International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 14, Issue 2, 2026

An Intelligent Machine Learning Framework for Accurate Prediction and Analysis of Campus Placement Outcomes

Sk.Shahina [1], Tripuramallu Lakshmi Aasritha [2], Davuluri Pravalika [3], Shaik Anwar Ahamad Basha [4], Kasukurthi Priyatham [5], Kunduru Ravindra [6]

[1]Assistant Professor, Department of CSE (AI & ML), Tirumala Engineering College,AP

[2,3,4,5,6] Students; Department of CSE (AI & ML), Tirumala Engineering College ,AP

Abstract: *Campus placement prediction has become an important application of machine learning for improving student employability and institutional placement performance. This paper presents an intelligent machine learning framework for predictive analysis of campus placement outcomes using academic, personal, and skill-based attributes such as CGPA, internships, backlogs, stream, and other relevant factors. The collected dataset is preprocessed through cleaning, encoding, and feature transformation techniques to improve data quality and model efficiency. Multiple machine learning algorithms including Decision Tree, Random Forest, Gradient Boosting, Hist Gradient Boosting, LightGBM/XGBoost, and CatBoost are trained and evaluated using performance metrics such as accuracy, precision, recall, and F1-score. Experimental analysis shows that the CatBoost model achieves the highest prediction accuracy among all compared models. A Flask-based web application is developed to provide a user-friendly platform where students can enter their details and instantly obtain placement predictions with useful insights. The proposed system helps students, educators, and placement officers make data-driven decisions and improve placement readiness.*

Index terms *- — Machine Learning, Campus Placement Prediction, Predictive Analytics, CatBoost, Random Forest, Gradient Boosting, Student Employability, Data Preprocessing, Classification Models, Flask Web Application.*

1. INTRODUCTION

Campus placements play a major role in shaping the career opportunities of students and the reputation of educational institutions. Predicting placement outcomes in advance helps students understand their readiness level and enables institutions to improve training strategies. Traditional placement evaluation methods mainly depend on academic marks or manual assessment, which may not accurately represent a student's employability potential. Therefore, intelligent data-driven systems are required to improve prediction accuracy and support better decision-making.

Machine learning provides powerful techniques for analysing historical student data and identifying patterns related to successful placements. By using factors such as CGPA, internships, technical skills, communication ability, backlogs, stream, and extracurricular achievements, predictive models can estimate the probability of a student getting placed. These models learn from past placement records and generate reliable predictions for future candidates.

This paper presents an intelligent machine learning framework for predictive analysis of campus placement outcomes. Multiple algorithms such as Decision Tree, Random Forest, Gradient Boosting, Hist Gradient Boosting, LightGBM/XGBoost, and CatBoost are implemented and compared to identify the best-performing model. Among them, CatBoost achieves superior performance with higher accuracy and better handling of categorical data.

A Flask-based web application is also developed to provide a user-friendly platform where students can enter their academic and personal details to obtain instant placement predictions. The proposed system assists students in improving employability skills and supports placement officers in making informed decisions. This framework contributes toward smarter, faster, and more efficient campus recruitment processes.

2. LITERATURE SURVEY

a) Placement Prediction and Analysis using Machine Learning

Campus placement is essential for all educational institutions to help students reach their goals. Large student datasets can have related data retrieved via machine learning classification. This test develops a predictive model that can predict the jobs for which students are qualified based on their prior academic and extracurricular accomplishments. In order to help students get ready for placement, the model will also suggest other skills that will be required for future hiring. Additionally, it provides ongoing trial outcomes and discoveries together with execution estimates anticipated for model approval, supporting the accomplishment of result-based training at

instructional foundations, which is acknowledged as the primary issue in the current setting.

b) Student Placement Prediction and Analysis using Machine Learning

Before finishing their degree, almost all students want to take advantage of a placement opportunity. The market offers a variety of existing and emerging placement options. To meet the demands of the company, a student must, however, carefully select his profession and skill set. Students can use a placement prediction model to forecast their chances of being placed or not depending on their academic and personal accomplishments. The five different machine learning classification methods that were taken into consideration for this project were the K-Nearest Neighbors [KNN] algorithm, Decision tree, Logistic regression, Naïve Bayes, and Random Forest. We assess the effectiveness of these algorithms based on the dataset, and each one independently predicts the results. Based on the student's credentials and job history, our prediction model may indicate how likely it is that he will get placed. Future academic planning for students or institutions may benefit from the use of such prediction models.

