https://doi.org/10.53656/math2025-6-4-obc

Education Technologies
Образователни технологии

OPTIMIZATION VS BOOSTING: COMPARISON OF STRATEGIES ON EDUCATIONAL DATASETS TO EXPLORE LOW-PERFORMING AT-RISK AND DROPOUT STUDENTS

Ranjit Paul[1)], Asmaa Mohamed[2)], Peren Canatalay[3)], Ashima Kukkar[4)], Sadiq Hussain[1)], Arun Baruah[1)], Jiten Hazarika[1)], Silvia Gaftandzhieva[5)], Esraa Mahareek[2)], Abeer Desuky [2)], Rositsa Doneva[5)]

[1)]Dibrugarh University, Dibrugarh (India)
[2)]Al-Azhar University (girls branch), Cairo (Egypt)
[3)]Istinye University, Istanbul (Turkey)
[4)]Chitkara University, Punjab (India)
[5)]University of Plovdiv "Paisii Hilendarski", Plovdiv (Bulgaria)

Abstract. The paper proposes a comprehensive student academic performance prediction approach by integrating machine learning with metaheuristic optimization. Initial models (Logistic Regression, Decision Tree, Random Forest, MLP) were refined using boosting techniques (Gradient Boosting, XGBoost, LightGBM), with XGBoost achieving 95.59% accuracy. Eight modern optimization algorithms were applied for feature selection to enhance model efficiency and interpretability, with the Grey Wolf Optimizer and the Heap-Based Optimizer outperforming others in key metrics. Support Vector Machine algorithms applied after feature selection strengthened the predictive capability of the selected feature subsets. The research outcomes demonstrate that uniting boosting approaches with feature selection algorithms enables the creation of reliable and scalable predictive models that detect student success and failure earlier.

Keywords: Machine Learning; Optimization Algorithms; Educational Data Mining; Ensemble Models; Boosting Algorithms.

1. Introduction

Student dropout rates significantly challenge higher education's role in fostering employment, social equity, and economic growth. Inconsistent

definitions and varied calculation methods (Xu & Kim, 2024) lead to reporting discrepancies, complicating efforts to implement effective student retention strategies. Higher education institutions (HEIs) use several monitoring techniques to assess student performance by tracking course advancement and analyzing academic standing each semester (Chen et al., 2014).

Technological advancements and increased data availability have established Educational Data Mining (EDM) as a specialized research field (Apriyadi & Rini, 2023). EDM uses data mining techniques to find actionable patterns in educational data. Its predictive models analyze student performance to help HEIs address dropout risks. However, standard predictive techniques still face challenges related to interpretability, scalability, and computational efficiency (Shekhar et al., 2020).

Data preprocessing, specifically feature selection, is crucial for optimizing data mining systems by removing redundant and noisy data. This process improves algorithm performance and enables classifiers to achieve higher accuracy. The two main types of feature selection are: filter methods, which are computationally efficient but cannot detect feature dependencies, and wrapper methods, like Linear Discriminant Analysis (LDA) and K-Nearest Neighbor (KNN), which are more effective at identifying complex dependencies but are computationally intensive and thus limited to smaller datasets. Identifying the optimal feature subset remains a challenge, as efficient search mechanisms (complete, random, or heuristic) risk overlooking optimal solutions (Hussain et al., 2020; Farissi et al., 2022; Punitha & Devaki, 2024; Ajibade et al., 2019). Heuristic search mechanisms offer an effective and efficient framework for problem-solving. Specifically, metaheuristic algorithms like Particle Swarm Optimization (PSO) and Ant Colony Optimization (ACO) demonstrate exceptional capability in feature selection. By replicating natural processes and employing probabilistic rules, these algorithms efficiently navigate large parameter spaces and escape local optima, making them well-suited for complex, high-dimensional datasets. This enables improved feature selection quality and more effective predictive models, particularly in applications like student performance prediction (Kukkar et al., 2023; Kukkar et al., 2024).

Using data to boost student retention is a key goal for university administrators. However, the sheer volume of student data can be overwhelming, requiring advanced tools to identify and help at-risk students proactively.

