

# Forecasting Modelling for Campus Placement: A Comparative Study of Machine Learning Algorithms

Shivangi Srivastava<sup>1</sup>, and Er. Vishal Bharati<sup>2</sup>

<sup>1,2</sup> Department of Computer Science and Engineering, Suyash Institute of Information Technology,  
Gorakhpur, Uttar Pradesh, India

Correspondence should be addressed to Shivangi Srivastava; [shivangisrii1409@gmail.com](mailto:shivangisrii1409@gmail.com)

Received: 16 March 2026;

Revised: 31 March 2026;

Accepted: 14 April 2026

Copyright © 2026 Made Shivangi Srivastava et al. This is an open-access article distributed under the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**ABSTRACT-** In this research paper, we investigate the use of machine learning algorithms to predict campus placement outcomes. Our study uses a dataset that includes various attributes of students, such as academic performance, background information, and employability test scores. By conducting extensive data pre-processing, exploratory data analysis (EDA), and predictive modelling techniques, we aim to create accurate predictive models and compare the performance of different algorithms. In addition, we have developed a recommendation system that offers actionable insights to students based on the model predictions.

The main objective of this placement prediction model is to develop skills in students, primarily by improving their academic performance and exam results. There by increasing their chances of getting placed. The use and study of such machine learning and AI-based models also helps colleges and educational institutions improve their placement departments and develop new technologies. This type of study also assists the faculty of these institutions in training students accordingly.

This research helps institutions understanding students learning pace and skill development track their progress and determine the requirements of recruiting companies or organizations, ensuring that these needs are met.

**KEYWORDS-** Campus Placement Prediction, Machine Learning, Educational Data Mining, Student Employability Forecasting, Predictive Analytics, Classification Models.

## I. INTRODUCTION

Today, in the era of big data, machine learning has proved to be a highly efficient technique for accurate prediction of student placement. It can be done by examining a wide range of data sets, including academic performance, skills, and previous placement history. Traditionally, this technique of predicting student placement has always been a time-consuming and subjective task. However, the introduction of machine learning as a technique for student placement prediction has opened doors for a more objective approach to this task.

The main objective of this research paper is to explore how machine learning models can be used to make precise predictions in student placement. This can be achieved by using vast data sets of student records, university data, and assessment data from the placement management system. Machine learning models can analyze these data sets to make precise predictions. In this way, it is possible to assess which factors are affecting student placement success.

This research paper focuses on the applicability of machine learning algorithms for predictive modelling in campus placement. Through an in-depth analysis of student information that includes academic performance, student background, and employability, this paper attempts to create effective predictive models that can predict student placement with accuracy. By making effective use of the predictive ability of machine learning algorithms, this paper attempts to achieve the following objectives:

### A. Evaluation of Predictive Models

The main objective of this paper is to evaluate the effectiveness of various machine learning algorithms in prediction. By comparing the effectiveness of various algorithms such as Logistic Regression, Support Vector Machines, Decision Trees, Random Forest, and Gradient Boosting, it is possible to identify the best algorithm for placement prediction.

### B. Exploration of Feature Importance

Moreover, it is essential to understand what affects placement in order to guide students effectively. Feature importance analysis is also conducted in the study in order to understand what determines placement success. This provides insights into academic, personal, and experiential aspects that affect placement.

### C. Development of Recommendation Systems

Besides predicting the placement outcomes, the paper also aims at developing recommendation systems that are capable of providing insights to the students using the results of the predictive modelling. This is expected to empower the students in their pursuit of placement opportunities through the provision of appropriate recommendations for skill development, networking, and career progression.

## II. LITERATURE REVIEW

The prediction of campus placement opportunities is one of the most important issues in the context of educational institutions and students. Machine learning and recommendation systems have been recognized as important tools in the context of predictive purposes. In this context, this literature review provides an overview of the recent advancements in the application of machine learning and recommendation systems in the context of predicting student placements.