c) Placement Prediction using Various Machine Learning Models and their Efficiency Comparison:

The purpose of a placement predictor is to determine the likelihood that a student will be hired by a firm based on the company's requirements. The placement predictor uses a number of characteristics that may be utilized to determine the student's competence level. Certain factors are derived from tests carried out within the placement management system itself, while others are derived from university-level data.

d) PREDICTION OF STUDENT PLACEMENT USING MACHINE LEARNING

Traditionally, data analysis relied on trial and error, which is unfeasible for big and diverse data sets. As a component of data science, machine learning offers rapid fixes for analyzing vast amounts of data. by developing data-driven models and quick, well-structured algorithms for real-time data processing. Institutions exist to give their students exceptional possibilities. One supervised machine learning tool that can help with this is linear regression. Based on the results of advance placement practice tests, it is a useful technique for forecasting future trends in student placement. The outcome of this will provide the pupils a better understanding of their areas of weakness so they can improve. Students can increase their number of placements in an institution by focusing on these topics.

e) Campus Placement Prediction:

Increasing people's employability can help combat India's high unemployment rates. College graduates make up a sizable share of the 20–24 age bracket, which is one of the biggest unemployment groups. By adding courses and altering the curriculum to assist graduates gain the skills that employers value, colleges may significantly lower the percentage of jobless graduates. We developed a mechanism that assists in comparing the skill levels of students who are placed and those who are not. It forecasts the likelihood that a student with a certain skill set will be able to find employment. It considers additional soft skills that are necessary to get a job in addition to technical talents. The SVM model achieved an accuracy of 87%, whereas the XGBoost model achieved an accuracy of 90%. We discovered that the most crucial factors are the student's technical proficiency, projects, completed certification courses, and internships. The outcomes are encouraging and will undoubtedly raise college placement rates.

3. METHODOLOGY

i) Proposed Work:

The proposed work introduces an intelligent machine learning framework for predicting campus placement outcomes with high accuracy using student academic, personal, and skill-related attributes. The system collects important parameters such as CGPA, age, gender, stream, internships, backlog history, certifications, and extracurricular activities. The collected data is preprocessed through cleaning, encoding, normalization, and feature selection techniques to improve the quality of input data and overall model performance.

Several advanced machine learning algorithms such as Decision Tree, Random Forest, Gradient Boosting, Hist Gradient Boosting, LightGBM/XGBoost, and CatBoost are trained and evaluated using performance metrics like accuracy, precision, recall, and F1-score. Based on comparative analysis, CatBoost provides the best prediction results due to its efficient handling of categorical data and strong boosting capability.

A Flask-based web application is developed to provide a simple and interactive platform where students can enter their details and instantly receive placement predictions. The system also offers useful insights for skill improvement, helping students increase employability and assisting institutions in making data-driven placement decisions.

ii) System Architecture:

The proposed system follows the Model-View-Controller (MVC) architecture to ensure organized

development, scalability, and efficient communication between components. The View layer is developed using HTML and CSS, which provides a user-friendly interface where students enter academic and personal details such as CGPA, internships, stream, and backlog status. This layer captures user inputs and displays the final placement prediction results in an understandable format.

The Controller layer is implemented using Flask, which acts as the central coordinator between the user interface and machine learning model. It receives user requests from the View, validates input data, performs preprocessing, and forwards the processed data to the prediction engine. After receiving the predicted result, the controller sends the response back to the interface for display.

The Model layer contains the trained machine learning algorithms such as Decision Tree, Random Forest, Gradient Boosting, Hist Gradient Boosting, LightGBM/XGBoost, and CatBoost. This component processes the input features and generates placement predictions based on learned historical patterns. Among all models, CatBoost achieved the highest performance and is selected for final deployment. The MVC architecture makes the system modular, efficient, and easy to maintain.

