



Campus Placement Eligibility & Recruitment Outcome Prediction Using Academic and Behavioural Attributes

Mrs. M. Khamar¹, D. Sripriya², B. Vyshnavi³, D. Gowthami⁴, G. Sushma⁵

Assistant Professor, Department of Information Technology, KKR & KSR Institute of Technology and Sciences, Guntur, India¹

Student, Department of Information Technology, KKR & KSR Institute of Technology and Sciences, Guntur, India²⁻⁵

Abstract: Campus placement outcomes influence both students career and also the academic situation of education institutions. In the recent years, employment market has become very competitive and difficult for training and placement departments to assess the students accurately for the placements using traditional evaluation methods. To address this challenge, this system is presenting a machine learning-driven framework for predicting the campus placements and its eligibility by evaluating academic and behavioural attributes. This proposed approach uses features such as cumulative grade point average (CGPA), students technical skills, internship experience, and also communication capabilities to estimate placement probability. The system follows a structured pipeline that includes data collection, data preprocessing, feature selection, train test split, machine learning models, model evaluation and Placement predicted outcome. Data preprocessing techniques can be used to reduce this. Correct inconsistencies, encode categorical variables, and normalize numerical features, enabling the models to be built. Thus, a number of different classification algorithms are implemented and tested to see which works best Prediction. The framework is implemented in the python programming language and the pandas library is used. And analysis, and Scikit-learn can be used for modelling and evaluating the approach. Experimental Analysis shows that combining both academic and behavioural data can yield better results. Accuracy prediction as compared to models that only use the academic performance for prediction. The result experimental and communication-related features play an important role in improving employability and placements preparation. The proposed system provides actionable insights that can help training and placement departments to help identify students that need direct skill instruction and to target those needing specific skills programs or grouping like skilled students together for training. By leveraging machine learning and data-driven decision making, this research contributes to improving placement preparation strategies and optimizes institutional placement outcomes.

Index Terms: campus placement eligibility, python, pandas, Scikit-learn, classification algorithms.

I. INTRODUCTION

Campus placements are very important phase in a students academic journey and it serves as a key factor for higher educational institutions. Every student dreams of getting placed before their academics comes to end and they constantly thrive to perform better and their respective institutions are also constantly trying to make their student placement ready with various training programs. With the rapid growth of the job market and increasing competition among graduates, placement decisions have become more and more complex and multi-dimensional. Nobody wants to hire a average student with basic skills, everyone are looking for better candidates hence, institutions are always working hard to make their students stand out in the market with multiple skills. Traditional placement assessment methods rely solely on manual evaluation and academic scores which may not accurately reflect a student's employability and also such approaches are time-consuming and often fail to identify the students who require additional training support in time. By the time the institutions recognize the students who require extra care or on what additional variables students need to be trained on it's already end of their academics and this obviously reduces the rate of placements for institutions and also students fail to acquire job. With the availability of large volumes of educational data, machine learning techniques provide and effective solution for predicting placement outcomes in a data-driven manner. This paper proposes a machine learning-based system to predict campus placement eligibility using both academic and behavioural attributes.



II. LITERATURE REVIEW

Educational data mining has gained significant attention for analysing student performance and predicting the academic and career outcomes. Early Studies focused on traditional evaluation methods such as Rao et al. [3] applied traditional classification methods and showed that academic performance indicators significantly influence placement probability. Kumar et al. [4] extended this approach by integrating multiple machine learning algorithms and observed that models trained on academic performance data that can yield reliable predictions, particularly when features such as CGPA and technical skills are included. Classification techniques such as Naive Bayes Theorem, Decision Trees, Random Forest and Logistic Regression to predict the placement status. But these approaches achieved very reasonable accuracy, they rely on limited features sets restricting predictive capability. Recent research focus on the inclusion of behavioural attributes such as communication skills, internships, and technical skills and Most of the existing systems are using learning methods, particularly Random Forest Classifiers, Sharmila et al. [2] employed a Random Forest classifier and demonstrated improved prediction accuracy by leveraging ensemble learning, concluding that tree-based models handle complex student attributes effectively. this improved accuracy and robustness due to reduced overfitting and better handling of feature interactions. Survey-based research highlights the importance of combining both the academic and non-academic features to capture the employability.