This study presents an integrated solution combining Machine Learning (ML) and metaheuristic methods to predict student academic performance. It equips teachers, administrators, and policymakers with a predictive tool for tracking at-risk students and implementing effective interventions. Such a tool transforms large-scale student data into actionable intelligence, enabling the strategic allocation of support services and the timely implementation of targeted interventions designed to improve student outcomes and reduce attrition. The proposed EDM approach addresses current deficiencies, providing a comprehensive framework for fostering academic success and reducing dropout rates in HEIs. To achieve this, the research focused on developing an accurate predictive analytics model using historical academic and demographic data; applying advanced data preprocessing techniques for improved data quality; implementing a range of advanced ML algorithms (Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), XGBoost, LightGBM) and metaheuristic techniques (Mud Ring Algorithm (MRA), Archimedes Optimization (AO), Jellyfish Search (JS), Ant Lion Optimizer (ALO), Grey Wolf Optimizer (GWO), Whale Optimization Algorithm (WOA), Heap-Based Optimizer (HO), Equilibrium Optimizer (EO)) to enhance prediction accuracy, conducting feature selection and importance analysis using RF; identifying 12th-grade percentage, CGPA, and gender as key predictors; performing a comparative analysis of boosting versus optimization techniques for feature selection to improve classifier efficiency and predictive accuracy; utilizing data visualization (histograms, heatmaps) to analyze patterns and relationships; statistically validating findings through cross-validation and comparisons with state-of-the-art methods; ensuring the model is computationally efficient and scalable for diverse educational datasets; providing data-driven insights for targeted interventions to improve learning outcomes.

The paper's structure includes Related Work (Section 2), Proposed Methodology (Section 3), Experimental Results (Section 4), Discussion (Section 5), and Conclusion (Section 6).

2. Related work

Ma (2024) enhanced student performance prediction by optimizing an RF Classifier with Electric Charged Particles Optimization (ECPO) and Artificial Rabbits Optimization. Analyzing 4,424 student records, their optimized model demonstrated higher predictive precision and better alignment with actual values, proving bio-inspired algorithms effective for educational decision-making.

Thaher et al. (2021) developed a Student Performance Predictive model using an enhanced WOA (EWOA) for automatic feature selection. Their approach integrated the Sine Cosine Algorithm, a Logistic Chaotic Map, and an Adaptive Synthetic Sampling to address data imbalances. This method, particularly with LDA, showed superior reliability and enhanced prediction accuracy compared to other classifiers and feature selection methods on real educational datasets.

Hasheminejad & Sarvmili (2019) introduced S3PSO, a discrete PSO method for forecasting student outcomes via rule-based prediction. Using Support, Confidence, and Comprehensibility metrics, S3PSO generated understandable rules from the Moodle dataset, achieving a 31% fitness improvement over standard methods like CART, C4.5, and ID3. It also outperformed benchmark algorithms (Support Vector Machine (SVM), KNN, Naive Bayes (NB), Neural Networks (NN), APSO) by 9% in student performance forecasting accuracy.

Turabieh et al. (2021) developed HHO-based dynamic controllers with KNN clustering to overcome early stagnation and local minima in student performance feature selection. Their HHO-enhanced model, particularly with Layered Recurrent NN and Artificial NN (ANN), achieved the highest accuracy on UCI data for early prediction of student outcomes.

Song (2024) integrated KNN with Honey Badger Optimization (HBO) and the Arithmetic Optimization Algorithm (AOA) to create the KNHB prediction system. This model excelled in both prediction and classification tasks for G1 and G3 datasets, demonstrating high accuracy (0.921) and

precision (0.92) for G3. The KNHB model also demonstrated exceptional precision as a G1 value forecaster, with accuracy and precision scores of 0.899 and 0.90, respectively, in the prediction phase.

Ren & He (2024) enhanced a NB model for student performance prediction using Leader Harris Hawk's Optimization and Alibaba and the Forty Thieves Algorithm. Their model achieved 0.891 accuracy and substantial precision, recall, and F1-scores, outperforming other tested models and improving student support. The improvement in prediction precision achieved through these methods helps educational institutions deliver better student support and improve academic results.

Hai & Wang (2024) improved Multilayer Perceptron Classification (MLPC) for student performance prediction by combining the Pelican Optimization Algorithm and the Crystal Structure Algorithm. Their MLPO2 approach, using appropriate fine-tuning and preprocessing, achieved a 95.78% success rate and effectively handled class imbalance and high dimensionality.

