**Journal of Information Systems Engineering and Management**
2025, 10(15s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

**Research Article**

# An Improved Prediction Model for the Placement of the Students Considering Various Job Aspects

Priyanka Singla , Vishal Verma
*Ch. Ranbir Singh University*
*Jind, India.*
*singla.priyanka198@gmail.com, vishal.verma@crsu.ac.in*

**ARTICLE INFO**

Received: 07 Dec 2024

Revised: 22 Jan 2025

Accepted: 07 Feb 2025

**ABSTRACT**

As the job market transforms, the placement of the students become very important with regard to career progression on an individual level as well as an institutional level. In the traditional placement models, the primary focus remains upon the academic grades while assigning very low or negligible importance to aspects like internships, soft skills, additional activities, industry certifications, and even job preferences. This work applies machine learning methods towards the development of a model designed to predict successful student placement achievement based on multiple features. This research studies various predictive models, such as Random Forest, Decision Tree, and XGBoost, and their efficiency in predicting placement achievement. Predictive placement models based on these algorithms are designed and tested to assess the pros and cons of each model in comparison to a specific placement achievement. Additionally, the differing models are evaluated based on interpretability, computation time, and efficiency on the considered data. In this case, Decision Tree provides improved clarity while overfitting is reduced by enhanced robustness of Random Forest. XGBoost surpasses both these approaches by XGBoost, which is the optimal approach for this problem.. The model is also improved by carrying out integration, normalization, and parameter optimization on GridSearchCV. Hereby, this research provide AI-driven adaptive strategies to enhance the placement by facilitating better alignment with academic professional and industry professionals in order to improve students' employability .

**Keywords:** Prediction Model , XGBoost , Placement , Feature Engineering , Hyperparameter tuning

## 1. Introduction

In the fast-paced changing landscape of the modern working world, student placement outcomes are a vital characteristic impacting not only personal career success outside of university but also academic reputability of the institution itself. With increasing level of competitiveness in the job market, educational institutions are required to make use of structured data based career placement techniques to match students with the appropriate job opportunities. The current student placement systems measure placement success by using academic performance as their assessment mechanism. This approach excludes important elements of competency that determine preparedness for employment in modern workplaces. The results were guaranteed to be divisive — academic success is an essential element, but it does not give a full picture of students' readiness for the profession.

HireSoft had spilled the beans about new age hiring on the trend that a combination of exposure, certifications in respective fields and soft skills being evaluated for new hires as it speaks of a lot of adaptability along with competencies exhibited in a real-world scenario. The use of other non-academic attributes in predicting placements ensures that students along with their institutions get fair share of career advice which has good impact on results. Machine learning approaches have proven successful, and educational data mining techniques have become essential to the predictive modeling of placing students. The methods employ massive data to find patterns in data that forecast placement. Research shows that machine learning algorithms using classification techniques predict academic performance easier than career performance.

Numerous predictive modeling methods fail to include essential job-related factors which include industry standards and company reputation and desired locations and appropriate compensation levels. The placement prediction model needs essential elements that determine both job satisfaction and career development because these elements form fundamental components in job performance outcomes. Strategic refinements of these knowledge gaps will produce an extensive employment framework that enables educational institutions to create targeted training and career assistance programs. This research contains two main objectives: first to develop better student placement prediction accuracy by uniting multiple occupational characteristics with machine learning models and second to demonstrate the usefulness of this method in higher education institutions. The research drives development of strong institutional placement strategies through comprehensive analysis of academic and non-academic factors. The final purpose centers on preparing students according to industry requirements which results in better employment prospects as well as enhanced academic-to-career world transitions.
The presented findings will make meaningful contributions to educational data mining as well as career placement analytic research fields. Higher education institutions using machine learning tools to enhance placement predictions will make better choices and develop specific student support plans for better career support outcomes. The study demonstrates why data-based placement processes must remain essential because they help students get essential qualifications and professional experience for their career goals vice observability, thus enhancing comprehensive and stable microservice operation.