Several works also stress the significance of proper data preprocessing which means the data cleaning, gaps filling and overall data manipulation and feature selection in improving model performance. Hirolikar et al. [5] conducted predictive analysis using supervised learning and emphasized that proper feature selection enhances model performance. However, many existing systems are limited to small datasets, which lacks interpretability, or insufficient focus on practical deployment. Foundational research in machine learning and data mining [11]–[13] provides essential theoretical support for applying supervised learning techniques in the educational domains.

This study is based upon existing literature by integrating both the academic and behavioural attributes within the structured predictive framework, evaluating multiple classifiers, and emphasizing actionable outcomes for placement preparation

III. METHODOLOGY

A. System Architecture

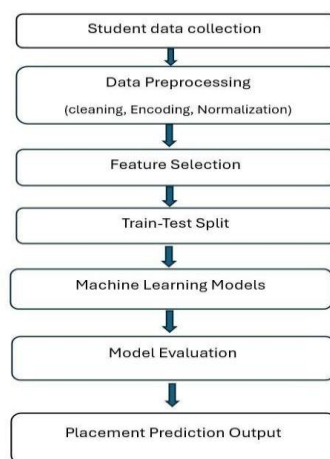


Fig. 1. Placement Prediction Using Machine Learning System Architecture

Input: Student Dataset D

Output: Placement Prediction

- 1) Collect student academic and behavioural data
- 2) Preprocess data (handle missing values, encoding, scaling)
- 3) Select relevant features
- 4) Split dataset into training and testing sets
- 5) Train classification models
- 6) Evaluate models using performance metrics
- 7) Predict placement outcome for new students



B. Dataset Description

Let the dataset be defined as: $\mathbf{D} = \{(X_i, y_i)\}^N$ where

- X_i represents the feature vector of the i^{th} student
- $y_i \in \{0, 1\}$ represents the placement outcome
- 1: Placed
- 0: Not Placed

This feature set consists of academic and behavioural patterns. Academic Attributes include CGPA (Cumulative grade point average). Behavioural Attributes include internship participations, communication skills, and other employability related characteristics. With the help of both academic and behavioural attributes, this dataset provides representation of important factors that directly or indirectly influence student placement outcomes.

C. Train - Test Split

After the cleaning and refinement of the dataset, the dataset is divided into two independent subsets as in training dataset and testing dataset to support the model development and evaluation. Generally, the larger portion of the data is allocated for training the model, where the learning algorithms study the relationship between student characters and placement outcomes as in which attribute is influencing the placement result in which manner which helps the machine model in forming decision patterns based on the training dataset. The test data is generally not exposed during training, so when this data is given to the machine model it gives genuine outcome based on the attributes given according to its decision patterns and predicts the student placement outcome. Maintaining a clear differentiation between the training and testing data will always give better results and gives accurate final placement predictions.

D. Data Preprocessing

Generally, the collected student data may often contain missing values and variations in a format that can influence model performance. If these are left unaddressed it affects the overall results. To overcome this, data preprocessing techniques are used before model training where all the missing values and variations and inconsistencies are addressed and modified appropriately to improve the overall data reliability. Certain attributes such as internships or skill level which can be categorical are converted into the numerical form which makes it easier for processing by machine learning algorithms. Numerical features like CGPA, attendance and previous academic percentiles are normalized for better results. These preprocessing operations help to create a clean, structured and complete dataset which makes it easier for the machine learning models to learn patterns and produce accurate predictions.

E. Feature Representation

Student is represented as a feature vector: $X_i = [x_{i1}, x_{i2}, \dots, x_{im}]$

where $X_i = [\text{CGPA}, \text{Internships}, \text{Technical Skills}, \text{Communication Skills}, \text{Attendance}]$

The features that are selected above reflects both the academic and behavioural attributes of a person which influences their placements. CGPA represents a student's academic history, Internships indicate work experience, Technical Skills represents Domain related knowledge along with these academic factors Behavioural attributes such as Communication skills which represents a person's nature and interpersonal ability professionally and Attendance which depicts discipline and commitment. All of these features are pre-processed and normalized to make sure uniform structured and effective learning by the machine learning models. This structured representation helps in predicting the accurate placement readiness of different student profiles.

