JRITM
JOURNAL OF RESEARCH AND INNOVATION IN TECHNOLOGY, COMMERCE AND MANAGEMENT

Journal of Research and Innovation in Technology, Commerce and Management
Vol. 3 Issue 9, September 2026, pp. 39046-39053
ISSN: 3049-3129 (Online)

College Student Performance Analytics and Placement Prediction Using Machine Learning

S. V. Mankar
Department of Artificial Intelligence and Data Science,
Pravara Rural Engineering College, Loni
Email: snehal.eleventh@gmail.com

S. B. Shirsath
Department of Artificial Intelligence and Data Science,
Pravara Rural Engineering College, Loni
Email: supriya.shirsath8@gmail.com

Cite as: S. V. Mankar, & S. B. Shirsath. (2026). College Student Performance Analytics and Placement Prediction Using Machine Learning. Journal of Research and Innovation in Technology, Commerce and Management, Vol. 3(Issue 9), 39046–39053. https://doi.org/10.5281/zenodo.22702108

DOI: https://doi.org/10.5281/zenodo.22702108

Abstract— Educational institutions generate large volumes of student data through academic records, attendance, assessments, skill development activities and placement processes. However, effective utilization of these data for academic and placement decision making remains challenging. This paper presents a Student Performance Analytics and Placement Prediction Framework integrating data analytics, machine learning, clustering and business intelligence visualization. The framework uses an initial dataset of 75 student records and 22 attributes containing academic, behavioral, employability, extracurricular and student-risk related information. A Placement variable is subsequently derived using predefined Previous CGPA, Aptitude Score and Coding Score criteria for the placement classification experiment. Linear Regression is applied for academic performance prediction, Logistic Regression for placement classification, Decision Tree for performance classification, and K Means for student segmentation. A Power BI dashboard is developed to visualize placement outcomes, academic performance, attendance, employability skills, performance categories and dropout risk. Logistic Regression achieved an accuracy of 93.33 percent with a weighted F1 score of 0.93 on the evaluated test set, while Linear Regression obtained an MSE of 1.5467 and an R^2 score of -0.4373. Decision Tree classification achieved an accuracy of 40.00 percent. Since the placement labels were derived using predefined academic and technical criteria, the placement classification result should be interpreted as an evaluation of the defined rule-based labeling scheme rather than as evidence of generalized real-world placement prediction.The results indicate the potential of the proposed framework for student segmentation, placement analysis and data driven academic decision support, while the limited dataset size restricts generalizability.

Keywords— student performance analytics; placement classification; machine learning; educational data analytics; Power BI; student segmentation; academic performance

I. Introduction

The increasing availability of digital educational data has created opportunities for applying data-

Vol. 3 Issue 9, September 2026, pp. 39046-39053

ISSN: 3049-3129 (Online)

driven methods to understand student learning and academic outcomes. Educational Data Mining (EDM) has developed as a research area concerned with extracting meaningful patterns from educational data and applying computational methods to educational problems **[1]**, **[2].** Recent EDM research has placed increasing emphasis on prediction and model-based analysis for understanding educational processes and outcomes **[2].**

Machine learning provides a set of supervised and unsupervised techniques that can be applied to prediction, classification, and pattern discovery tasks. Libraries such as Scikit-learn provide implementations of machine learning algorithms for supervised and unsupervised learning, supporting their application in practical data-analysis workflows **[5]**. In the context of student analytics, academic performance and employability can be examined using attributes such as attendance, marks, previous CGPA, aptitude, coding ability and other skill-related indicators.

The present work develops a College Student Performance Analytics and Placement Prediction Framework that integrates data analytics, machine learning, student segmentation, and business intelligence visualization. The framework uses an initial dataset of 75 student records and 22 attributes containing academic, behavioral, employability, extracurricular and student-risk related information. A Placement variable is subsequently derived using predefined Previous CGPA, Aptitude Score and Coding Score criteria for the placement classification experiment. Linear Regression is applied for Previous CGPA estimation, Logistic Regression for placement classification, Decision Tree for Performance Category classification, and K Means for student segmentation.