- Romero and Ventura [1] provided a comprehensive review of Educational Data Mining techniques, including classification, clustering, and recommendation systems. Their study highlighted the importance of predictive modelling in improving learning outcomes.
- Michael P. O'Mahony et al.[2] describes the development of a recommender system for the online enrolment application at University College Dublin. The study covers the factors that affect student decisions and offers solutions to the key considerations in the selection process. The study uses historical student enrolment data to evaluate the system and finds that it holds promising results in the decision process for students.
- Irene Treasa Jose et al.[3] focuses on placement prediction system using machine learning methods. This work aims to help educational institutions make better predictions about student placement by estimating the probability that a given student will get placed. They employ various machine learning classification algorithms (e.g., K-Nearest Neighbors, Support Vector Machine) to study student information and placement likelihoods. This prompt can be used for assessing different parameters (academic, skills, etc.) and compares how efficient the model is for prediction tasks.
- Olaf Zawacki-Richter et al[4]., is a thorough review of the existing literature in the field. The paper analyzes 58 different studies where machine learning has been implemented to profile and forecast educational outcomes, which include decision-making about admissions, course planning, and student performance. It is highlighted that the majority of the research is quantitative in nature and that different machine learning algorithms are implemented to perform tasks such as pattern recognition.
- Baker and Invent ado [5] discussed the role of learning analytics and machine learning in predicting student performance, dropout risk, and engagement levels, emphasizing data-driven decision-making.
- Cortez and Silva [6] applied data mining algorithms to predict student performance in secondary education using classification techniques such as Decision Trees and Neural Networks.
- Kotsiantis [7] investigated the use of machine learning methods for forecasting student grades and demonstrated the impact of algorithm selection on prediction accuracy.
- Al-Barrak and Al-Razgan [8] utilized decision tree models to predict students' final GPA,

demonstrating the effectiveness of predictive analytics in identifying students at risk and supporting early intervention strategies.

- Shreyas Harinath et al. [9] proposed a classification algorithm was applied in this study to predict whether a student's placement would be positive or negative. This complementing with the existing work focuses on data-driven methods in enhancing placement process efficiency and making it easier for institutions to locate students who need improvement training. The results showed that machine learning models let you find accurate predictions and support better decision-making in campus recruitment.
- Sheikh Adil Habib et al.[10] proposed that fossil modelling would provide more high accuracy for wind, renewable power, and hydropower is produced. Natural gas plant estimates are also improving, although the margin of error remains considerable, especially for smaller facilities.
- S. Banerje et al. [11] they have used five different machine learning algorithms to classy organic food.
- M.S. Spoorthi et al. [12] provided an extensive overview of placement management systems, emphasizing the aspects related to the architecture of these systems, the functions performed by these systems, and the advantages associated with them. This paper elaborates on the role of digital tools in optimizing the placement process through the effective management of student data, firm requirements, and recruiting activities.

Each of these studies helps build the corpus of knowledge in terms of machine learning and recommendation systems and how they can be applied to improve educational outcomes, with a specific focus on how they can be used to optimize student placement. They demonstrate a range of methodologies, evaluations, and implementations of these types of systems within the educational environment. Collectively, these studies form a basis for future research in the area of predictive modelling and recommendation systems within higher education.

## III. PROPOSED METHODOLOGY

In this study, machine learning methods are used to predict a student's placement status based on historical placement information. The predictive parameters considered for this study are high school percentage, senior secondary percentage, B.Tech percentage, previous work experience, e-test percentage, certifications, and the number of current backlogs. To perform the predictions for a student's placement, a range of machine learning algorithms are implemented, including Logistic Regression, Random Forest Classifier, K-Nearest Neighbors Classifier, Support Vector Machines, Decision Tree Classifier, and Gradient Boosting Classifier.

Depending on the predictions, recommendations are provided to the students that help them improve their chances of getting placed in a good organization.

### A. Data Collection and Pre-processing

Dataset Description: In this study, a dataset from a

university's placement cell is used. The information in the dataset contains various attributes that are important for predicting the placement of a student. The attributes of the dataset are secondary school percentage (ssc\_p), higher secondary percentage (hsc\_p), percentage of the degree (degree\_p), percentage from employability tests

(etest\_p), whether the student possesses work experience (Workex), whether the student possesses any certifications, gender of the student, and whether the student got a placement or not (See the table 1).

Table 1: Dataset Description

sl_no	gender	ssc_p	ssc_b	hsc_p	hsc_b	degree_p	Branch	Workex	Certifications	etest_p	Backlogs	status	salary	
0	1	M	67.00	Others	91.00	Others	58.00	CSE	No	4	55.00	0	Placed	270000.0
1	2	M	79.33	Central	78.33	Others	77.48	CSE	Yes	1	86.50	0	Placed	200000.0
2	3	M	65.00	Central	68.00	Central	64.00	ECE/EN	No	2	75.00	1	Placed	250000.0
3	4	M	56.00	Central	52.00	Central	52.00	CSE	No	0	66.00	2	Not Placed	NaN
4	5	M	85.80	Central	73.60	Central	73.30	ECE/EN	No	3	96.80	0	Placed	425000.0
5	6	M	55.00	Others	49.80	Others	67.25	CSE	Yes	0	55.00	1	Not Placed	NaN
6	7	F	46.00	Others	49.20	Others	79.00	ECE/EN	No	0	74.28	1	Not Placed	NaN
7	8	M	82.00	Central	64.00	Central	66.00	CSE	Yes	1	67.00	0	Placed	252000.0

Data Pre-processing: Before constructing the model, certain steps are taken for the pre-processing of the data. The missing values are handled, and all the values are present. The columns that are not important for the prediction of the placement is removed from the dataset. The columns that are removed are sl\_no and salary. The categorical values are encoded and then converted to a numerical format. The categorical values that are encoded are ssc\_b, hsc\_b, Branch, Worker, and gender.