Li & He (2024) applied ML dimensionality reduction and optimized an Extra-Trees Classifier with Gorilla Troops Optimizer and Reptile Search Algorithm for student success prediction. Their ETRS model achieved 97.5% accuracy in G1 mathematics course prediction, demonstrating the promise of bio-inspired optimization for educational outcomes.

Goran et al. (2024) used metaheuristic optimization with a modified Sinh Cosh Optimizer to enhance Adaptive Boosting (AdaBoost) and XGBoost for student dropout risk prediction. Their approach demonstrated superior performance on real-world binary and multi-class datasets, with SHAP and SAGE explainability methods identifying key dropout triggers for targeted retention programs.

Cheng et al. (2024) evaluated various ML techniques (RF, DT, KNN, MLP, XGBoost) and ANNs for student performance prediction. Their SVM–SMOTE data- balancing process significantly improved results, with an Enhanced Artificial Ecosystem-Based Optimization + XGBoost hybrid model achieving 0.9417 accuracy and a 0.9413 F1-score, confirming the success of combining ML with metaheuristics for precise student performance prediction.

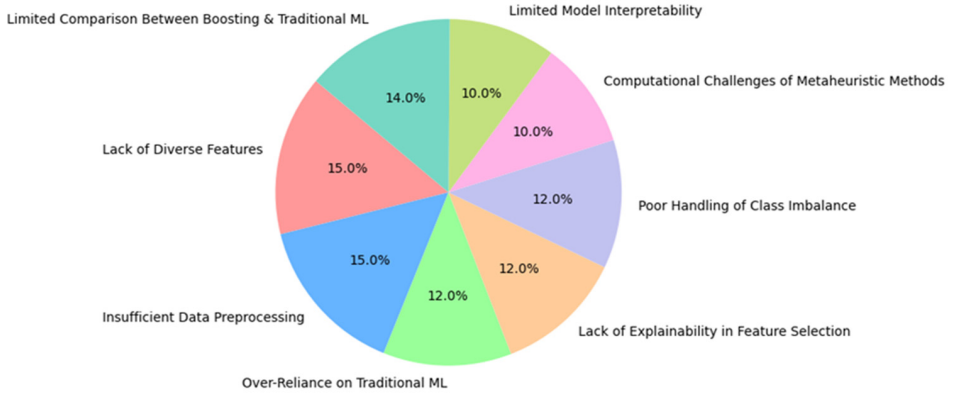

Figure 1. Percentage Distribution of the Identified Research Gaps

Previous studies often neglected student demographic and socioeconomic factors, used limited datasets, and inadequately addressed missing values, outliers, and feature engineering (see Fig. 1). In contrast, the current study utilizes a diverse dataset encompassing enrollment, 10th/12th-grade scores, demographics (gender, caste), and specialized program information. To ensure strong model foundations, we establish a robust data preprocessing pipeline, including categorical encoding, missing-value imputation, and IQR-based outlier detection. While prior research frequently employed NB, DT, RF, and some bio-inspired algorithms for optimization, our investigation specifically examines the predictive excellence of AdaBoost and Gradient Boosting. Furthermore, unlike studies using metaheuristic algorithms (e.g., ECPO, WO, ACO) for feature selection without explaining feature importance, this work utilizes Random Forest for feature importance analysis, identifying 12th-grade percentage, CGPA, and gender as the most influential predictors. The issue of imbalanced class distribution—often inadequately addressed in related works, leading to biased outcomes—is resolved in our method through SVM–SMOTE, improving recall and F1-scores. Recognizing the computational intensity of some metaheuristic optimization techniques (e.g., Jellyfish Search, HBO) that limit their application to large datasets, our research prioritizes computational

efficiency and scalability through performance evaluation and training time measurements. Finally, while current models struggle to identify key performance factors and lack direct comparative analysis between popular ML and metaheuristic optimization techniques, our study performs an extensive comparative analysis. In essence, the research fills these gaps by integrating extensive datasets, advanced preprocessing, optimized ML models, metaheuristic-based feature selection and classification with class balancing, and efficiency checks. These enhancements yield predictive models with improved accuracy, interpretability, and scalability for student academic performance assessment.