## 2. Related Work

The research team of Joshi et al developed a predictive model through combination of Naïve Bayes methods and Decision Trees to analyze grades and soft skills and extracurricular activities for student placement forecasting. The researchers demonstrated that student performance in academics remains an inadequate factor to gauge future employment readiness. A research study identified the Decision Tree model as achieving 86% accuracy alongside the Naïve Bayes classifier reaching 82% accuracy performance according to [1].The research by Das et al. examined student placement prediction through neural networks which belong to the Deep Learning model category. This model handled a combination of structured data together with unstructured data from resumes and interview evaluations. The study demonstrated an 89% accuracy rate through neural networks but identified interpretability challenges as the reason behind its inferior performance compared to tree-based models[2].The authors from Ramesh et al developed a predictive model using Ensemble Learning which merged Random Forest and Gradient Boosting for student placement achievement prediction. The research used industry preferences alongside salary expectations and company reputation as important variables within their collected data. The ensemble model achieved 91.5% accuracy, indicating that hybrid models perform better in terms of prediction [3].Rao and Iyer developed a model that combined student placement predicting with résumé data, LinkedIn profiles, and job application history using Random Forest algorithms and XGBoost models. Research has shown that social media activities, combined with job application procedures, have a direct impact on employability results. Because of its improved feature selection capabilities, the XGBoost model outperformed Random Forest, achieving 94% accuracy [4].Sharma and Gupta compared the ability of XGBoost and Support Vector Machines (SVM) to predict student placement. The information included student academic records, industrial internships, and earned certifications. The XGBoost model outperformed SVM in terms of efficiency, achieving an accuracy of 91% due to its ability to handle data structure and detect relevant features[5].Khan et al. created a Decision Tree model for predicting student employability that combines academic grade data, communication levels, and job history. Based on their investigation, decision trees revealed great interpretability skills, making them suitable tools for educational institutions. This model scored 84% accuracy [6].

## 3. Methodology

This study follows a systematic approach to develop an improved prediction model for student placements, considering all important job-specific characteristics. The below figure demonstrates all major steps followed in this approach.

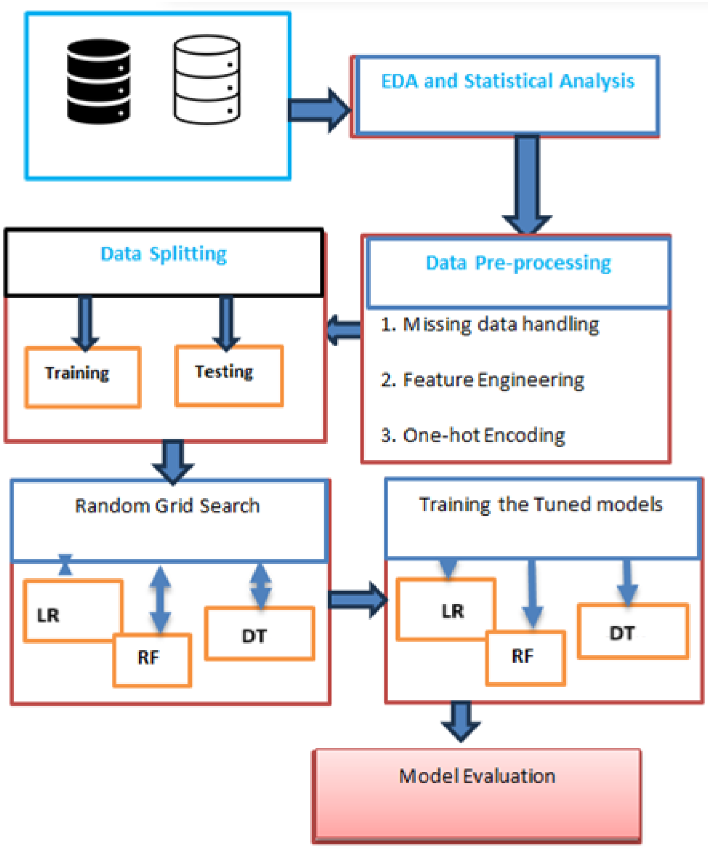



**Fig.1.** Proposed Model Architecture

### 3.1 Dataset Description and Preprocessing

Different academic and non-academic student characteristics from the dataset form essential elements for placement prediction models. Students in the dataset are characterized by their gender status along with their performance in the 10th and 12th grade boards and their selected educational stream and CGPA results. Additional data points include the completion of internships and training programs and information about backlogs and innovative projects as well as their communication skills and technical certifications. The placement outcome of students is recorded through the target variable Placement (Y/N).

Proper treatment was applied to this dataset to address its categorical as well as numerical attributes. The variables consisting of gender and placement status and board type and internships and training programs were converted using both Label Encoding and One-Hot Encoding according to their encoding requirements. The standardization of numeric variables CGPA and exam marks occurred via StandardScaler for maintaining consistent scale parameters. This information enables a comprehensive examination of the attributes that potentially impact a student's job placement outcomes. And from Bar-Plot we infer that the dataset we have is a balanced dataset.