F. Machine Learning Model Formulation

1) *Logistic Regression*: Logistic Regression is supervised machine learning algorithm which is commonly used for binary classification tasks. This algorithm helps in modelling the relationship between the input and output using sigmoid function for linear combination. In this campus placement prediction model, Logistic Regression estimates the probability of student being placed in campus placements based on their academic and behavioural attributes which include CGPA, internships, communication skills, technical skills and attendance. This algorithm helps in efficient and clear analysis of feature selection which influences the placement outcome. Because of its reliability and simplicity Logistic Regression is known for effective baseline model for placement outcome prediction. This algorithm plays a major role in estimating the probability of placement using sigmoid function.

$$p = \frac{1}{1 + e^{-z}}$$



The decision rule is based on

$$y = \begin{cases} 1 & \text{if } p \geq 0.5 \\ 0 & \text{otherwise} \end{cases}$$

2) *Random Forest Classifier*: Random Forest is a learning algorithm which improves classification by combining multiple decision trees. Each tree is trained based on randomly selected subset of dataset and the features. The final placement outcome prediction outcome is obtained by the majority voting between all of these decision trees that are combined which results in stable and accurate outcome. The various decision trees include different attributes like CGPA, Attendance, Communication skills, technical skills and internships by combining all these it trains the model. In placement prediction, Random Forest Classifier captures all the complex and non-linear relationships between the academic and behavioural attributes and how they influence the placement outcomes directly or indirectly. The Random Forest Classifier algorithm reduces the noise and overfitting making it more efficient and suitable for real-world datasets. It combines predictions from multiple decision trees and gives accurate outcome.

$$y = \text{mode}\{T_1(X), T_2(X), \dots, T_k(X)\}$$

where $T_k(X)$ denotes the prediction of the k^{th} decision tree

G. Model Evaluation

After the training the machine model using training dataset, performance of the developed machine learning models is very carefully observed using the testing dataset. Model Evaluation focuses on measuring how accurate the models can predict the placement outcome for each and every student. Accuracy, precision, recall and F1-score are used to assist different aspects of the machine model performance.

In Model Evaluation, By comparing all these measures across the multiple models, the system identifies most reliable classifier for the placement prediction. This evaluation process enables that selected model does not perform well only on training data but also gives the accurate outcome for the placement prediction when applied to the real-world student dataset.

H. Placement Prediction Output

The final stage of the system is generating the placement prediction outcome according to the trained and evaluated machine learning model. For each student record, this system produces classification about whether a student get placed or not placed. This output is derived from decision patterns that the model learning in training phases between the academic performance and behavioural attributes and the historical student details who are placed and not placed.

The prediction outcomes serve as a decision-support tool for training and placement departments to improve and enhance their training selectively for students based on missing skills that are required placements. By these outcomes the training departments can identify the students who require extra care and support for them to get placed. These outputs also help in monitoring overall placement readiness and planning focused training programmes and also gives students an overall view of their employability skills so that they can improve themselves, ultimately contributing to improved student employability and efficient placement strategies

I. Working

The proposed system is designed to predict campus placements as in to predict whether a student is likely to get placed or not and their eligibility with the help of academic and behavioural data such as CGPA, attendance, internships, technical skills, communication skills and participation in cocurricular activities. This raw data may contain missing or incorrect values sometimes so it is cleaned and normalized to remove the inconsistencies to make the data suitable for the analysis.

After the preprocessing, the next process starts by selecting important features that influence the placements are selected, then the dataset is divided into training and testing datasets. Using these test datasets and training datasets we train machine learning algorithms such as decision tree, Random Forest or Logistic Regression. The machine models are trained based on the historical student details and their behavioural attributes as in the characteristics of a student who got placed and a student who is not placed and form structured patterns including all the influencing attributes. These models learn patterns that differentiate placed and non-placed students.



