Placement Status, which indicates whether a student is Placed or Not Placed. The independent variables include CGPA, Internships, Projects, Workshops/Certifications, Aptitude Test Score, Soft Skills Rating, Extracurricular Activities, Placement Training, SSC Marks, and HSC Marks. These variables are considered important indicators of a student's academic ability, technical competence, and employability.

Before model development, the dataset is preprocessed by handling missing values, removing duplicate records, encoding categorical variables, and normalizing numerical features. The processed data are then divided into training and testing sets for model development.

Three supervised machine learning algorithms Logistic Regression, Random Forest, and XGBoost are implemented to predict student placement. The performance of these models is evaluated using Accuracy, Precision, Recall, F1-Score, and Confusion Matrix. The algorithm with the best evaluation results is identified as the most effective model for predicting placement outcomes and supporting educational institutions in making data-driven decisions to improve student employability.

3.4 Statistical tools and Machine learning models

Statistical analysis and machine learning techniques are used to analyze the dataset and predict student placement. Descriptive statistics summarize the characteristics of the data, while classification algorithms are applied to build predictive models. The performance of each model is evaluated using standard classification metrics.

3.4.1 Descriptive Statistics

Descriptive statistics are employed to summarize and analyze the numerical characteristics of the student placement dataset. Statistical measures such as mean, median, minimum, maximum, and standard deviation are computed for variables including CGPA, internships, projects, workshops/certifications, aptitude test score, soft skills rating, SSC marks, and HSC marks. These measures provide valuable insights into the central tendency, dispersion, and overall distribution of the data, helping to identify patterns and variations among student records.

Before developing the machine learning models, the dataset undergoes a comprehensive preprocessing stage to ensure data quality and consistency. Missing values are identified and appropriately handled, duplicate records are removed, categorical variables are encoded into numerical values, and numerical features are normalized to maintain a uniform scale. These preprocessing techniques reduce data inconsistencies, improve model efficiency, and enhance the overall prediction accuracy. The descriptive statistical analysis also assists in understanding the characteristics of the dataset and provides a strong foundation for building reliable student placement prediction models.

3.4.2 Machine Learning Models

This study employs three supervised machine learning algorithms Logistic Regression, Random Forest, and XGBoost to predict student placement. These algorithms were selected because they are widely used for binary classification problems and have demonstrated strong performance in educational data mining.

The dataset is divided into 80% training data and 20% testing data. The training dataset is used to develop the prediction models by learning the relationship between the independent variables CGPA, Internships, Projects, Workshops/Certifications, Aptitude Test Score, Soft Skills Rating, Extracurricular Activities, Placement Training, SSC Marks, and HSC Marks—and the dependent variable, Placement Status. The trained models are then evaluated using the testing dataset to assess their ability to predict placement outcomes for unseen student records.

Logistic Regression is used as a baseline classification model to estimate the probability of a student being placed based on the input features. Random Forest, an ensemble learning algorithm, constructs multiple decision trees and combines their predictions to improve accuracy while reducing overfitting. XGBoost is an advanced boosting algorithm that sequentially builds decision trees to minimize prediction errors and improve classification performance.

The prediction results of all three models are compared to determine the most suitable algorithm for student placement prediction. The comparison helps identify the model that provides the highest predictive accuracy and can effectively support educational institutions in improving placement strategies and enhancing student employability.

3.4.3 Comparison of the Models

The performance of the three machine learning models Logistic Regression, Random Forest, and XGBoost is compared to determine the most effective algorithm for predicting student placement. Each model is trained using the same training dataset and evaluated on the same testing dataset to ensure a fair and unbiased comparison. The models are assessed using standard classification metrics, including Accuracy, Precision, Recall, F1-Score, and Confusion Matrix.

Logistic Regression serves as the baseline model and provides reliable predictions for linearly separable data. However, its performance may be limited when the relationship between the input features and placement status is complex. Random Forest improves prediction accuracy by combining multiple decision trees, reducing overfitting, and effectively capturing non-linear relationships among student attributes. XGBoost, an advanced ensemble learning algorithm, further enhances prediction performance by sequentially minimizing classification errors through gradient boosting techniques.

The evaluation results are compared to identify the model with the highest predictive performance. The algorithm that achieves superior values in Accuracy, Precision, Recall, and F1-Score is considered the most suitable for predicting student placement. The comparison also provides insights into the strengths and limitations of each model, enabling educational institutions to adopt the most effective predictive approach for improving placement strategies and supporting student career development.

IV. RESULTS AND DISCUSSION

4.1 Results of Descriptive Statics of Study Variables

Table 4.1: Descriptive Statistics of Study Variables

Variable	Minimum	Maximum	Mean	Std. Deviation
CGPA	6.0	9.1	7.71	0.72
Internships	0	2	1.01	0.81
Projects	1	3	2.00	0.82
Workshops/Certifications	0	3	1.49	1.12
Aptitude Test Score	65	90	79.94	8.10

Table 4.1 presents the descriptive statistics of the variables used in this study. The analysis includes the minimum value, maximum value, mean, and standard deviation of the selected academic and employability-related attributes.