The main contribution of this paper is as follows:

i. A College Student Performance Analytics and Placement Prediction Framework is developed by integrating Data Analytics, Machine Learning, clustering and business intelligence visualization for analyzing academic performance and placement outcomes.

ii. Multiple machine learning techniques are applied for different analytical tasks, including Linear Regression for academic performance estimation, Logistic Regression for placement classification, Decision Tree for performance classification and K Means for student segmentation.

iii. A Power BI dashboard is developed to provide interactive visualization of academic performance, attendance, placement outcomes, employability skills, performance categories and dropout risk**[10].**

iv. Experimental evaluation is performed on a dataset containing 75 student records. Experimental evaluation is performed on a dataset containing 75 student records. Logistic Regression achieved 93.33 percent accuracy with a weighted F1 score of 0.93 on the evaluated test set, while Linear Regression and Decision Tree demonstrated limited performance with the selected features.

II. DATASET AND FEATURE DESCRIPTION

A. *Dataset Description*

The initial dataset comprises 75 student records and 22 attributes containing academic, behavioral, employability, extracurricular and student-risk related information. Academic attributes include Marks, Attendance, Study Hours, Study Hours Per Week, Backlogs and Previous CGPA, while employability attributes include Aptitude Score, Coding Score, Mock Interview Score, Resume Score and Soft Skills Score. Student development is further represented through Leadership, Teamwork, Club Participation, Hackathon Count and Library Usage. A Placement variable is subsequently derived using predefined criteria

Vol. 3 Issue 9, September 2026, pp. 39046-39053
ISSN: 3049-3129 (Online)

based on Previous CGPA, Aptitude Score and Coding Score for the placement classification experiment.Previous CGPA is considered as the target variable for regression analysis, the derived Placement variable is used for binary classification, and Performance Category is used for student classification. Dropout Risk is used as an analytical indicator for identifying students requiring additional academic support.

B. *Feature Description*

The dataset consists of numerical and categorical features representing different dimensions of student performance. Numerical features include academic scores, attendance, study effort, aptitude, coding ability, interview performance, resume quality, soft skills, teamwork and extracurricular participation. Categorical features include Gender, Leadership, Club Participation, Dropout Risk and Performance Category. The Placement variable is subsequently derived from predefined Previous CGPA, Aptitude Score and Coding Score criteria for binary classification. The combination of these attributes enables joint analysis of academic performance, employability and placement readiness.

TABLE I MAJOR DATASET FEATURE

Feature Group	Representative Attributes
Academic	Marks, Attendance, Study Hours, Backlogs, Previous CGPA, Improvement Rate
Aptitude and Technical	Aptitude Score, Coding Score
Employability	Mock Interview Score, Resume Score, Soft Skills Score
Participation	Leadership, Teamwork, Club Participation, Hackathon Count
Academic Resources	Library Usage
Target and Risk	Previous CGPA,Derived Placement, Performance Category, Dropout Risk

C. *Data Preprocessing*

The dataset was checked for missing values, and numerical missing values were handled using mean imputation. Categorical attributes were converted into numerical representations where required for analysis and machine learning. A Placement variable was derived using predefined Previous CGPA, Aptitude Score and Coding Score criteria. The dataset was divided into training and testing subsets using an 80:20 ratio for supervised learning experiments.

D. *Exploratory Data Analysis*

Exploratory Data Analysis was performed to examine the distribution and relationships among academic, behavioral, employability and placement attributes. Correlation analysis was used to study relationships between numerical features, while scatter plots were used to analyze the association between Attendance and Previous CGPA. The influence of Study Hours, Aptitude Score, Coding Score and Soft Skills Score on placement outcomes was also examined using appropriate graphical representations.

Placement distribution and the distributions of Performance Category and Dropout Risk were considered to support interpretation of student groups and placement-related characteristics.

E.Power BI Dashboard

A Power BI dashboard was developed to provide interactive visualization of the analyzed student data. The dashboard incorporates placement distribution, academic performance, attendance, employability skills, Performance Category and Dropout Risk. Interactive filters enable analysis of specific student groups and departments.