After Training, The model is then tested using the testing dataset with new set of data which was not introduced to the model during the training so that it can observe and define a student characteristics based on the training data set through which the machine learning model formed learning patterns and predicts the outcomes as placed or not placed according to these learning patterns and if not placed then it also exhibits the skills that are lacking and where the student should focus on in order to get placed. After testing model evaluation is performed where this is rechecked in order to get accurate outcome for the real-world datasets. This also defines skill gap and lacking skills helping the institutions to identify the students who require extra guidance and they can also classify the students into groups with like skills and arrange better selective training programmes for students according to their learning ability without wasting time and efforts and helps the students to understand the placement readiness and areas for improvement. These results help the training and placement cell to plan effective preparation strategies to assist students in more accurate and right way and also gives a clarity for students to improve their skills according to the market requirements. This model doesn't focus only on academic but also on behavioural attributes.

IV. RESULTS AND PERFORMANCE EVALUATION

This model predicts if a student will be placed or not placed in campus placements based on their academic and behavioural attributes data. This system provides accurate results with good reliability and helps institutions to identify students who need extra care for better placements and train them accordingly. This system is evaluated using standard classification metrics

A. Accuracy

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

B. Precision

$$Precision = \frac{TP}{TP + FP}$$

C. Recall

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

D. F1-Score

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

This experimental results state that internships and communication skills play very crucial role in influencing campus placement outcomes along with the academic and behavioural attributes through CGPA, industrial exposure and good interpersonal skills which show major chances of securing placements. This observation states that current recruitment methodologies and the skills that market requires, where students gain job-ready skills along with the academic hike. These findings highlight the importance of assisting employability through both the academic and behavioural attributes rather than only relying on academic attribute.

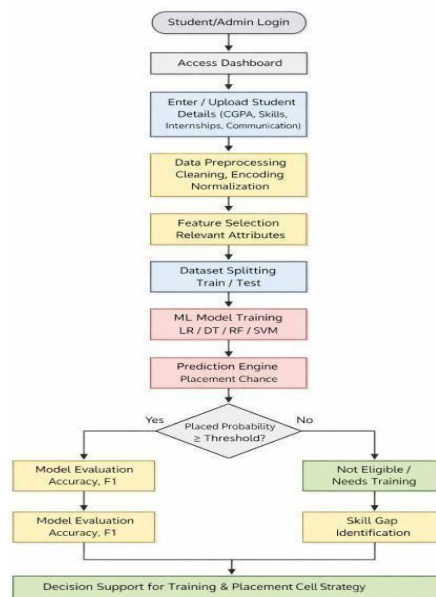


Fig 2. System Workflow Diagram

V. DISCUSSION

This system analysis shows that along with the academic attributes such as CGPA, internships and communication skills also have a very strong influence on placement outcomes. For a student to get placed there are various attributes that effects directly and indirectly behavioural attributes are one of them which influence students placements directly in one way or other. These days having only good number in theories doesn't get you nothing along with good academics it is very important for students to build their interpersonal skills such as communication skills and it is important for them to have industrial exposure once or twice through the internships. Students who gain exposure through these and have strong communication abilities are more likely to get placed during their academic year without much struggle because these are the skills that market is asking for nowadays and as long as you meet the market requirements you are good to work. This clearly highlights the increasing importance of practical and industry level skills along with the theoretical knowledge in today's recruitment process.

Through this system, we can understand that evaluating employability should not be limited to academics alone but also should include balanced involvement of both academic and behavioural attributes. Among the evaluated models, The Random Forest Classifier gives better generalized performance compared to the individual classification techniques. With the combination of multiple decision trees, Random Forest reduces the noise and overfitting and gives more reliable and accurate results while working with diverse student dataset. We use different data preprocessing methods to handle the data to fill the missing values and normalize the dataset and we use Python and the Scikit-learn library which allows efficient model development and testing.

This system helps in early identification of students who may need additional support and extra guidance for them to achieve placements and allowing placement cells to identify those students to introduce target training programs such as communication training, skill development and come up with more internship opportunities. Based on the data-driven insights, the framework reduces biased objective and improves transparency in decision-making. Overall, the integration of machine learning models with Python frameworks makes this system more suitable and reliable for real-world academic analysis giving reliably accurate results to improve placements and employability skills.