3. Proposed methodology

This study develops a robust predictive model for student academic performance by integrating diverse historical academic and demographic data obtained from educational institutions. The process involves dataset description, data preprocessing, experimental setup, model development, performance evaluation, and comparative analysis.

3.1. Dataset description

The dataset, provided by multiple educational institutions, comprises 19 numerical and 17 categorical features. It includes academic performance metrics like 10th and 12th standard examination scores and CGPA, alongside demographic details (gender, caste), program-specific data (major and minor subjects), and institutional identifiers (enrollment number, college name). Table 1 provides a detailed description of the attributes.

Table 1. Feature descriptions

Feature	Description
ENROLLMENT	Unique enrollment number of each student.
Programme	The program the student is enrolled in, such as B.A.
College Name	The name of the college.
MAJOR	The major subject chosen by the student, such as Education.
MINOR	The minor subject chosen by the student, such as Sociology or Political Science.
GENDER	Gender of the student: MALE or FEMALE.

Feature	Description
AGE	Age of the student, expressed in years, months, and days.
CASTE	The caste of the student: UR (Unreserved), ST (Scheduled Tribes), SC (Scheduled Castes), OBC (Other Backward Class)
X PASSING YEAR	The year the student passed their 10th standard examination.
X PERCENTAGE	Percentage scored in the 10th standard examination.
XII PASSING YEAR	The year the student passed their 12th standard examination.
XII STREAM	The stream chosen by the student in the 12th standard examination.
XII MAXIMUM MARKS	The maximum possible marks in the 12th standard examination.
XII MARKS OBTAINED	Marks obtained by the student in the 12th standard examination.
XII PERCENTAGE	The percentage scored in the 12th standard examination.
XII SUB 1, MAX MARK 1, OBTAINED MARK 1	Subject, maximum marks, and obtained marks for the first subject in the 12th standard.
XII SUB 2, MAX MARK 2, OBTAINED MARK 2	Subject, maximum marks, and obtained marks for the second subject in the 12th standard.
XII SUB 3, MAX MARK 3, OBTAINED MARK 3	Subject, maximum marks, and obtained marks for the third subject in the 12th standard.
XII SUB 4, MAX MARK 4, OBTAINED MARK 4	Subject, maximum marks, and obtained marks for the fourth subject in the 12th standard.
XII SUB 5, MAX MARK 5,	Subject, maximum marks, and obtained marks for the fifth subject in the 12th standard.

Feature	Description
OBTAINED MARK 5	
XII SUB 6, MAX MARK 6, OBTAINED MARK 6	Subject, maximum marks, and obtained marks for the sixth subject in the 12th standard.
CGPA	Cumulative Grade Point Average of the student.
STATUS	Status of the student, which can be 1, 2, or 3. Status 1 means dropout 2 means at-risk students, and 3 means passed students.

3.2. Data Preprocessing

The accuracy of predictions depends heavily on maintaining data quality. The preprocessing steps include:

– *Handling Missing Values*: The imputation method was used to handle missing values while preserving the complete dataset structure.

– *Categorical Encoding*: Variables such as gender and caste were transformed using One-Hot Encoding to convert them into a numerical format suitable for machine learning algorithms.

– *Outlier Detection and Adjustment*: Outliers were identified using the Interquartile Range (IQR) method. Data points with zeros or unusually high values (e.g., in "XII MAXIMUM MARKS" or CGPA) were carefully reviewed and adjusted to mitigate the effects of data entry errors.

– *Feature Engineering*: New attributes (e.g., average subject marks) were generated to help the model capture a more holistic view of a student's academic performance beyond individual grades.

In addition to creating new attributes, the process also involved standardizing existing numerical features to ensure they are on a comparable scale for the machine learning algorithms and numerical features were standardized using techniques such as StandardScaler. Fig. 2 depicts a feature importance plot.

