ISSN: 2096-3246
Volume 58, Issue 2, 2026

AN EVALUATION OF MACHINE LEARNING ALGORITHMS IN PREDICTING STUDENT PERFORMANCE

Shruti Mehta[1], Dr Mukta Agarwal[2]
1 Indus University, Ahmedabad, Gujarat, India
2 Indus University, Ahmedabad, Gujarat, India
1 shrutimehta.21.rs@indusuni.ac.in
2 muktaagarwal.dcs@indusuni.ac.in

Abstract

This study investigates machine learning algorithms to predict student performance using a publicly available Kaggle dataset containing academic, behavioral, and socio-demographic attributes. Four algorithms—Logistic Regression, Random Forest, Decision Tree, and Gradient Boosting—were evaluated using cross-validation for reliability and accuracy assessment.

The Gradient Boosting classifier emerged as the best-performing model, achieving an accuracy of 96%. Its interpretability and simplicity make it well-suited for educational data analysis. Random Forest and Decision Tree provided competitive results, while Logistic Regression demonstrated lower performance due to the dataset's non-linear patterns.

These results highlight the critical role of algorithm selection in student performance prediction and underscore the potential of machine learning in enhancing educational decision-making. Future work could explore advanced models and feature engineering to further improve prediction accuracy.

Keywords: Machine Learning, Logistic Regression, Random Forest, Gradient Boosting, Decision Tree, Pedagogical issues, Teaching/learning strategies

1. Introduction

Modern educational institutions navigate a highly competitive and complex environment. Analyzing performance, delivering high-quality education, devising strategies to evaluate student outcomes, and identifying future needs are key challenges faced by most educational organizations today. To address these issues, schools and universities implement intervention plans to support students in overcoming challenges during their studies. Predicting student performance at the entry level and throughout subsequent stages enables institutions to effectively design and refine these intervention strategies. These performance prediction initiatives benefit both management and educators by providing valuable insights to enhance academic outcomes and support student success.

The National Education Policy (NEP) emphasizes competency-based learning and the importance of allowing students to learn at their own pace, fostering a stress-free environment that supports holistic development. This approach not only reduces pressure but also enables children to grasp concepts more effectively, leading to better academic outcomes. Predictive modeling plays a pivotal role in this framework by helping to anticipate student performance and identify areas that need targeted intervention. By leveraging advanced prediction techniques, educators can design personalized learning paths tailored to each student's unique strengths and weaknesses. This personalized approach ensures that students receive the support they need to achieve their full potential, aligning educational strategies with individual learning needs and goals.

Quality education, as emphasized in the Sustainable Development Goals (SDGs), strives to ensure inclusivity and equity for all children, regardless of their background. It aims to provide equal learning opportunities to every child, whether they come from rural or urban areas, belong to any caste, or face challenging family circumstances, such as parental illiteracy or socio-economic disadvantages. With the integration of machine learning, we can further enhance this vision by personalizing educational experiences to meet the unique needs of each child. Machine learning can analyze vast amounts of data to identify gaps in learning, predict performance, and recommend tailored strategies for improvement. By addressing individual challenges and adapting educational approaches, machine learning can transform classrooms into inclusive spaces where every child thrives, thereby fostering equity and lifelong learning for sustainable development.

2. Literature Review:

An adaptive recommendation system was proposed to predict suitable education paths for students in their preparatory year at Al-Azhar University's Faculty of Engineering. The system optimized pre-processing and architectural parameters, including F-measure (for imbalanced datasets), numeric feature discretization, and machine learning algorithm selection (SVM, QDA, KNN, LR, RF). The model adaptively identified optimal features and algorithms for different departments, achieving the highest F-measures with QDA for Urban (0.91), RF for Mechanical (0.78) and Civil (0.79), KNN for Architectural (0.89) and Electrical (0.77), LR for Mining and Petroleum (0.91), and SVM for Computer (0.73). With an average F-measure of 82.57%, the system demonstrated effective adaptability and accuracy.