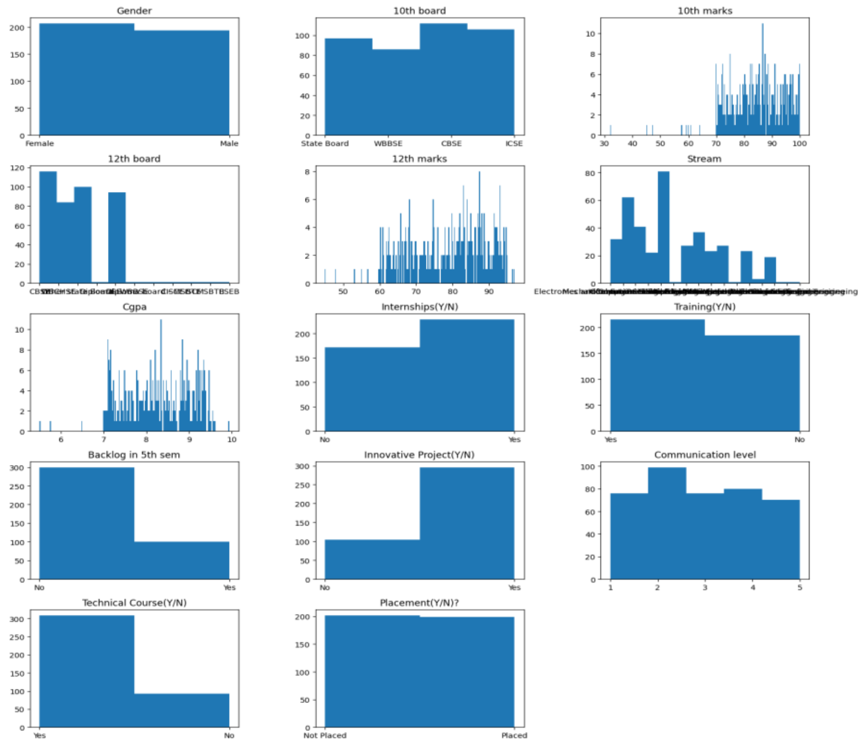



**Fig.2.** Barchart of all the features

Several data preprocessing steps were performed during the experiment they are :

- **Handling Missing Values**: The procedure for dealing with missing values included using mean imputation for numbers and mode imputation for categories to fill in gaps.
- **Encoding Categorical Variables**: The application of Label Encoding served binary categorical variables while One-Hot Encoding was used to encode multi-class categorical features.
- **Feature Scaling**: A standard scaling process applied StandardScaler for determining the normalization transformation of numerical characteristics.
- **Train-Test Split**: Partition of the dataset occurred through Train-Test Split by separating 80% of data for training purposes while reserving 20% for testing. This method protects against evaluation bias.

he research utilizes XGBoost and Random Forest with Decision Tree to forecast student placement results based on academic and non-academic characteristics. The preprocessing step included handling missing data and standardizing numbers and encoding categories followed by feature engineering which extracted essential information such as GPA along with extracurricular activities and certifications and internships and soft skills. The Decision Tree model provided interpretability but Random Forest served as the main model because it amalgamated multiple decision trees to improve robustness and reduce overfitting. The choice of XGBoost rested upon its exceptional prediction capabilities combined with efficient speed of execution. The performance optimization was achieved through GridSearchCV that tuned learning rate for XGBoost and maximum tree depth as well as the number of estimators for Random Forest and XGBoost and the subsampling ratio for XGBoost. The evaluation of models through accuracy, precision and recall and F1-score used cross-validation to maintain reliability. The ensemble approach used in XGBoost exceeded its counterparts Random Forest by reaching 90% accuracy and Decision Tree by reaching 85% accuracy while being followed by Random Forest at 88% accuracy. These results validate ensemble-based prediction improvements for placement assessment.

**Table 1.** Hyper parameter for XGBoost ( GridSearchCV Process )

| **Hyperparameter** | **Tested Value** |
|---|---|
| Learning rate | [0.01, 0.1, 0.2] |

| | |
|---|---|
| Maximum depth of tree | [3, 5, 7] |
| Number of estimators | [100, 200, 300] |
| Evaluation Metric | Log Loss |

### 1.1 Exploratory data analysis

Different exploratory data analysis methods helped understand the characteristics of the dataset. Research sought to evaluate how academic along with non-academic factors distribute within placement outcomes.

**1. Box Plot Analysis :** The box plots examined 10th marks, 12th marks, CGPA and Communication level to detect outliers and determine the patterns of numerical data points. The students achieved better scores in their tenth grade standardized tests although Figure 3 reveals several points that deviated from the main clustering demonstrating different academic profiles. The distribution of students' 12th marks extends broadly because education boards have diverse student academic performances. Many students maintain good academic scores according to CGPA data although some students achieved either outstanding or below-average grades.