VI. CONCLUSION

This study focuses on introducing a machine learning-based framework which is designed to predict campus placement eligibility by considering both the academic and behavioural attributes of students including their academic history through CGPA and internships, communication skills, technical skills and attendance. By combining all these factors which basically includes their grades, skills and overall student involvement and commitment, this system improves the accuracy of placement predictions in a practical and more meaningful manner



This framework helps in training the machine models to assist the training and placement cells to gain early insights about students placement readiness and also identifying the students who require extra guidance in early stages itself rather than regretting at the end of their academic year. This helps the training departments to train the students according to their skill gap and group them for targeted training programmes which helps students to improve their skills well in advance. Using the educational data analytics and various classification techniques, this system reduces the need of time-consuming manual evaluations and supports more reliable and data-driven evaluation and enhanced decision-making.

Another Advantage of this proposed approach is its reliability, flexibility and scalability, which makes it more suitable for real-world educational data analysis. By identifying skill gaps in the earlier stages, institutions can prepare their students in a better way for the industrial recruitment processes. Future enhancements may include the usage of deep learning models, real-time performance tracking and training larger datasets for further accuracy in prediction, robustness and practical usefulness. Based on the results, it can be concluded that the proposed system effectively predicts student placement outcomes based on both the academic and behavioural attributes. This system gives a practical and reliable approach for the early placement readiness assistance and supports focused training programmes.

REFERENCES

- [1]. I. A. Margay and G. K. Sodhi, "Predicting student placement in college using machine learning," *International Journal for Research in Applied Science & Engineering Technology*, vol. 12, no. 2, pp. 1384–1387, 2024.
- [2]. S. Sharmila, R. Karthik, and P. Anitha, "Campus placement prediction using random forest classifier," *Journal of Emerging Technologies and Innovative Research*, vol. 11, no. 11, pp. 122–127, 2024.
- [3]. A. S. Rao, S. R. Patil, and M. Kulkarni, "Student placement prediction model using data mining techniques," *International Journal of Recent Technology and Engineering*, vol. 8, no. 3, pp. 2497–2499, 2019.
- [4]. M. Kumar, R. Singh, and S. Verma, "Predicting college students placements based on academic performance using machine learning approaches," *International Journal of Modern Education and Computer Science*, vol. 15, no. 6, pp. 1–9, 2023.
- [5]. D. S. Hirolikar, A. J. Patil, and S. R. Jadhav, "Campus placement predictive analysis using machine learning," *International Journal of Creative Research Thoughts*, vol. 12, no. 5, pp. 264–267, 2024.
- [6]. A. Pathak, N. Sharma, and P. Gupta, "A machine learning framework for predicting student placement outcomes," *Ingenierie des Systems d'Information*, vol. 30, no. 7, pp. 1715–1721, 2025.
- [7]. S. Kumar and R. Thangavel, "Applications of data mining techniques in educational systems: A survey," *International Journal of Computer Applications*, vol. 10, no. 3, pp. 1–5, 2010.
- [8]. T. Byagar and R. M. Shastri, "Campus placement prediction using classification algorithms," *Procedia Computer Science*, vol. 172, pp. 321–326, 2020.
- [9]. R. Agrawal and A. Gupta, "Educational data mining for employability prediction using machine learning," *International Journal of Engineering and Advanced Technology*, vol. 9, no. 4, pp. 546–551, 2020.
- [10]. P. S. Patil and S. A. Shinde, "Student performance and placement prediction using supervised learning techniques," *International Journal of Computer Sciences and Engineering*, vol. 8, no. 6, pp. 45–50, 2020.
- [11]. J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. San Francisco, CA, USA: Morgan Kaufmann, 2012.
- [12]. T. M. Mitchell, *Machine Learning*. New York, NY, USA: McGraw-Hill, 1997.
- [13]. I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*, 4th ed. Burlington, MA, USA: Morgan Kaufmann, 2016.
- [14]. S. B. Kotsiantis, "Supervised machine learning: A review of classification techniques," *Informatica*, vol. 31, no. 3, pp. 249–268, 2007.
- [15]. P. Cortez and A. Silva, "Using data mining to predict secondary school student performance," *IEEE Transactions on Education*, vol. 52, no. 1, pp. 1–8, Feb. 2009.