The dashboard complements the machine learning models by providing an interactive analytical interface for examining trends and identifying students requiring academic or placement support.

Journal of Research and Innovation in Technology, Commerce and Management
Vol. 3 Issue 9, September 2026, pp. 39046-39053
ISSN: 3049-3129 (Online)

Fig. 1. Power BI dashboard for student performance and placement analysis.

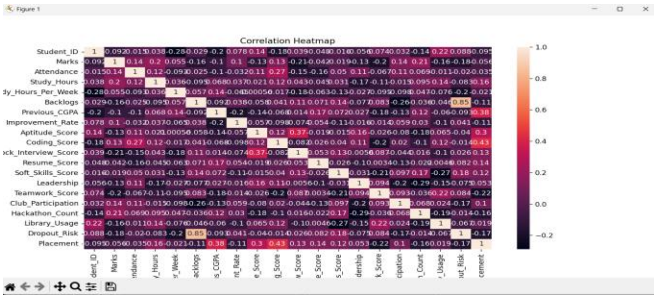

Fig. 2. Correlation heatmap of academic and employability attributes.

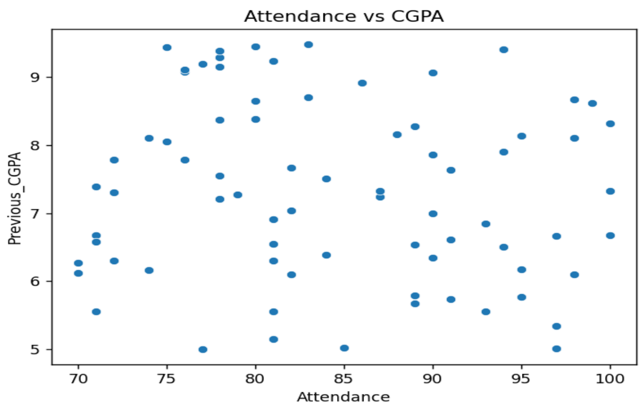

Fig. 3. Relationship between attendance and previous CGPA.

III. PROPOSED MACHINE LEARNING FRAMEWORK

The proposed framework integrates supervised and unsupervised learning techniques for student performance analysis, placement classification, and student segmentation. The workflow consists of data preprocessing, feature selection, model training, prediction, evaluation and visualization. Linear Regression is applied for academic performance estimation, Logistic Regression for placement prediction, Decision Tree for performance classification and K Means for student segmentation.

A. Linear Regression for Academic Performance Prediction

Linear Regression is applied to estimate Previous CGPA using Attendance, Marks and Study Hours Per Week. The model is trained using 80 percent of the dataset and evaluated on the remaining 20 percent using Mean Squared Error (MSE) and R squared

$$z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$$

$$\text{Previous CGPA} = \beta_0 + \beta_1(\text{Attendance}) + \beta_2(\text{Marks}) + \beta_3(\text{Study Hours})$$

(R^2). The model obtained an MSE of 1.5467 and an R^2 score of -0.4373, indicating limited predictive capability for CGPA estimation with the selected features.

TABLE II LINEAR REGRESSION RESULTS

Evaluation Metric	Result
Mean Squared Error	1.5467
R squared Score	-0.4373

B. Logistic Regression for Placement Classification

Logistic Regression is applied to classify the derived Placement status using Previous CGPA, Aptitude Score, Coding Score and Soft Skills Score. The Placement labels are generated using predefined criteria based on Previous CGPA, Aptitude Score and Coding Score. The model is evaluated on the test set using accuracy, precision, recall and weighted F1 score.

$$z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$$

$$P(Y = 1) = \frac{1}{1 + e^{-z}}$$

TABLE III LOGISTIC REGRESSION RESULTS

Evaluation Metric	**Result**
Accuracy	93.33 %

Weighted F1 Score	0.93
Precision — Placed	1.00
Recall — Placed	0.75
F1 Score — Placed	0.86

C. Decision Tree Classification

A Decision Tree classifier is applied to classify students according to Performance Category using Attendance, Marks, Study Hours Per Week, and Backlogs. The model achieved an accuracy of **40.00 percent** on the evaluated test set. The result indicates limited classification capability with the selected features and the available dataset. Additional predictive features and a larger dataset may improve the reliability of Performance Category classification.