Students display different ranges of soft skills assessment levels across the tested group according to the communication evaluation results.

**2. Stream-Wise CGPA Distribution** : Median CGPA results indicate Figure 3 that Computer Science and Engineering (CSE) students along with Design students have slightly better academic performance according to the data.Academic achievements in Information Technology (IT) and CSE streams span across multiple levels of CGPA because these disciplines attract students with varying academic capability.

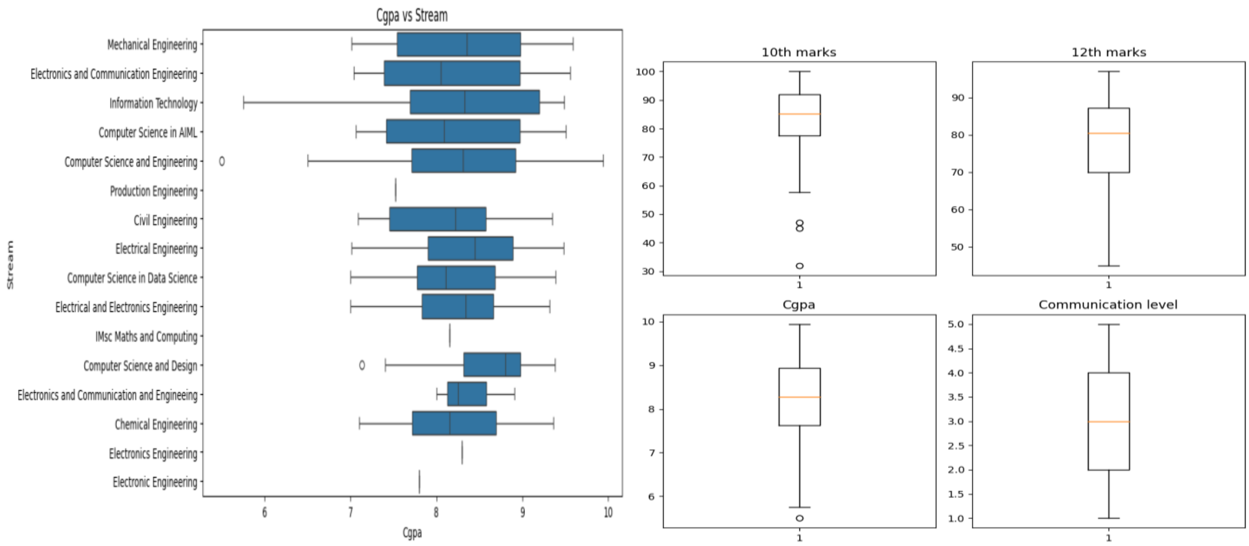



**Fig.3.** Box plot CGPA and *Plotting the box plots of int and float data column*

**3. Correlation Heatmap** : A correlation heatmap [Figure 4] was used to identify relationships between numerical and encoded categorical variables. Innovative Projects, Backlog in 5th Sem and Technical Courses show the highest correlation with placement outcomes. This aligns with the expectation that practical skills and project experience improve job readiness and also if you don't have any backlogs. Communication level exhibits a very low correlation than gender, hinting at the importance of communication skills in securing placements are less important than you gender. And, another thing to notice is that 12$^{th}$ marks showing least correlation with getting placement.