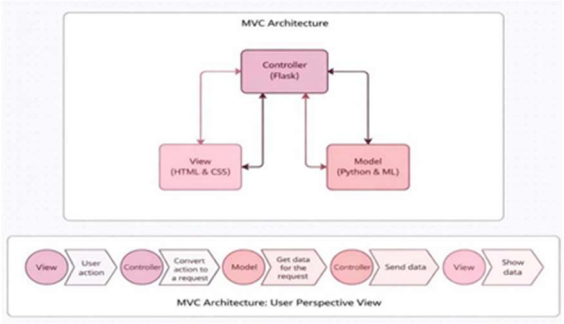

Fig1 System architecture

iii) Modules:

1. Data Collection Module

This module collects student-related data such as CGPA, age, gender, academic stream, internships, backlog history, certifications, and extracurricular activities. The collected dataset acts as the primary input for training and testing the placement prediction system.

2. Data Preprocessing Module

This module cleans the raw data by handling missing values, removing duplicates, encoding categorical attributes, and normalizing numerical values. Proper preprocessing improves data quality and enhances machine learning model performance.

3. Feature Selection Module

This module identifies the most relevant attributes that influence placement outcomes. It reduces unnecessary features, decreases computation time, and improves prediction accuracy by selecting important parameters.

4. Model Training Module

This module trains multiple machine learning algorithms such as Decision Tree, Random Forest, Gradient Boosting, Hist Gradient Boosting, LightGBM/XGBoost, and CatBoost using historical placement data. The system learns patterns from previous student records for prediction.

5. Prediction Module

This module accepts new student input data and uses the trained model to predict whether the student is likely to be placed or not placed. It generates quick and accurate placement probability results.

6. Web Application Module

This module provides a Flask-based user interface developed with HTML, CSS, and JavaScript. Students can enter details, submit data, and view prediction results through an interactive web platform.

7. Result Analysis Module

This module compares algorithm performance using metrics such as accuracy, precision, recall, and F1-score. It identifies the best-performing model, where CatBoost achieved the highest accuracy.

iv) Algorithms:

1. Decision Tree Algorithm

Decision Tree is a supervised learning classification algorithm that divides the dataset into smaller branches based on feature conditions. It creates a tree structure with root nodes, internal nodes, and leaf nodes to make predictions. It is easy to understand and suitable for campus placement classification tasks.

2. Random Forest Algorithm

Random Forest is an ensemble learning algorithm that combines multiple decision trees and gives the final result through majority voting. It improves prediction accuracy and reduces overfitting problems found in a single decision tree. It performs well for placement prediction using student datasets.

3. Gradient Boosting Algorithm

Gradient Boosting builds models sequentially where each new tree corrects the errors of the previous model. It uses boosting techniques to increase prediction performance and is highly effective for classification tasks with structured data.

4. Hist Gradient Boosting Algorithm

Hist Gradient Boosting is an advanced version of gradient boosting that converts continuous data into histogram bins before training. This reduces training time and computational complexity while maintaining strong prediction accuracy.

5. LightGBM / XGBoost Algorithm

LightGBM and XGBoost are optimized boosting algorithms designed for faster training and better

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 14, Issue 2, 2026

efficiency. They use parallel processing, regularization, and advanced tree optimization techniques to produce accurate predictions for placement outcomes.

6. CatBoost Algorithm

CatBoost is a boosting algorithm specially designed for handling categorical features efficiently. It uses ordered boosting and symmetric trees to reduce overfitting and improve model stability. In this project, CatBoost achieved the highest accuracy among all algorithms and was selected as the final prediction model.

4. EXPERIMENTAL RESULTS

The proposed campus placement prediction system was tested using a student placement dataset containing academic, personal, and skill-based attributes. The dataset was preprocessed using data cleaning, categorical encoding, and feature normalization techniques before model training. Multiple machine learning algorithms such as Decision Tree, Random Forest, Gradient Boosting, Hist Gradient Boosting, LightGBM/XGBoost, and CatBoost were implemented and evaluated using standard performance metrics including accuracy, precision, recall, and F1-score.

Experimental comparison shows that ensemble boosting algorithms performed better than basic classifiers due to their ability to learn complex patterns from placement data. Among all tested models, CatBoost achieved the highest prediction accuracy of 88.66%, making it the most reliable algorithm for campus placement prediction. Random Forest and Gradient Boosting also produced competitive results, while Decision Tree showed comparatively lower performance.