The results indicate that the average CGPA of students is 7.71, suggesting satisfactory academic performance among the sampled students. The mean number of internships completed is 1.01, while the average number of projects completed is 2.00. Students have also participated in an average of 1.49 workshops or certifications, indicating moderate engagement in skill-development activities.

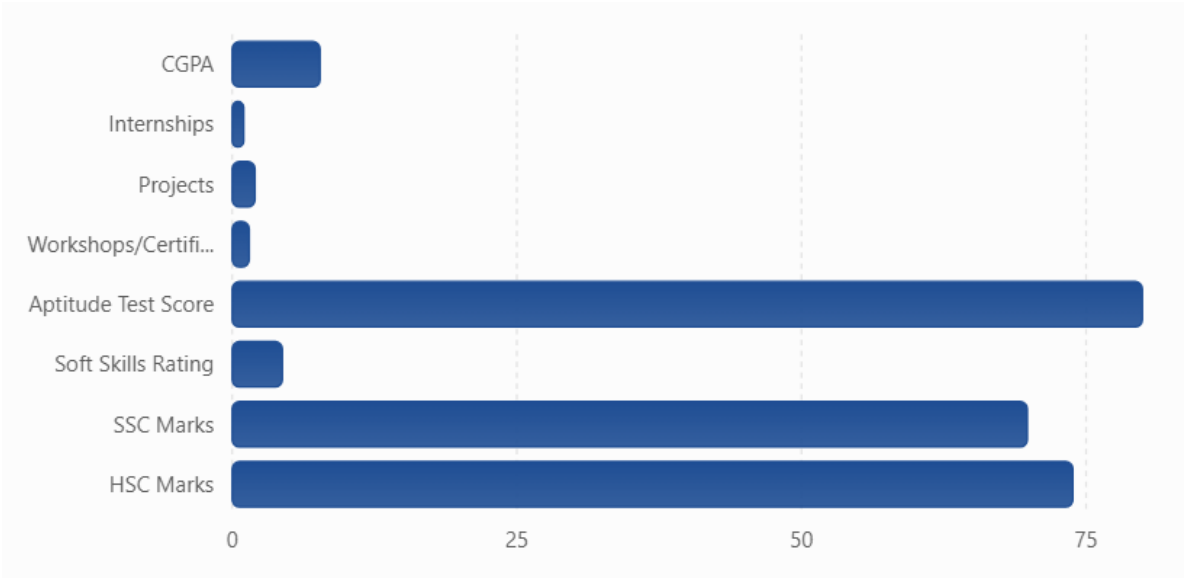
The mean aptitude test score is 79.94, and the average soft skills rating is 4.40, reflecting good technical and communication abilities among the students. The average SSC marks and HSC marks are 69.84 and 73.84, respectively.

The standard deviation values show that there is reasonable variation among student records, particularly in aptitude scores and school-level marks. This variability is useful for machine learning analysis because it allows the prediction models to learn meaningful patterns from the data.

Overall, the descriptive statistics confirm that the dataset contains diverse academic and skill-related characteristics, making it suitable for developing and evaluating machine learning models for student placement prediction.

Figure 4.1: Mean Values of Study Variables

Average values calculated from the student placement dataset.



V. Acknowledgment

The authors express their sincere gratitude to the Management and Principal of Srinivas Institute of Technology, Mangaluru, for providing the necessary facilities and encouragement to carry out this research work. The authors are thankful to the faculty members of the Department of Master of Computer Applications for their valuable guidance, continuous support, and constructive suggestions throughout the study.

The authors also acknowledge the Kaggle platform for providing the publicly available student placement dataset used in this research. Finally, the authors extend their heartfelt thanks to their family members, friends, and well-wishers for their constant encouragement and support during the completion of this research work.

REFERENCES

[1] K. Rai, "Student Placements Prediction Using Machine Learning Algorithm," *South Asia Journal of Multidisciplinary Studies (SAJMS)*, vol. 8, no. 5, June 2022.

[2] K. M. Naresh Patel, N. M. Goutham, K. A. Inzamam, V. K. Suraksha, and V. R. Vineet Sharan, "Placement Prediction and Analysis Using Machine Learning," *International Journal of Engineering Research & Technology (IJERT)*, ICEI Conference Proceedings, 2022.

[3] S. Harinath, A. Prasad, H. S. Suma, A. Suraksha, and T. Mathew, "Student Placement Prediction Using Machine Learning," *International Research Journal of Engineering and Technology (IRJET)*, vol. 6, no. 4, Apr. 2019.

[4] V. K. Harihar and D. G. Bhalke, "Student Placement Prediction System Using Machine Learning," *SAMRIDDHI*, vol. 12, Special Issue 2, 2020.

[5] N. M. Goutham, K. A. Inzamam, V. K. Suraksha, V. R. Vineet Sharan, and K. M. Naresh Patel, "Placement Prediction and Analysis Using Machine Learning," *International Research Journal of Modernization in Engineering Technology and Science*, vol. 4, no. 7, July 2022.

[6] P. Archana, D. Pravallika, P. Sindhu Priya, S. Sushmitha, and S. Amitha, "Student Placement Prediction Using Machine Learning," *Journal of Survey in Fisheries Sciences*, 2023.