TABLE IV DECISION TREE CLASSIFICATION RESULT

Evaluation Metric	Result
Accuracy	40.00 %

D. K Means Clustering for Student Segmentation

K Means clustering is applied for unsupervised student segmentation using Previous CGPA, Aptitude Score and Coding Score. Three clusters are generated to identify groups of students with similar academic and technical characteristics. The resulting clusters are interpreted using feature distributions and cluster centroids rather than assigning direct performance labels to the numerical cluster identifiers.

TABLE V SAMPLE K MEANS CLUSTER OUTPUT

Previous CGPA	Cluster
6.91	0
9.15	0
8.62	2
9.44	1
8.65	2

IV. RESULTS AND DISCUSSION

The experimental results demonstrate that model performance varies across prediction tasks and depends on the selected features and dataset characteristics. Logistic Regression achieved the strongest reported classification performance for the predefined placement labels, whereas Linear Regression and Decision Tree showed limited predictive capability with the selected features. K Means provided an additional segmentation-based view of student characteristics.

TABLE VI SUMMARY OF EXPERIMENTAL RESULTS

Model	**Task**	**Evaluation Metric**	**Result**
Linear Regression	CGPA Prediction	Mean Squared Error	1.5467
Linear Regression	CGPA Prediction	R squared Score	-0.4373
Logistic Regression	Placement Prediction	Accuracy	93.33%
Logistic Regression	Placement Classification	Weighted F1 Score	0.93
Decision Tree	Performance Classification	Accuracy	40.00 %
K Means	Student Segmentation	Number of Clusters	3

The Logistic Regression model achieved 93.33 percent accuracy and a weighted F1 score of 0.93 on the evaluated test set using Previous CGPA, Aptitude Score, Coding Score and Soft Skills Score. Since the placement labels were derived using predefined criteria, this result evaluates the classification of the defined labeling scheme rather than generalized real-world placement outcomes.

Linear Regression obtained an **MSE of 1.5467** and an **R^2 score of −0.4373**, indicating limited predictive capability of Attendance, Marks, and Study Hours Per Week for estimating Previous CGPA. The Decision Tree classifier achieved an **accuracy of 40.00 percent**, indicating limited classification capability using Attendance, Marks, Study Hours Per Week, and Backlogs. These results highlight the importance of additional predictive features and larger datasets.

K Means generated three clusters based on similarities in Previous CGPA, Aptitude Score, and Coding Score. The clustering results provide an additional perspective for identifying groups of students with similar academic and technical characteristics and complement the supervised learning results.

V. KEY INSIGHTS AND INSTITUTIONAL IMPLICATIONS

[1] The analysis indicates that academic and employability attributes can support classification of the predefined placement labels. The Logistic Regression result demonstrates promising classification performance within the evaluated dataset; however, the small test set and rule-based label construction limit the generalizability of the result.

[2] The framework can support institutions in identifying students requiring additional mentoring by considering attendance, backlogs, academic performance, dropout risk and study effort jointly. Placement departments can use aptitude, coding, soft skills, mock interview, and resume-related indicators to design targeted preparation programs.

[3] Student segmentation using K Means provides an additional perspective by identifying groups with similar academic and employability characteristics. The Power BI dashboard enables department-wise and category-wise examination of these patterns and supports data-driven academic and placement review.

[4] The proposed system is intended as a decision-support framework rather than an automated replacement for academic or placement professionals. Predictive outputs should be interpreted with institutional context and appropriate human review.