**B. Exploratory Data Analysis (EDA)**

Descriptive Statistics: The primary exploration of the dataset started by computing a number of summary statistics that would help us understand the central tendency and variability of the features. Next, a detailed analysis of the features and how they are distributed was carried out, such that the shapes of the features, spreads, and skewness could be understood. Additionally, outliers of the features were identified by making use of box plots.

- The box plots are designed to display the various distributions of key metrics such as secondary

school percentage (ssc\_p), higher secondary percentage (hsc\_p), undergraduate percentage (degree\_p), and employability percentage (etest\_p). By displaying the data in such a way, it becomes easier to detect outliers in the data, as well as get a better understanding of the spread of the data for each of the metrics.

- The outliers existing in the data in the hsc\_p column are identified and removed using the Interquartile Range (IQR) method. Once again, data is visualized in order to observe how the data has changed due to the removal of outliers.
- Correlation Analysis: In this section, a correlation analysis was carried out. The purpose of this analysis was to understand the relationship that exists between the various features and the target variable, which in this case was the status. To better understand the relationship that exists between these features and the target variable, a heatmap was created. The heatmap displays the correlation that exists between all the features.

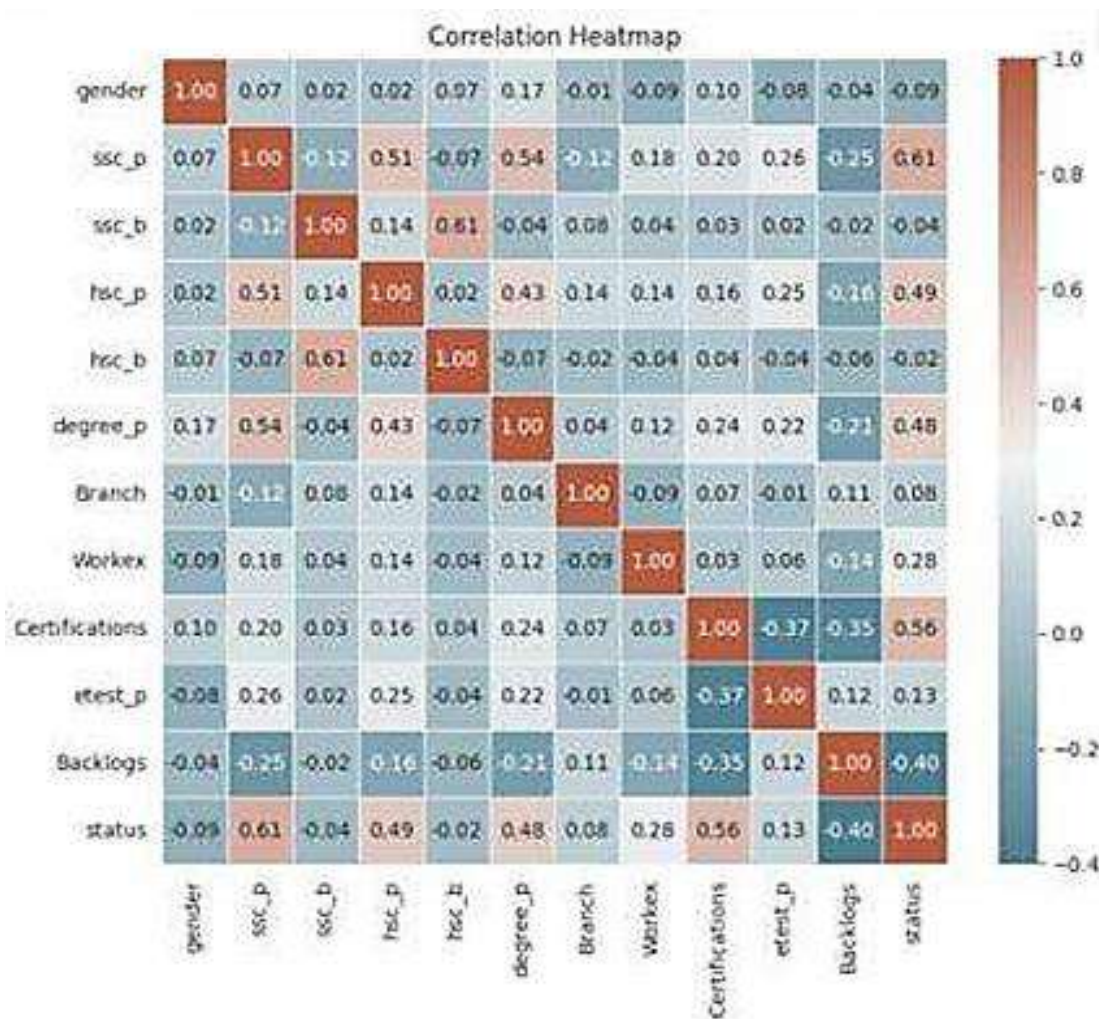


Figure 1: Correlation Heatmap

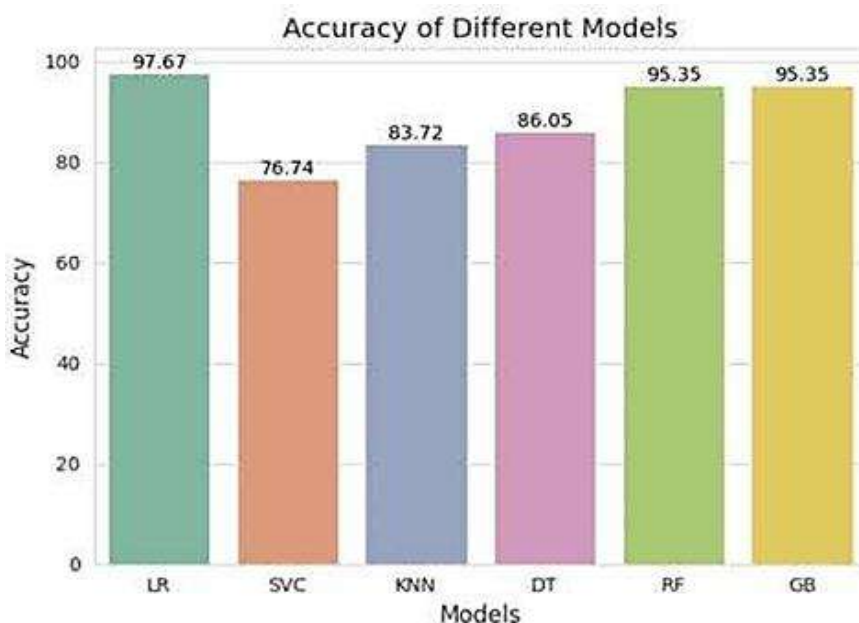


Figure 2: Accuracy of different Models

**C. Predictive Modelling**

- Model Selection: For the predictive modelling task, six machine learning algorithms were selected to train models on the data and compare their