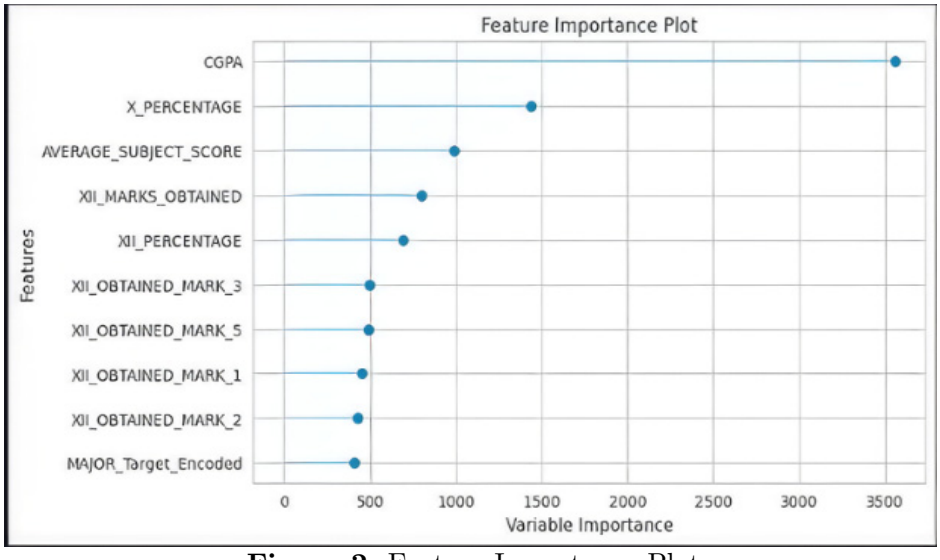

Figure 2. Feature Importance Plot

3.3. Data Visualisation and Reporting

To support interpretability and assess model performance, various visualization techniques were utilized.

A *Confusion Matrix* illustrates true versus false classifications for the best-performing models. In a multi-class problem like this one (with classes: Dropout, At-Risk, Passed), the matrix is an n × n table, where 'n' is the number of classes. Typically, each row represents the actual class, while each column represents the predicted class. The cells along the main diagonal show the number of correct predictions, where the predicted class matches the actual class. The cells off the diagonal show the errors or misclassifications.

To calculate performance metrics for a multi-class model, each class is typically evaluated in a "one-vs-all" manner. For any given class, we can define four basic outcomes: **True Positive (TP), True Negative (TN), False Positive (FP),** and **False Negative (FN)**. These four outcomes are used to calculate several key metrics that measure a model's performance from different perspectives: **Accuracy** (measures the proportion of all predictions that the model got right – calculated as the sum of all correct

predictions (the diagonal) divided by the total number of predictions), **Precision** (measures the accuracy of the positive predictions - Precision = TP / (TP + FP)), **Recall** (measures the model's ability to find all relevant instances of a class – Recall = TP / (TP + FN)), **F1-Score** (provides a single score that balances both precision and recall - F1-Score = 2 * (Precision * Recall) / (Precision + Recall)), **Geometric Mean** (useful for imbalanced datasets because it measures the balance between the classification performance on both the majority and minority classes – calculated as the root of the product of the sensitivity (recall) of each class), **Matthews Correlation Coefficient (MCC)** (considered a very reliable evaluation metric because it produces a high score only if the prediction did well in all four categories (TP, TN, FP, FN) – its value ranges from −1 to +1, where +1 indicates a perfect prediction, 0 represents a random prediction, and −1 indicates a total disagreement between prediction and observation), **Log Loss** (measures the difference between the predicted probabilities and the actual outcomes, penalizing the model more heavily for being confident in an incorrect prediction - for this metric, a lower value is better, with a perfect model having a log loss of 0).

ROC Curves evaluate the trade-offs between sensitivity and specificity.

Training and Validation Graphs allow monitoring model convergence over epochs or iterations.

Comparison Tables visually represent the impact of each model compared to other techniques.

4. Experimental setup and model development

To ensure a robust evaluation of the predictive models, the experimental process was divided into two main phases.

4.1. Experiment 1: Evaluation of baseline Machine Learning algorithms

We implemented LR, DT, RF, and Multi-Layer Perceptron as baseline models for comparative analysis and robust results. Ten-fold cross-validation ensured reliability, while RF-based feature importance analysis identified 12th-grade percentage, CGPA, and gender as key predictors of student performance. These predictors serve as key indicators that can help

institutions identify students who may be at risk of falling behind. Detailed results are presented in Figs. 3–6.