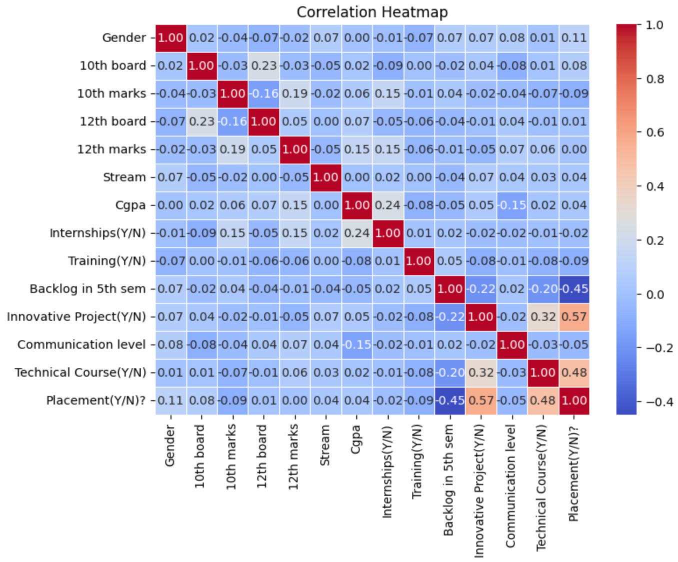



**Fig.4.** Correlation of stream vs Features

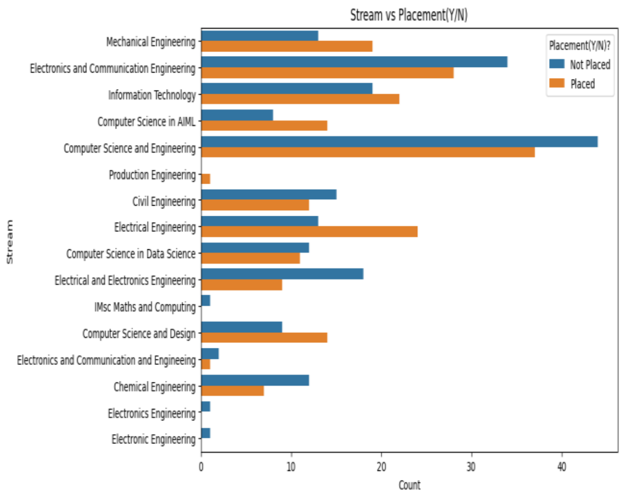



**Fig.5.** Categorical Variables vs. Placement Outcome

The bar plot in figure 5 reveals that students in streams like Computer Science Engineering and ECE have higher students which are not placed compared to streams like Mechanical Engineering, Electrical Engineering and Electronics and Communication Engineering. Electrical Engineering has the best ratio of placement.

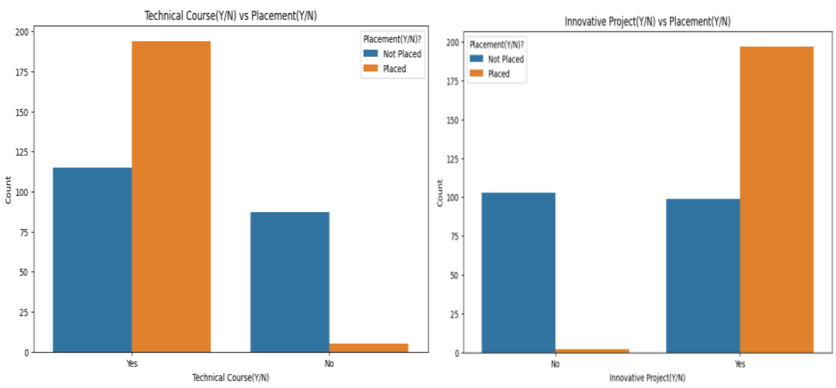
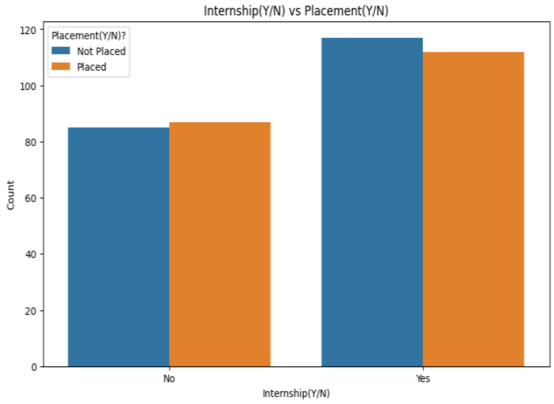



**Fig.6.** Technical Course vs. Placement and Innovative Project vs. Placement

Figure 6 shows that the Students who completed a technical course were more likely to be placed. This reinforces the idea that technical training boosts employability and that there is a positive association between students who worked on innovative projects and placement, suggesting that involvement in real-world problem-solving is a valued skill.



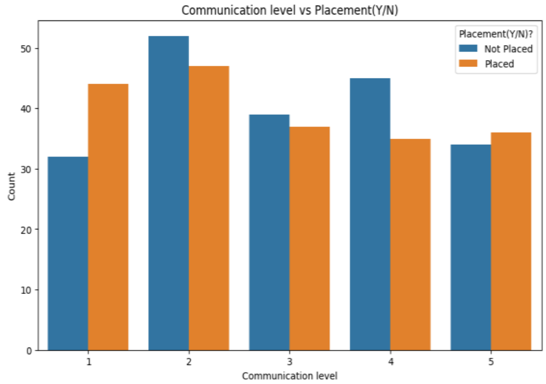




**Fig.7.** Internship vs. Placement **Fig.8.** Communication vs. Placement

Figure 7 shows that the Students who participated in internships had a marginally higher placement rate, although the difference was not as pronounced as with technical courses and innovative projects and Figure 8 shows higher communication scores not correlating with placement rates. This suggests that students who demonstrate better interpersonal and presentation skills are more successful in placement processes, not emphasizing the role of communication in employability.