VI. LIMITATIONS AND FUTURE WORK

A. Limitations

1. The primary limitation of the study is the dataset size of 75 student records, which restricts the generalizability of the experimental results. A larger dataset covering multiple academic years and departments would provide more reliable model training and evaluation.
2. Model performance is also dependent on feature selection and data quality. The Linear Regression and Decision Tree results indicate that the selected features were insufficient for certain prediction tasks. Additional academic, behavioral, technical and placement-related attributes may improve predictive performance.
3. The Logistic Regression accuracy of 93.33% was obtained from a test set containing only 15 observations. In addition, the Placement labels were derived using predefined Previous CGPA, Aptitude Score and Coding Score criteria. Therefore, the result should not be interpreted as a general guarantee of real-world placement prediction performance. Cross-validation and evaluation using larger independent datasets with externally validated placement outcomes are required for stronger validation.
4. Finally, the framework is based primarily on historical student data. Changes in curriculum, student behavior, recruitment practices and industry skill requirements may affect model performance over time.
5. The K Means clustering implementation uses Previous CGPA, Aptitude Score and Coding Score without feature standardization. Since these

variables have different numerical ranges, feature scaling may improve the fairness of distance-based clustering and should be investigated in future experiments.

B. Future Work

1. Future development will focus on integrating larger and continuously updated institutional datasets covering multiple academic years and departments. Advanced algorithms such as Random Forest, Gradient Boosting and XGBoost can be evaluated with hyperparameter optimization and cross-validation to improve model reliability.
2. Feature importance and explainable machine learning techniques can be incorporated to identify influential factors and provide interpretable predictions. Ensemble learning can also be investigated to combine multiple models for improved predictive performance.
3. The framework can further be extended into a web-based decision-support application for authorized users. Integration with institutional systems could enable automated incorporation of attendance, assessment and placement data. Personalized recommendations for aptitude practice, coding development, communication training, mentoring, and placement preparation can also be incorporated based on identified student needs.

VII. CONCLUSION

1. This paper presented a College Student Performance Analytics and Placement Prediction Framework integrating data analytics, machine learning, student segmentation and Power BI visualization. The framework analyzes academic, behavioral, technical, extracurricular and employability attributes to support student performance assessment and placement-related decision making.
2. Linear Regression was applied for CGPA estimation, Logistic Regression for placement prediction, Decision Tree for performance classification, and K Means for student segmentation. Logistic Regression achieved 93.33% accuracy with a weighted F1 score of 0.93, demonstrating promising classification performance for the predefined placement labels using Previous CGPA, Aptitude Score, Coding Score and Soft Skills Score. Because the placement labels were derived using predefined criteria, the result should not be interpreted as evidence of generalized real-world placement prediction capability.Linear Regression obtained an MSE of 1.5467 and an R^2 of −0.4373, while Decision Tree achieved **40.00% accuracy**, indicating limitations associated with the selected features and dataset size.
3. The Power BI dashboard provides an interactive visualization layer for analyzing placement outcomes, academic performance, employability skills, performance categories and dropout risk. The integration of predictive analytics and visualization provides a foundation for identifying students requiring support and designing targeted academic and placement interventions.
4. Although the current results are limited by dataset size and validation constraints, the proposed framework demonstrates the potential of educational data analytics for data-driven student support and institutional planning. Larger datasets, advanced machine learning models, explainable predictions and real-time institutional data integration can further enhance the effectiveness of the proposed framework.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the faculty members and the Department of Artificial Intelligence and Data Science, Pravara Rural Engineering College Loni, for their guidance and support during the development of this work. The authors also acknowledge the academic and technical resources that supported the implementation of the proposed student performance analytics and placement classification framework.

REFERENCES

[1] C. Romero and S. Ventura, “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, Nov.2010.

[2] R. S. J. d. Baker and K. Yacef, “The state of educational data mining in 2009: A review and future visions,” *Journal of Educational Data Mining*, vol. 1, no. 1, pp. 3–17, 2009.

[3] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York, NY, USA: Springer, 2009.

[4] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.

[5] F. Pedregosa *et al.*, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[6] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, pp. 5–32, 2001.

[7] J. Han, J. Pei, and H. Tong, *Data Mining: Concepts and Techniques*, 3rd ed. Waltham, MA, USA: Morgan Kaufmann, 2012.

[8] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*, 4th ed. Cambridge, MA, USA: Morgan Kaufmann, 2016.

[9] T. M. Mitchell, *Machine Learning*. New York, NY, USA: McGraw-Hill, 1997.

[10] Microsoft, “Microsoft Power BI documentation,” Microsoft Learn, 2026.