A comparative analysis of algorithms reveals that K-Nearest Neighbor (KNN) outperforms Decision Tree in accuracy for predicting student performance. GPA emerges as a critical factor in determining final outcomes, with most students categorized as "Pass," while 19% fall under "Excellent" and 31% under "Good." Early prediction can help improve student performance, such as upgrading students from "Good" to "Excellent" or preventing failures, which currently account for 6%. Additionally, gender has been identified as a significant attribute, with female students often performing better in certain categories due to effective learning and management skills. Another crucial factor is travelling time, which impacts students' energy and study time, as longer commutes may hinder academic focus. These findings highlight that GPA, gender, and travel time are pivotal attributes in determining prediction accuracy and influencing students' overall academic outcomes. Adjusting these factors can significantly enhance predictions and improve results.

A detailed analysis of data for stream selection at the higher secondary level demonstrates the effectiveness of machine learning methods such as Support Vector Machines (SVM), Artificial Neural Networks (ANN), and Neural Networks (NN). Among these, Neural Networks outperformed the other algorithms, achieving a classification accuracy of 86.72%, along with a sensitivity rate of 0.91, specificity rate of 0.82, and Matthews Correlation Coefficient (MCC) of 0.72. These results highlight the reliability and efficiency of Neural Networks in accurately predicting suitable academic streams for students. The findings demonstrate the potential of Neural Networks as a robust tool for stream classification, supporting their application in educational decision-making processes at the higher secondary level.

Another study explored the use of K-Nearest Neighbors (KNN) and Naive Bayes as primary classification algorithms to develop a student performance prediction model. The study compared the two algorithms using three evaluation metrics: accuracy, recall, and precision. Results revealed that Naive Bayes outperformed KNN, achieving an impressive accuracy of 93%. These findings highlight the effectiveness of Naive Bayes in student performance prediction tasks, demonstrating its potential for accurate and reliable classification in educational datasets. The study emphasizes the importance of comparing multiple metrics to determine the most suitable algorithm for predictive modeling in academic contexts.

To address the issue of poor academic performance in computer science subjects, a study compared four machine learning techniques—Artificial Neural Networks (ANN), Naive Bayes, Decision Tree, and Logistic Regression—focusing on their classification accuracy. Among these, ANN achieved the highest accuracy, demonstrating its potential in predictive tasks for academic performance improvement. Meanwhile, Decision Tree analysis provided valuable insights by identifying that not all attributes significantly contribute to the classification process, emphasizing the importance of feature selection. This comparative analysis highlights the effectiveness of ANN in solving performance-related challenges while showcasing the interpretability of Decision Tree models for optimizing prediction strategies. These findings are instrumental in enhancing academic outcomes, particularly in computer science education.

3. Material and Methodology

3.1 Data Collection:

The dataset was obtained from the Kaggle dataset repository, comprising 33 attributes and 395 rows. The average of G1 and G2 was computed, and the outcomes were categorized into three groups: Below average (<=12), Average (>=12 and <=16), and Above average (>=16 and <=20). This categorization was performed with the intention of tailoring pedagogical strategies to students' comprehension levels. Assignments can be assigned separately to enhance their learning experiences.

3.2 Data Conversion

The dataset has been transformed into a numeric format, facilitating the ease of calculating statistical values and correlations. The average of Grade 1 and Grade 2 has been computed, and the results have been categorized into Below Average, Average, and Above Average.

Table 1: Attribute Conversion

SNo	Attribute	Description	Conversion
1	Sex	Student'gender	M=0, F=1
2	Address	Urban or Rural	U=0, R=1
3	Famsize	Family size	GT3=0, LE3=2
4	Pstatus	Parent Living apart or together	A=0,T=1
5	Medu	Mother's education	none=0,primary education(4th grade)=1, 5th to 9th grade=2, secondary education=3,higher education=4
6	Fedu	Father's education	none=0,primary education(4th grade)=1, 5th to 9th grade=2, secondary education=3,higher education=4
7	Mjob	Mother Education	services=0, health=1,teacher=2,at_home=3,other=4
8	Fjob	Father Education	services=0, health=1,teacher=2,at_home=3,other=4
9	reason	Reason to choose the school	course=0,home=1,reputation=2,other=3
10	guardian	Mother, Father or Other	mother=0,father=1,other=3
11	traveltime	Travelling time to school	<15 min=1,15 to 30 min =2,30 min to 1 hour=3 and >1 hour=4
12	studytime	Time spends by the students on studies	<2 hours=1,2 to 5 hours=2, 5 to 10 hours=3,>10 hours=4
13	failure	No of past class failures	0,1,2,3,4