The developed Flask web application was also tested successfully by accepting user inputs such as CGPA, stream, internships, age, and backlog details, then instantly generating placement predictions. The results demonstrate that the proposed system is effective, accurate, and practical for helping students, faculty members, and placement officers make informed placement decisions.

Accuracy: The ability of a test to differentiate between healthy and sick instances is a measure of its accuracy. Find the proportion of analysed cases with true positives and true negatives to get a sense of the test's accuracy. Based on the calculations:

Accuracy = TP + TN /(TP + TN + FP + FN)

$$Accuracy = \frac{(TN + TP)}{T}$$

Precision: The accuracy rate of a classification or number of positive cases is known as precision. Accuracy is determined by applying the following formula:

Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$\Pr e\,cision = \frac{TP}{(TP + FP)}$$

Recall: The recall of a model is a measure of its capacity to identify all occurrences of a relevant machine learning class. A model's ability to detect class instances is shown by the ratio of correctly predicted positive observations to the total number of positives.

$$Recall = \frac{TP}{(FN + TP)}$$

mAP: One ranking quality statistic is Mean Average Precision (MAP). It takes into account the quantity of pertinent suggestions and where they are on the list. The arithmetic mean of the Average Precision (AP) at K for each user or query is used to compute MAP at K.

$$mAP = \frac{1}{n}\sum_{k=1}^{k=n} AP_k$$

$$AP_k = \text{the AP of class } k$$

$$n = \text{the number of classes}$$

F1-Score: A high F1 score indicates that a machine learning model is accurate. Improving model accuracy by integrating recall and precision. How often a model gets a dataset prediction right is measured by the accuracy statistic..

$$F1 = 2 \cdot \frac{(Recall \cdot \Pr e\,cision)}{(Recall + \Pr e\,cision)}$$

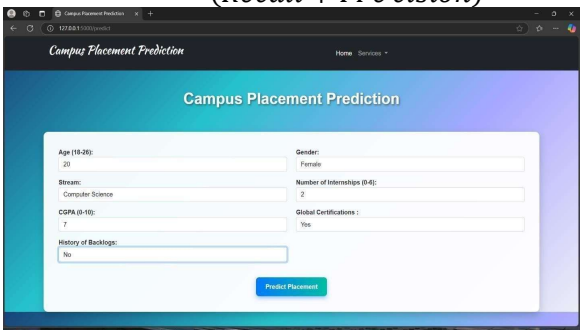

Fig 2 input data

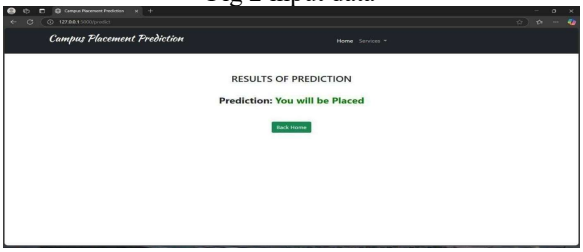

Fig3 results

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 14, Issue 2, 2026

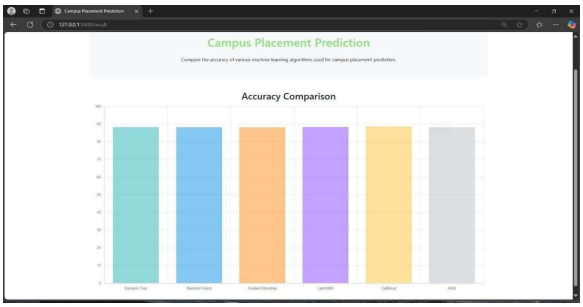

Fig 4 comparision graph

5. CONCLUSION

The study used machine learning algorithms to forecast student placement and company based on academic and personal traits.

A data collection of 215 Indian engineering college students with 13 attributes and two labels was employed. Data was preprocessed by managing missing values, encoding category categories, and scaling numerical features. Logistic regression, support vector classifier, K neighbors classifier, decision tree classifier, random forest classifier, and gradient boosting classifier were evaluated.