Accuracy: 60.43%
Classification Report:

	precision	recall	F1-score	support
1.0	0.59	0.01	0.02	1883
2.0	0.57	0.81	0.67	3501
3.0	0.65	0.73	0.69	3252
accuracy			0.60	8636
macro avg	0.61	0.52	0.46	8636
weighted avg	0.61	0.60	0.53	8636

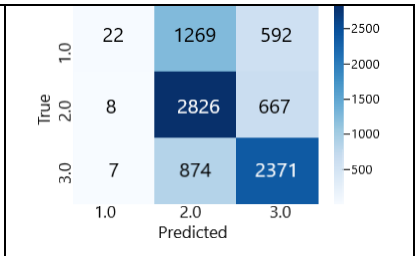

Figure 3. Evaluation metrics and confusion matrix related to logistic regression

The experimental process comprised three main phases, with Experiment 1 dedicated to evaluating baseline models. Ten-fold cross-validation was meticulously employed to ensure robust results and a reliable model evaluation framework.

Accuracy: 88.83%
Classification Report:

	precision	recall	F1-score	support
1.0	0.73	0.77	0.75	1883
2.0	0.94	0.93	0.93	3501
3.0	0.93	0.92	0.92	3252
Accuracy			0.89	8636
macro avg	0.87	0.87	0.87	8636
weighted avg	0.89	0.89	0.89	8636

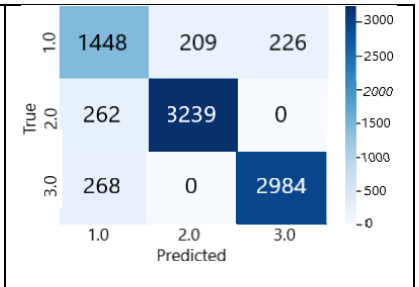

Figure 4. Evaluation metrics and confusion matrix related to the decision tree

One of the noteworthy outcomes of this study was the identification of influential predictors of student performance through feature importance analysis, particularly using the Random Forest (RF) algorithm. The feature importance analysis, using RF, revealed 12th-grade percentage, CGPA, and gender as key predictors of student performance, highlighting their significance for predictive models. We evaluated LR, DT, RF Classifier, and

Multi-Layer Perceptron: while LR and DT offered interpretability, RF and Multi-Layer Perceptron achieved higher predictive accuracy. This study underscores the importance of selecting optimal algorithms and feature combinations for accurate student performance evaluation in educational predictive modeling.

Accuracy: 91.15%
Classification Report:

	precision	recall	F1-score	support
1.0	0.88	0.68	0.77	1883
2.0	0.94	1.00	0.97	3501
3.0	0.89	0.95	0.92	3252
accuracy			0.91	8636
macro avg	0.91	0.88	0.89	8636
weighted avg	0.91	0.91	0.91	8636

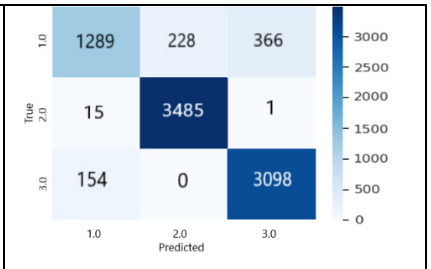

Figure 5. Evaluation metrics and confusion matrix related to Random Forest

Accuracy: 91.67%
Classification Report:

	precision	recall	F1-score	support
1.0	0.93	0.67	0.78	1883
2.0	0.94	1.00	0.97	3501
3.0	0.89	0.97	0.93	3252
accuracy			0.92	8636
macro avg	0.92	0.88	0.89	8636
weighted avg	0.92	0.92	0.91	8636

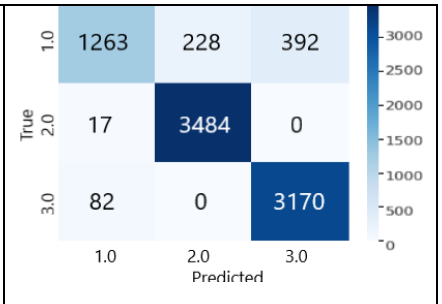

Figure 6. Evaluation metrics and confusion matrix related to MLP

4.2. Experiment 2: Evaluation of Boosting Algorithms

To establish optimal baseline performance, we assessed four boosting algorithms in the first phase: Gradient Boosting Classifier (GBC), XGBoost, and LightGBM. These models collectively demonstrated high performance with an average Accuracy of 92.86%, Precision of 93.31%, Recall of 92.99%, and F1 Score of 92.55%.