performance. These models include logistic regression, support vector machines, k-nearest neighbors, decision trees, random forests, and gradient boosting. Each of them has its advantages

in solving the problem, which allows for a comprehensive analysis of their performance on the data.

- **Model Training and Evaluation:** The dataset was split into separate subsets for training and testing show in [Figure 1](#). Each algorithm was trained on the training subset, followed by evaluation on the testing subset, with accuracy as the sole measure of effectiveness for all the algorithms.

#### IV. RESULTS AND DISCUSSION

- **Model Performance:** The performance of each algorithm in terms of effectiveness was determined based on the accuracy score of each algorithm when tested. The comparative analysis showed [Figure 2](#) that different algorithms had different predictive capabilities
- Our findings indicate that the Logistic regression classifier and Random Forest classifier performed best in terms of predictive accuracy, achieving an accuracy score of 97.67% and 95.35% respectively. Feature importance analysis revealed that in random forest classifiers certifications, employability test, academic performance in secondary and higher secondary education, and prior work experience, was significant predictors of placement success. By identifying these key predictors, our model empowers educational institutions to make more informed decisions and provide targeted support to students.
- **Recommendation System: Recommendation System Implementation:** A recommendation system was developed based on the predictive model's output to provide personalized recommendations to students based on their placement prediction.

#### V. CONCLUSION

Our study aimed to develop a predictive model for student placement based on historical placement data. Through the analysis of various machine learning algorithms and feature engineering techniques, we have gained valuable insights into the factors influencing student placement outcomes.