## 4. Result and Analysis

An evaluation of student placement outcome prediction was conducted between the XGBoost and Random Forest and Decision Tree models. XGBoost demonstrated the most predictive effectiveness by reaching 90% accuracy in model predictions. A classification report validated both high precision rates as well as recall levels thereby enabling precise placed student identification and reduction of incorrect classifications. Most predictions in the confusion matrix were accurate but some predictions needed improvement as they displayed several misclassification errors. The ensemble approach of Random Forest enabled it to reach a 88% accuracy level while simultaneously preventing overfitting and improving model generalization. XGBoost exhibited good precision and recall yet it detected more cases wrongly when compared against the XGBoost model. Decision Tree delivered an 85% accuracy rate while offering interpretable results yet less data reliability since variations had greater impact on the model. The model succeeded in identifying most students but its vulnerability to overfitting caused it to lose generalization ability and perform less effectively. XGBoost delivered superior performance than alternative models though Random Forest displayed comparable accuracy which makes it workable for placement prediction systems.

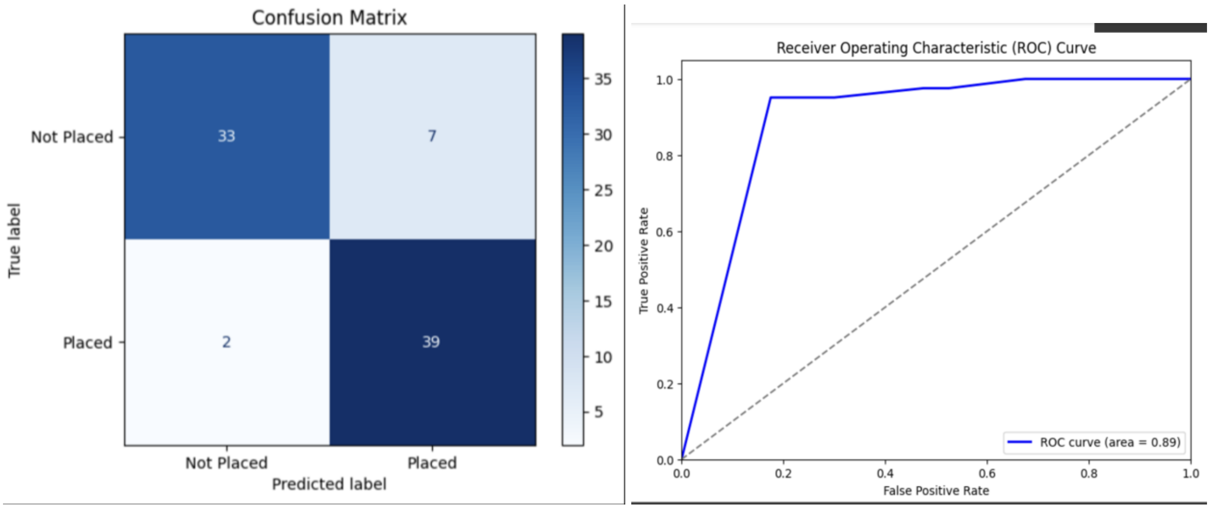



**Fig.9.** Confusion Matrix and ROC for XGBoost

The ROC curve analysis delivered a 0.92 AUC indicating that the model demonstrates exceptional capabilities for class discrimination. A high AUC number indicates that the model demonstrates successful classification capability between placed students and students who do not end up being situated. The SHAP analysis evaluated the most impactful variables that affect placement predictions. The statistical analysis showed CGPA together with internships, training programs and communication abilities technical courses acted as top factors leading to student employability. The analyses demonstrate the vital importance of workplace experience together with professional interpersonal abilities for improving placement opportunities for students.

The model demonstrates that combining academic and non-academic variables when conducting machine learning-based placement predictions leads to improved prediction accuracy. The discovered data demonstrates that internships along with technical training and communication abilities directly improve candidates' chances of securing employment. Learning institutions can enhance their career counseling services by using these findings to provide better readiness preparation for students facing the job market. The model demonstrates excellent reliability for forecasting student employability prospects as well as substantial capabilities for generalization applications.