Accuracy: 92.86%
Classification Report:

	precision	recall	F1-score	support
1.0	0.96	0.70	0.81	1883
2.0	0.94	1.00	0.97	3501
3.0	0.90	0.99	0.94	3252
Accuracy			0.93	8636
macro avg	0.94	0.89	0.91	8636
weighted avg	0.93	0.93	0.92	8636

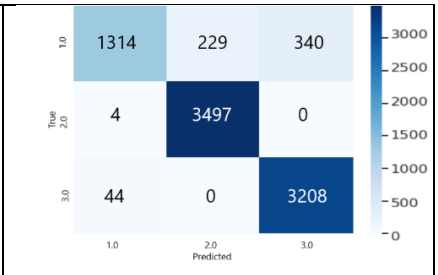

Figure 7. Evaluation metrics and confusion matrix related to GBC

Accuracy: 95.59%
Classification Report:

	precision	recall	F1-score	support
1.0	0.99	0.80	0.89	1883
2.0	0.94	1.00	0.97	3501
3.0	0.96	1.00	0.98	3252
accuracy			0.96	8636
macro avg	0.96	0.93	0.94	8636
weighted avg	0.96	0.96	0.95	8636

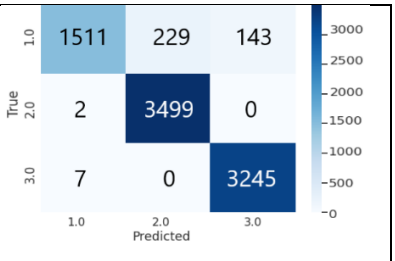

Figure 8. Evaluation metrics and confusion matrix related to XGBoost

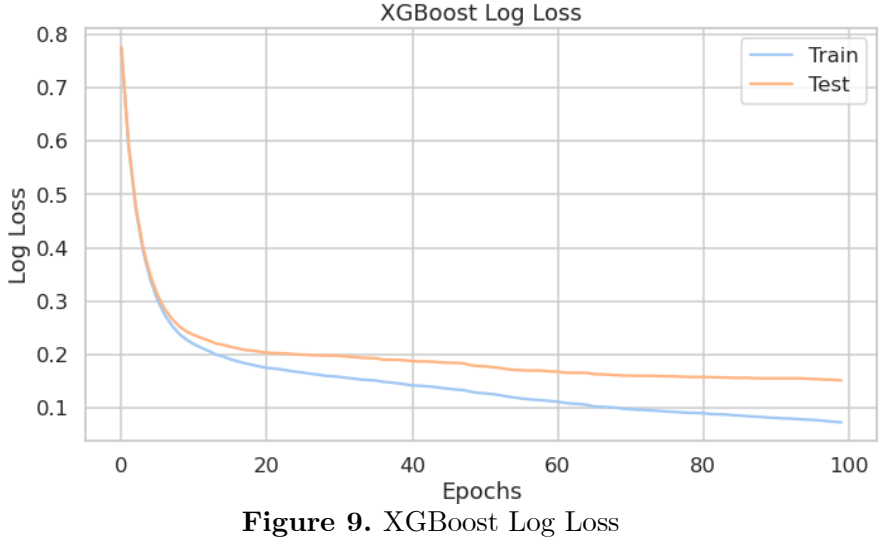

Figure 9. XGBoost Log Loss

Accuracy: 93.86%
Classification Report:

	precision	recall	F1-score	support
1.0	0.97	0.74	0.84	1883
2.0	0.94	1.00	0.97	3501
3.0	0.93	0.99	0.96	3252
accuracy			0.94	8636
macro avg	0.94	0.91	0.92	8636
weighted avg	0.94	0.94	0.94	8636