The methods were built in Python and scikit-learn and assessed using accuracy, precision, recall, F1-score, confusion matrix, feature significance, and permutation importance. Gradient boosting classifier has the highest accuracy, precision, recall, and F1-score of the six techniques.

The research also indicated that secondary education, degree, and job experience were the best predictors of placement status. The research also discovered that degree specialization, employability test percentage, and B.Tech specialization best predicted the firm.

A detailed and comparative review of placement prediction algorithms contributes to machine learning. The study also suggested ways for students and schools to improve academic performance, choose a degree and MBA focus, and obtain job experience. The study proposed adding data, feature engineering, algorithms, and labeling to future work.

6. FUTURE SCOPE

The placement prediction project will refine predictive models with advanced machine learning techniques, integrate dynamic features for trend analysis, incorporate external data sources for richer insights, develop personalized recommendation systems, evaluate long-term career trajectories, address ethical considerations in model development, deploy the model in real-world recruitment settings, and prioritize user-centric design.

Future machine learning research can enhance the placement prediction system in numerous ways. Adding student resumes and job descriptions to predictive models, fine-tuning model hyperparameters, and researching ensemble learning can improve forecast accuracy.

Deep learning architectures can capture complicated data patterns, whereas explainable AI can improve model interpretability. Dynamic updating and adaption methods keep the system relevant, while individualized recommendation algorithms tailor ideas to students' requirements.

Future research should evaluate performance on varied datasets, address ethical concerns, and improve user experience and interface design for usability and efficacy. The placement prediction system can enhance and help students choose careers through various channels.

REFERENCES

[1]. Naresh Patel K M, Goutham N M, Inzamam K A, “Placement Prediction and Analysis using Machine Learning” International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181, pp. 224 – 227, Special Issue – 2022.

[2].Pratiksha Khamkar, Rutuja Lagad, Priyanka Shinde, “Students Placement Prediction System International Journal for Research in Applied Science & Engineering Technology” (IJRASET) ISSN: 2321-9653, pp. 145 – 150, November – 2022.

[3].Irene Treesa Jose, Daibin Raju, Jeebu Abraham Aniya Kunju , “Placement Prediction using Various Machine Learning Models and their Efficiency Comparison” International Journal of Innovative Science and Research Technology (IJISRT) ISSN: 2456-2165, pp.1005-1009, 20 May – 2020.

[4].G.Gautami Monam Hayat, Pallab Banarjee, Biresh Kumar,“PREDICTION OF STUDENT PLACEMENT USING MACHINE LEARNING ALGORITHM”International Journal Of Creative Research Thoughts(IJCRT) ISSN: 2320-2882, pp.1379- 1384, 5 May – 2020.

[5]. S. K. Sharma and A. Sharma, “Placement Prediction using Machine Learning Algorithms,” International Journal of Engineering Research & Technology (IJERT), vol. 9, no. 3, pp. 175-180, 2020.

[6]. J. Shah, S. Kochrekar, N. Kale, S. Patil and A. Godbole, “Campus Placement Prediction,” in Recent Trends in Communication and Intelligent Systems, Springer,
pp. 369-378, 2022.

[7].A. Rao, S. K. Singh and A. K. Singh, “Student Placement Prediction Model: A Data Mining Perspective for Outcome-Based Education System,” International Journal of Recent Technology and Engineering (IJRTE), vol. 8, no. 3, pp. 7229-7234, 2019.

[8].Gupta, S., & Mehta S. “Predicting Student Placements Using Machine Learning Algorithms”.

ISSN 2347–3657
Volume 14, Issue 2, 2026

International Journal of Advanced Computer Science and Applications vol.11 sr. (2) pp.127-13, 2020.
[9]. Aggarwal, V., & Madan S., “Predicting Student Placements Using Machine Learning International Journal of Computer Sciences and Engineering”, vol.7 sr.(6), pp.136-141, 2019.
[10]. Manvitha, Pothuganti, and Neelam Swaroopa. "Campus placement prediction using supervised machine learning techniques." Int J Appl Eng Res vol.14.9 pp.2188- 2191, 2019.
[11].Thangavel, S.Bkaratki, P. Sankar, “Student placement analyzer: A recommendation system using machine learning”, Advances in Computing and Communication (ICACCS- 2017) International Conference on. IEEE, pp.101109,2017.