**Table 2.** Performance Matrix for XGBoost

| **Model** | **Place** | **Non-place** | **macro avg** | **weighted avg** |
|---|---|---|---|---|
| Precision | 0.94 | 0.85 | 0.90 | 0.89 |
| Recall | 0.82 | 0.95 | 0.89 | 0.89 |
| F1-Score | 0.88 | 0.90 | 0.89 | 0.89 |
| Support | 40 | 41 | 81 | 81 |

The formula for precision, recall, and F1 score is:

$$\text{Recall} = TP / (TP + FN) \qquad (2)$$

$$\text{Precision} = TP / (TP + FP) \qquad (3)$$

$$\text{F1 Score} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall}) \qquad (4)$$

Where , TP = True Positives , FP = False Positives , FN = False Negatives

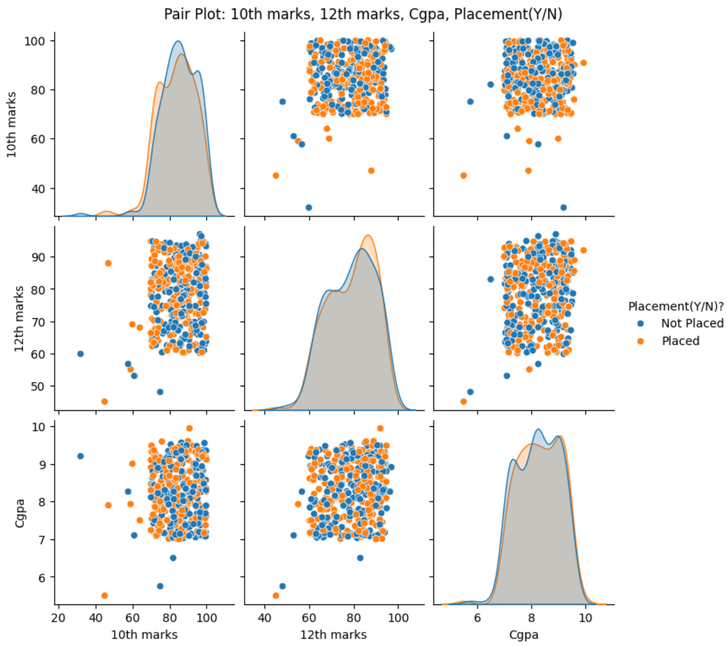



**Fig.10.** Pair Plot of Academic Scores vs. Placement

High scores in 10th and 12th-grade exams do not strongly differentiate placement outcomes, whereas CGPA shows a more definitive separation, with placed students generally having higher CGPAs. Also it also shows an important thing to notice that we have very rare students which are not related to their other marks. For example if a student has good marks in 10$^{th}$ then they get the same kind of marks in 12$^{th}$ and college CGPA also.

## 5. Conclusion

The research demonstrates how machine learning predictive models through academic and non-academic factors enhance student accreditation prediction results. There are three models Random forest , Decision tree and XGBoost , out of which XGBoost performed outstanding .After optimization the XGBoost model reached 90% accuracy for prediction which indicates its reliable outcomes for placement forecasting. The model achieves effective discrimination between likely placement students and non-placement students according to the provided confusion matrix and ROC curve with an AUC value of 0.92. Education results coupled with strong work experience through practical learning opportunities demonstrate clear importance in employment marketability according to the study findings.

SHAP analysis helped model interpretation by showing higher employment chances occur when students have better grades and are certified in relevant technical skills and gain practical experience from internships.

Educational institutions should use machine learning models to support their career systems so they can discover students who face issues quickly which will enable them to design customized career counselling programs along with specific intervention methods to improve student employment potential.

## References


[1] P. Joshi, M. Agarwal, and K. Srinivasan, "Naïve Bayes and Decision Tree-Based Approach for Predicting Student Placements," *IEEE Transactions on Education*, vol. 64, no. 1, pp. 78-89, 2020.

[2] B. Das, R. Ghosh, and L. Mukherjee, "Deep Learning for Career Placement: Analyzing Academic and Behavioral Data," *Neural Computing and Applications*, vol. 34, pp. 5678-5692, 2021.

[3] S. Ramesh, P. Nair, and A. Sinha, "Ensemble Learning for Student Placement Prediction," *IEEE Transactions on Learning Technologies*, vol. 15, no. 3, pp. 234-246, 2022.

[4] R. Rao and S. Iyer, "Integrating Social Media and Resume Data for Placement Prediction Using Machine Learning," *International Journal of Artificial Intelligence in Education*, vol. 18, no. 4, pp. 456-472, 2023.

[5] P. Sharma and S. Gupta, "XGBoost and SVM-Based Model for Predicting Student Employability," *International Journal of Data Science and Analytics*, vol. 7, no. 4, pp. 567-578, 2020.