SNo	Attribute	Description	Conversion
14	schoolsup	Extra Educational support	0,1
15	famsup	Extra family support	0,1
16	paid	Extra paid classes	0,1
17	activities	Extra-Curricular activities	0,1
18	nursery	Attended pre school	0,1
19	higher	Attended higher education	0,1
20	internet	Access of internet at home	0,1
21	romantic	Romantic or Practical	0,1
22	famrel	Family Relation	1,2,3,4,5
23	freetime	Free time after school	1,2,3,4,5
24	goout	Going out with friends	1,2,3,4,5
25	Dalc	Workday alcohol consumption	1,2,3,4,5
26	Walc	Weekend alcohol consumption	1,2,3,4,5
27	Health	Current health status	1,2,3,4,5
28	Absence	No. Of school absence	0 to 93
29	Avg(G1,G2)	Average of Grade 1 and Grade 2	0 to 20
30	Res	Performance of student	Below Average=0, Average=1, Above average=2

Existing studies have utilized machine learning techniques to achieve classification accuracies ranging from 89% to 91%, often using a limited set of attributes. Our objective is to explore classification techniques that can surpass 91% accuracy by leveraging all available attributes from open-source datasets. This approach aims to maximize the predictive potential of the data and improve performance beyond existing benchmarks.

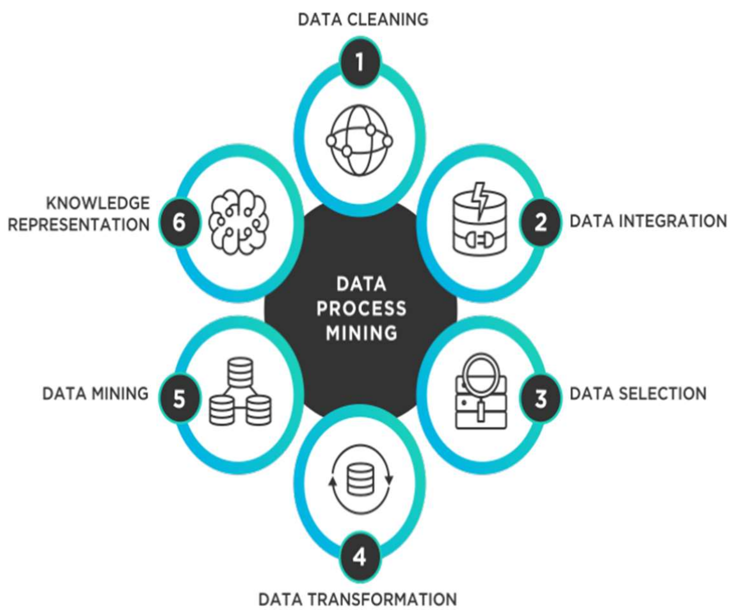

Fig 1: Cycle of Data Processing

3.3 Machine Learning Technique used

3.3.1 Logistic Regression

Logistic regression is a widely used classification technique for predicting a binary outcome. It is effective in analyzing the relationship between a binary dependent variable and one or more independent variables, which may be nominal, ordinal, interval, or ratio in nature. As a predictive classification method, logistic regression models the probability of an observation belonging to one of the two categories, making it a valuable tool for binary classification problems.

However, we have classified data into three categories such as "below average" (0), "average" (1), and "above average" (2), so we utilized extension of logistic regression called multinomial logistic regression.

The classification report of the Logistic Regression machine learning technique after applying on the student database taken from open-source dataset is as follows:

Table 2: The classification report of the Logistic Regression machine learning technique

	precision	recall	F1-score	support
0	0.95	0.96	0.95	75
1	0.88	0.85	0.86	33
2	0.91	0.91	0.91	11
accuracy			0.92	119
macro avg	0.91	0.91	0.91	119
weighted avg	0.92	0.92	0.92	119

The confusion matrix for the logistic regression machine learning model is as follows:

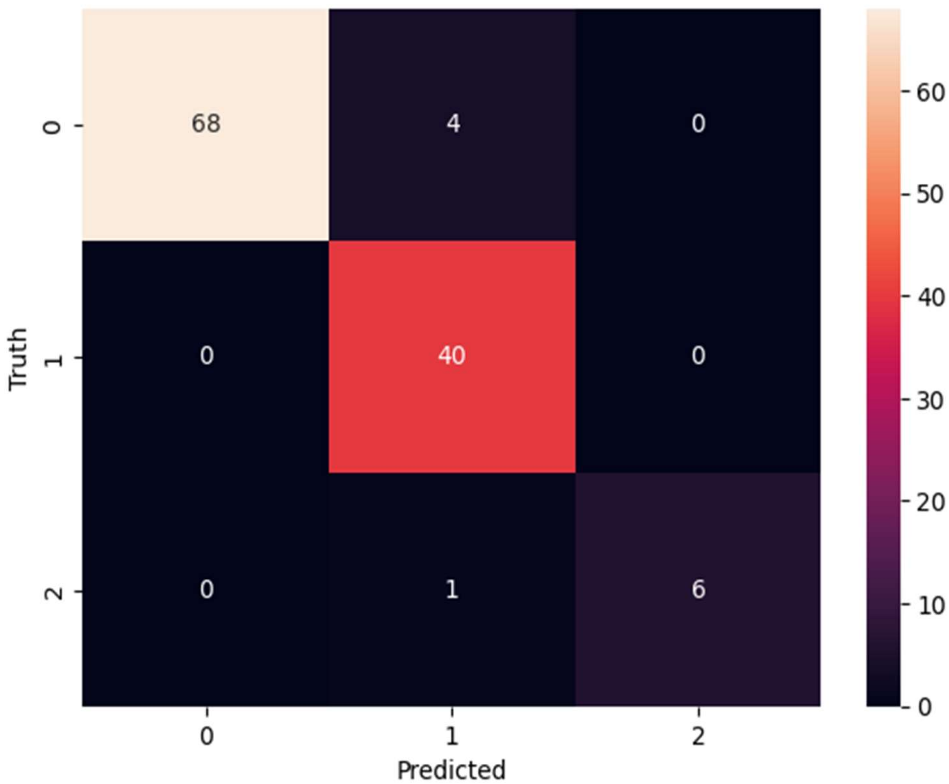

Figure 2: confusion matrix for the logistic regression

3.3.2 Decision Tree

This technique constructs a decision tree by calculating a gain ratio, a measure that assigns specific weights to the attributes within a dataset. By doing so, it identifies the most significant attributes that contribute to the classification process. The gain ratio evaluates the importance of each attribute based on how well it splits the data, resulting in a clear and interpretable tree of rules. These rules provide insights into the relationships among the attributes and help streamline the decision-making process for classification tasks.

Once the decision tree is built, the model can be reused to predict the desired class for new and unseen data. This reusability makes the model highly efficient for handling future predictions while maintaining accuracy and consistency.

The classification report of the decision tree machine learning technique after applying on the student database taken from open-source dataset is as follows:

Table 3: The classification report of the decision tree machine learning technique

	precision	recall	F1-score	support
0	0.97	0.97	0.97	75
1	0.91	0.88	0.89	33
2	0.83	0.91	0.87	11
accuracy			0.94	119
macro avg	0.90	0.92	0.91	119
weighted avg	0.94	0.94	0.94	119

The confusion matrix for the decision tree machine learning model is as follows:

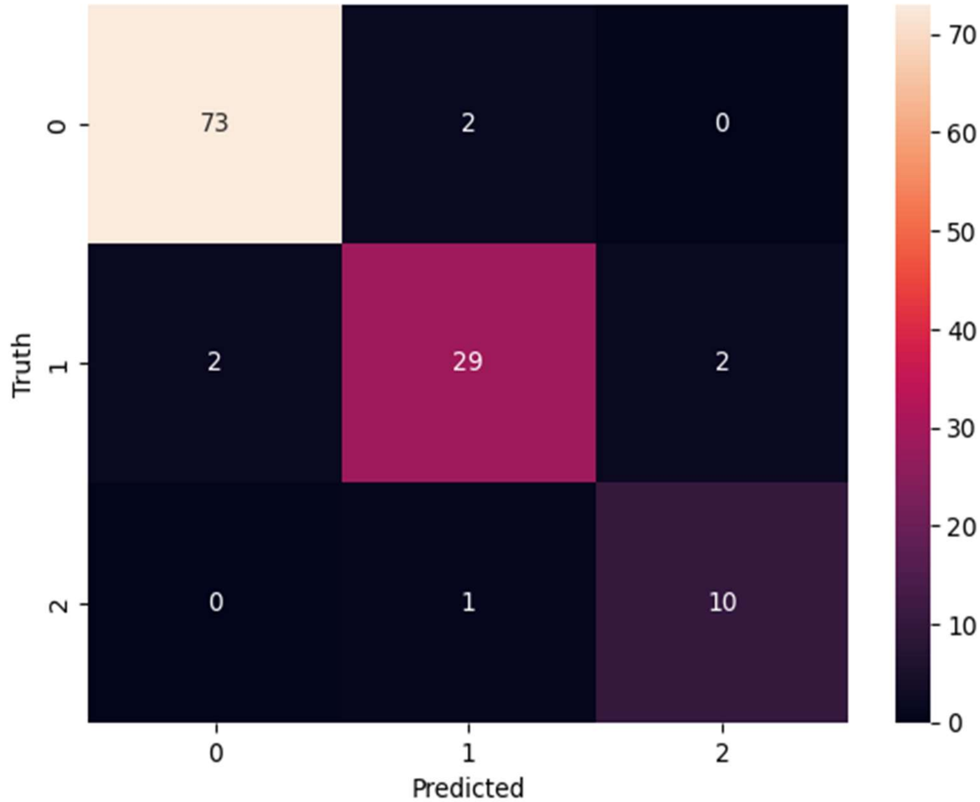

Figure 3: confusion matrix for the decision tree

3.3.3 Random Forest

Random Forest is a robust machine learning algorithm used for classification and regression. It builds multiple decision trees during training, and for classification tasks, predicts the output class based on the majority vote from all trees. By using random subsets of data and features for each tree, it ensures diversity, reduces overfitting, and improves model robustness.

A key advantage of Random Forest is its ability to handle high-dimensional data and provide insights into feature importance. It is resilient to noise and missing values, making it suitable for complex datasets. While it performs well in most scenarios, its computational complexity can be a limitation for smaller datasets or real-time predictions. The classification report of the random forest machine learning technique after applying on the student database taken from open-source dataset is as follows:

Table 4: The classification report of the random forest machine learning technique

	precision	**recall**	**F1-score**	**support**
0	0.95	0.97	0.96	75
1	0.75	0.91	0.82	33
2	1.00	0.18	0.31	11
accuracy			0.88	119
macro avg	0.90	0.69	0.70	119
weighted avg	0.90	0.88	0.86	119

The confusion matrix for the random forest machine learning model is as follows:

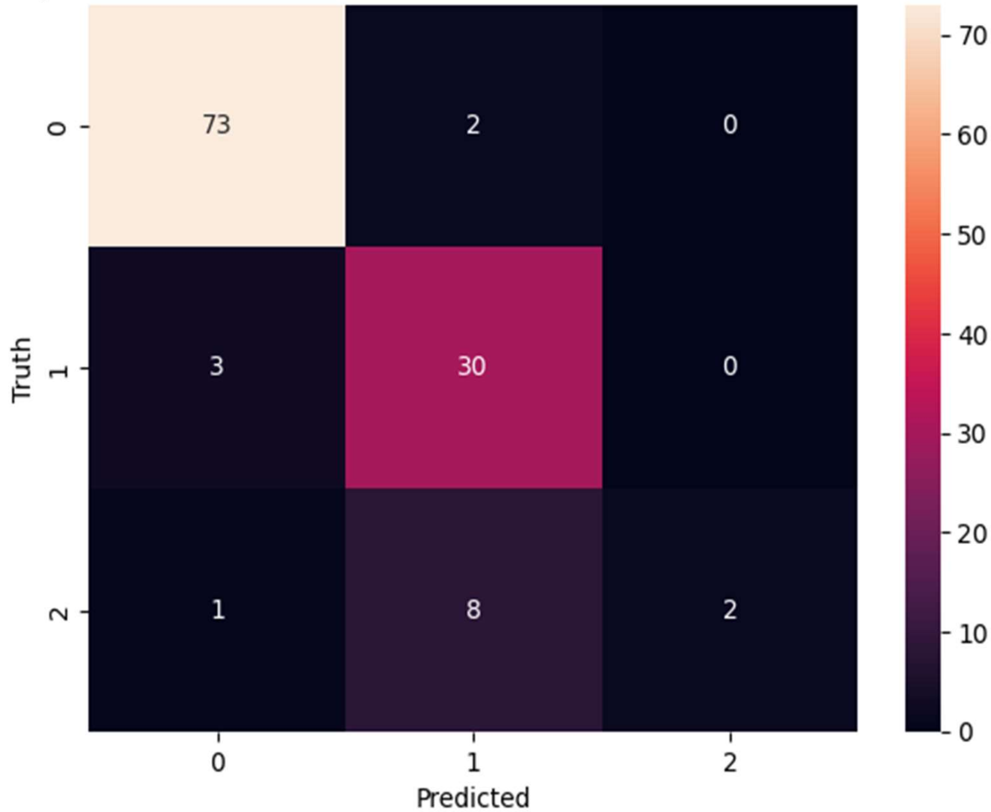

Figure 4: confusion matrix for the random forest

3.3.4 Gradient Boosting

Gradient Boosting Classifiers are a powerful group of machine learning algorithms designed to create strong predictive models by combining multiple weak learners, typically decision trees. The process involves building trees sequentially, where each subsequent tree corrects the errors of the previous ones. Gradient boosting optimizes a loss function by minimizing the residual errors in predictions, ensuring that the model becomes more accurate with each iteration. Unlike traditional ensemble methods that rely on simple voting or averaging, gradient boosting focuses on reducing errors step by step, making it particularly effective for complex datasets with non-linear relationships.

Decision trees serve as the base learners in gradient boosting because of their simplicity and flexibility. Each tree in the sequence is shallow, capturing only basic patterns, but when combined iteratively, they create a robust model capable of handling both regression and classification tasks. Gradient boosting is particularly effective for datasets with imbalanced classes or complex feature interactions. However, it requires careful parameter tuning, such as the learning rate and number of iterations, to prevent overfitting. Despite its computational demands, gradient boosting is widely used in various applications, from finance to healthcare, due to its high accuracy and interpretability.

The classification report is as follow:

Table 5: The classification report of the gradient boost machine learning technique

	precision	**recall**	**F1-score**	**support**
0	0.99	0.99	0.99	77
1	0.92	1.00	0.96	35
2	1.00	0.55	0.87	11
accuracy			0.97	119
macro avg	0.97	0.85	0.89	119
weighted avg	0.97	0.97	0.96	119

The confusion matrix of Gradient Boosting:

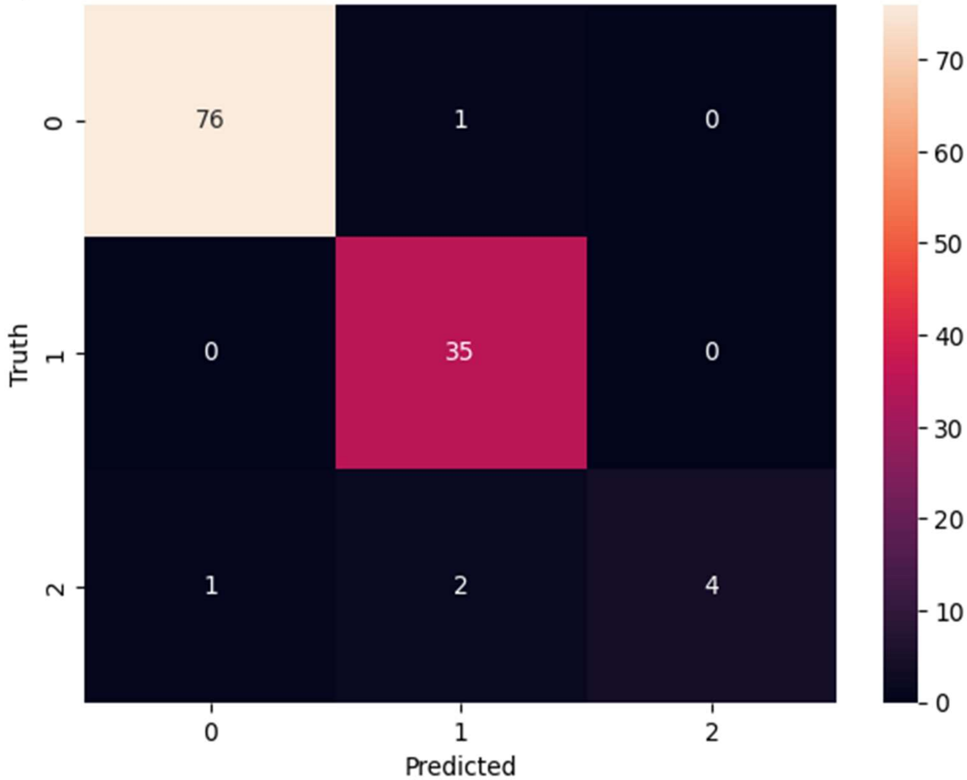

Figure 5: confusion matrix for the gradient boost

3.3.5 K-Fold Cross-Validation

It is a statistical technique used to evaluate the performance of machine learning models in a more reliable and robust manner. It works by splitting the dataset into K equal parts (or folds). The model is trained on K-1 folds and tested on the remaining fold, repeating this process K times such that each fold serves as a test set exactly once.

The results from all iterations are then averaged to provide a comprehensive estimate of the model's performance. This approach minimizes bias and variance in performance evaluation, as every data point gets used for both training and testing. K-Fold Cross-Validation is particularly useful when the dataset is small or when there is a risk of overfitting, as it ensures that the model is tested across diverse subsets of data.

The choice of K (commonly 5 or 10) balances computational efficiency and the reliability of the evaluation. By providing insights into how the model generalizes to unseen data, K-Fold Cross-Validation is an essential tool for model selection and hyperparameter tuning in machine learning.

This technique minimizes overfitting and variance in the evaluation process, particularly important given the moderate size of the dataset (395 rows). The average performance across all folds was then computed for each algorithm to obtain a more generalized accuracy estimate. The following are the average accuracies obtained through 10-fold cross-validation:

Table 6: 10-fold cross-validation

Algorithm	**Average Accuracy**
Logistic Regression	91.6%
Decision Tree	93.4%
Random Forest	91.8%
Gradient Boosting	96.1%

These results reflect not just the performance on a single test set but a more consistent measure of how well each model would perform on unseen data. Gradient Boosting consistently outperformed the other models, reinforcing its capability to handle complex, non-linear relationships in student data.

4. Result and Discussion

The performance of each machine learning model was evaluated using multiple metrics—accuracy, precision, recall, and F1-score—along with their respective confusion matrices. Gradient Boosting emerged as the most effective algorithm, achieving an accuracy of 96.1% with high precision across all three performance categories: below average, average, and above average.

Random Forest and Decision Tree models also performed well, achieving 91.8% and 93.4% accuracy, respectively. However, Random Forest showed limitations in classifying the "above average" category, likely due to class imbalance or overfitting to majority classes. Logistic Regression, although effective in simpler classification tasks, underperformed in this context with an accuracy of 91.6%, struggling to capture the dataset's non-linear patterns.

The high performance of Gradient Boosting can be attributed to its iterative learning process, which focuses on minimizing errors at each stage. Its ability to combine weak learners into a strong ensemble classifier makes it particularly suitable for education datasets where multiple features interact in complex ways.

These findings confirm that model selection significantly impacts prediction accuracy in educational data mining. Gradient Boosting, with its superior predictive ability and adaptability, offers promising implications for academic intervention strategies and personalized education plans.

5. Conclusion

This study demonstrates that the choice of machine learning algorithm plays a critical role in accurately predicting student performance. Among the models evaluated, Gradient Boosting proved to be the most effective, delivering the highest accuracy and consistent precision across all performance categories. Its robust performance highlights the strength of ensemble methods in handling complex, feature-rich educational datasets.

While Decision Tree and Random Forest also showed strong results, their limitations—particularly in classifying high-performing students—underscore the importance of addressing class imbalance and model tuning. Logistic Regression, though a reliable baseline model, fell short in capturing the non-linear relationships inherent in the data.

Overall, Gradient Boosting stands out as a valuable tool for educational data mining, offering actionable insights that can support early interventions and personalized learning strategies. Future work may explore hybrid models or deep learning techniques to further enhance predictive accuracy and uncover deeper patterns in student behavior and outcomes.

References

[1] B. Albreiki, N. Zaki, and H. Alashwal, "A Systematic Literature Review of Student' Performance Prediction Using Machine Learning Techniques," Education Sciences 2021, Vol. 11, Page 552, vol. 11, no. 9, p. 552, Sep. 2021, doi: 10.3390/EDUCSCI11090552.