This study provides findings that offer further insight into the degree of effectiveness that is possible with machine learning algorithms in relation to forecasting placement results. By examining the data provided on the effectiveness of different models, it is possible to understand more nuanced levels of effectiveness that could be applied in relation to helping educational institutions understand placement results. Therefore, there is a significant degree of value that could be applied in relation to helping inform decision-making processes with placement assistance services.

Furthermore, the integration of a recommendation system adds a practical dimension to our research, offering personalized guidance to students on how to enhance their placement prospects. By leveraging the insights gleaned from our predictive model, students can proactively take steps to strengthen their qualifications and increase their chances of securing placement opportunities.

While our study has yielded valuable findings, it is not without its limitations. The scope of our research is confined to a specific dataset and may not fully capture the nuances present in other educational contexts. Future research endeavors could explore the scalability and generalizability of our model across diverse student populations and institutional settings.

#### VI. FUTURE WORK

Future research could explore advanced machine learning techniques, ensemble methods, and feature engineering strategies to further improve predictive accuracy and robustness.

#### CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

#### REFERENCES

- [1] Romero, C. and Ventura, S., "Educational data mining: A review of the state of the art," *IEEE Transactions on Systems, Man, and Cybernetics—Part C: Applications and Reviews*, vol. 40, no. 6, pp. 601–618, 2010. Available from: <https://ieeexplore.ieee.org/abstract/document/5524021>
- [2] O'Mahony, M. P. and Smyth, B., "A recommender system for online course enrolment: An initial study," in *Proc. ACM Conf. Recommender Systems (RecSys '07)*, 2007, pp. 133–136. Available from: <https://doi.org/10.1145/1297231.1297254>
- [3] Jose, I. T., Raju, D., Aniyam Kunju, J. A., James, J., and Vadakkal, M. T., "(Title not specified)," *International Journal of Innovative Science and Research Technology*, vol. 5, no. 5, May 2020.
- [4] Zawacki-Richter, O., Marín, V. I., Bond, M., and Gouverneur, F., "Systematic review of research on artificial intelligence applications in higher education—Where are the educators?" *International Journal of Educational Technology in Higher Education*, vol. 16, no. 1, Art. no. 39, 2019. Available from: <https://link.springer.com/article/10.1186/s41239-019-0171-0>
- [5] Baker, R. S. and Inventado, P. S., "Educational data mining and learning analytics," in *Learning Analytics: From Research to Practice*, J. A. Larusson and B. White, Eds. Springer, 2014, pp. 61–75. Available from: <https://doi.org/10.1002/9781118956588.ch16>
- [6] Cortez, P. and Silva, A., "Using data mining to predict secondary school student performance," in *Proc. 5th Future Business Technology Conf. (FUBUTEC 2008)*, 2008, pp. 5–12. Available from: <https://tinyurl.com/4s9ykzkz>
- [7] Kotsiantis, S. B., "Use of machine learning techniques for educational purposes: A decision support system for forecasting students' grades," *Artificial Intelligence Review*, vol. 37, no. 4, pp. 331–344, 2012. Available from: <https://link.springer.com/article/10.1007/s10462-011-9234-x>
- [8] Al-Barrak, M. A. and Al-Razgan, M., "Predicting students' final GPA using decision trees: A case study," *International Journal of Information and Education Technology*, vol. 6, no. 7, pp. 528–533, 2016. Available from: <https://tinyurl.com/3j7w8rdj>
- [9] Harinath, S., Prasad, A., Suma, H., and Suraksha, A., "Student placement prediction using machine learning," *International Research Journal of Engineering and Technology (IRJET)*, vol. 6, no. 4, Apr. 2019. Available from: <https://tinyurl.com/ycyzhezmm>

- [10] S. A. Habib and D. Kumar, "Exploratory data analysis of global power plants using various machine learning algorithms," *International Journal of Innovative Research in Computer Science and Technology (IJIRCST)*, vol. 10, no. 4, pp. 52–61, 2022. Available from: <https://doi.org/10.55524/ijircst.2022.10.4.7>
- [11] S. Banerjee, D. Palsani, and A. C. Mondal, "Forecasting organic crops based on machine learning algorithms," *International Journal of Innovative Research in Engineering and Management (IJIREM)*, vol. 11, no. 5, pp. 98–103, 2024. Available from: <https://doi.org/10.55524/ijirem.2024.11.5.14>
- [12] Spoorthi, M. S., Kavana, V., Koushik, M. N., and Veena, M., "A review on placement management system," *International Journal of Creative Research Thoughts (IJCRT)*, vol. 9, no. 7, Jul. 2021. Available from: <https://www.ijraset.com/files/serve.php?FID=32641>