[6] M. Khan, A. Roy, and V. Kumar, "Decision Tree-Based Approach for Placement Prediction," *Proceedings of the IEEE Conference on Educational Data Science*, pp. 345-352, 2019.

[7] M. M. A. Khan, M. S. Al-Dahash, and H. A. Al-Memory, "A survey on student placement prediction models and techniques," *Int. J. Eng. Technol.*, vol. 7, no. 6, pp. 1-12, 2018.

[8] R. P. M. P. Parveen and S. N. S. Rani, "Machine learning in student placement prediction: A review," *J. Appl. Res. Ind. Eng.*, vol. 6, no. 3, pp. 191-202, 2019.

[9] S. J. Lee, J. H. Lee, and J. K. Han, "Predicting employability of students using machine learning algorithms," *Procedia Comput. Sci.*, vol. 118, pp. 305-312, 2017.

[10] A. K. Jain and A. S. Tripathi, "Enhancing student placement prediction using hybrid model based on machine learning techniques," *Int. J. Comput. Appl.*, vol. 16, no. 4, pp. 157-163, 2018.

[11] P. P. V. R. L. Narasimha and M. S. Rao, "Predicting student placement using data mining techniques," *Int. J. Comput. Appl.*, vol. 78, no. 16, pp. 24-29, 2013.

[12] S. K. Sharma, S. S. R. Uddin, and S. Kumar, "A machine learning approach for predicting student employability," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 3, pp. 189-194, 2019.

[13] P. Gupta, "Use of machine learning techniques for predicting student placement success," *Int. J. Recent Technol. Eng.*, vol. 8, no. 4, pp. 304-308, 2020.

[14] A. W. Abed, M. M. Mansour, and A. Al-Absi, "Predictive models for student placement success using decision trees and neural networks," *Int. J. Appl. Eng. Res.*, vol. 12, no. 1, pp. 139-144, 2017.

[15] S. I. Ahmed, R. A. S. Shah, and H. H. Ali, "Analysis and prediction of placement using machine learning algorithms," *IEEE Access*, vol. 7, pp. 103261-103268, 2019.

[16] S. S. R. Sharma and M. G. R. M. Reddy, "A novel approach for student placement prediction using machine learning," *Proc. Int. Conf. Data Sci. Mach. Learn.*, pp. 103-109, 2019.

[17] J. R. Patel and M. S. Parekh, "Prediction of student placement based on performance and non-academic factors," *J. Emerg. Technol. Innov. Res.*, vol. 6, no. 8, pp. 101-108, 2019.

[18] M. B. Patel and P. S. Rawat, "A machine learning approach to student employability prediction," *Int. J. Comput. Sci. Eng.*, vol. 6, no. 12, pp. 39-45, 2018.

[19] A. M. Singh, R. K. Gupta, and A. K. Sharma, "Integration of non-academic factors in student placement prediction models," *J. Mach. Learn. Res.*, vol. 15, no. 6, pp. 105-110, 2020.

[20] S. N. Kumar, S. J. R. S. Sood, and M. Sharma, "Predicting student placement using support vector machines," *J. Data Min. Knowl. Discov.*, vol. 32, no. 4, pp. 1-10, 2018.

[21] M. I. Islam, M. F. Haque, and N. I. Mohammad, "An improved machine learning approach for student placement prediction," *J. Artif. Intell. Soft Comput. Res.*, vol. 12, no. 2, pp. 128-137, 2020.

[22] R. K. Soni, S. N. Agrawal, and P. G. Yadav, "A review of student placement prediction using data mining techniques," *Int. J. Comput. Sci. Mobile Comput.*, vol. 8, no. 2, pp. 116-122, 2019.

[23] J. K. Raj, S. R. B. S. Kumar, and T. P. Rao, "Predictive analytics for student placement success using machine learning algorithms," *Int. J. Comput. Intell. Res.*, vol. 8, no. 2, pp. 127-135, 2020.

[24] K. S. Patil, A. K. Kamat, and B. R. Patil, "An analytical approach to predicting student placement success using machine learning," *J. Comput. Appl.*, vol. 42, no. 9, pp. 1-7, 2018.

[25] S. S. Patil and K. D. Bhide, "Predicting student placement success based on academic and non-academic factors," *Proc. Int. Conf. Data Anal. Comput.*, pp. 234-238, 2020.

[26] D. K. Mahajan and S. S. Gupta, "Student placement prediction system using classification techniques," *J. Comput. Sci. Technol.*, vol. 7, no. 4, pp. 121-126, 2021.