[2] M. Ezz and A. Elshenawy, "Adaptive recommendation system using machine learning algorithms for predicting student's best academic program," Educ Inf Technol (Dordr), vol. 25, no. 4, pp. 2733–2746, Jul. 2020, doi: 10.1007/S10639-019-10049-7/METRICS.

[3] "PREDICTION OF STUDENTS PERFORMANCE USING KNN AND DECISION TREE- A MACHINE LEARNING APPROACH," Strad Research, vol. 7, no. 9, Sep. 2020, doi: 10.37896/SR7.9/018.

[4] K. Sethi, V. Jaiswal, and M. D. Ansari, "Machine Learning Based Support System for Students to Select Stream (Subject)," Recent Advances in Computer Science and Communications, vol. 13, no. 3, pp. 336–344, Nov. 2018, doi: 10.2174/2213275912666181128120527.

[5] I. A. A. Amra and A. Y. A. Maghari, "Students performance prediction using KNN and Naïve Bayesian," ICIT 2017 - 8th International Conference on Information Technology, Proceedings, pp. 909–913, Oct. 2017, doi: 10.1109/ICITECH.2017.8079967.

[6] M. M. Elsaid Khoudier et al., "Prediction of student performance using machine learning techniques," 5th Novel Intelligent and Leading Emerging Sciences Conference, NILES 2023 - Proceedings, pp. 333–338, 2023, doi: 10.1109/NILES59815.2023.10296766.

[7] H. Srava, N. Reddy, S. Tk, Y. Reddy, R. Ch, and R. Chinthala, "Prediction of Student Performance Using Logistic Regression," vol. 5, 2018, Accessed: Mar. 06, 2025. [Online]. Available: www.jetir.org

[8] Y. S. Alsalman, N. Khamees Abu Halemah, E. S. Alnagi, and W. Salameh, "Using Decision Tree and Artificial Neural Network to Predict Students Academic Performance," 2019 10th International Conference on Information and Communication Systems, ICICS 2019, pp. 104–109, Jun. 2019, doi: 10.1109/IACS.2019.8809106.

[9] L. H. Alamri, R. S. Almuslim, M. S. Alotibi, D. K. Alkadi, I. Ullah Khan, and N. Aslam, "Predicting Student Academic Performance using Support Vector Machine and Random Forest," ACM International Conference Proceeding Series, vol. PartF168981, pp. 100–107, Dec. 2020, doi: 10.1145/3446590.3446607.

[10] S. O. Oppong, "Predicting Students' Performance Using Machine Learning Algorithms: A Review," Asian Journal of Research in Computer Science, vol. 16, no. 3, pp. 128–148, Jul. 2023, doi: 10.9734/AJRCOS/2023/V16I3351.

[11] T. T. Wong, "Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation," Pattern Recognit, vol. 48, no. 9, pp. 2839–2846, Sep. 2015, doi: 10.1016/J.PATCOG.2015.03.009.

[12] V. Teodorescu and L. Obreja Brașoveanu, "Assessing the Validity of k-Fold Cross-Validation for Model Selection: Evidence from Bankruptcy Prediction Using Random Forest and XGBoost," Computation 2025, Vol. 13, Page 127, vol. 13, no. 5, p. 127, May 2025, doi: 10.3390/COMPUTATION13050127.

[13] T. J. Bradshaw, Z. Huemann, J. Hu, and A. Rahmim, "A Guide to Cross-Validation for Artificial Intelligence in Medical Imaging," Radiol Artif Intell, vol. 5, no. 4, Jul. 2023, doi: 10.1148/RYAI.220232,.

[14] I. Saini, D. Singh, and A. Khosla, "QRS detection using K-Nearest Neighbor algorithm (KNN) and evaluation on standard ECG databases," J Adv Res, vol. 4, no. 4, pp. 331–344, Jul. 2013, doi: 10.1016/J.JARE.2012.05.007.

[15] E. Kee, J. J. Chong, Z. J. Choong, and M. Lau, "A Comparative Analysis of Cross-Validation Techniques for a Smart and Lean Pick-and-Place Solution with Deep Learning," Electronics 2023, Vol. 12, Page 2371, vol. 12, no. 11, p. 2371, May 2023, doi: 10.3390/ELECTRONICS12112371.