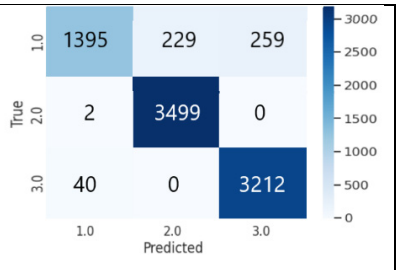

Figure 10. Evaluation metrics and confusion matrix related to LightGBM

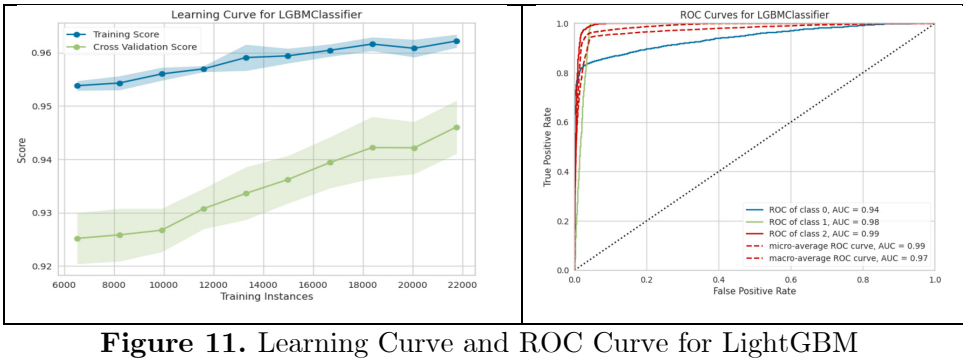

Figure 11. Learning Curve and ROC Curve for LightGBM

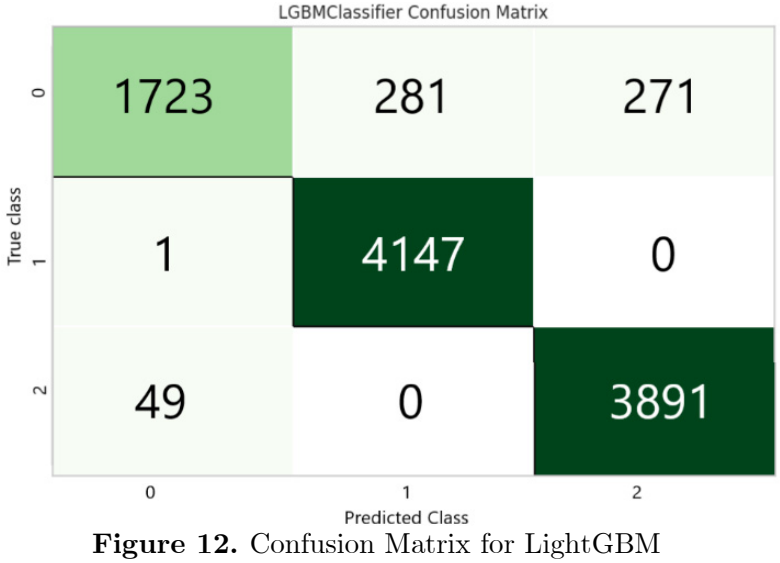

Figure 12. Confusion Matrix for LightGBM

Table 2. Comparison of outcome of different boosting algorithms based on their evaluation metrics

Model	**Accuracy%**	**Precision%**	**Recall%**	**F1 Score%**
Gradient Boosting	92.99	93.31	92.99	92.55
XGBoost Model	95.59	0.96	0.96	0.95
Tuned Lightgbm	94.19	94.45	94.20	93.90

The XGBoost model (see Table 2), with a 96% precision and recall, effectively identifies student statuses (Dropout, At-Risk, Passed). Its high precision means that when the model predicts a status, it's correct 96% of the time, leading to a low false positive rate. The 96% recall signifies it finds 96% of all students in each category, resulting in a low false negative rate. An F1-score of 95% indicates a strong, reliable balance between these two metrics, making the model a highly effective tool for identifying student statuses.

Experiment 2 analyzed the baseline performance of Gradient Boosting Classifier (GBC), XGBoost, and LightGBM in predictive modeling. Performance was assessed using accuracy, precision, recall, and F1-score (Figs. 7, 8, 10), with XGBoost log loss shown in Fig. 9, and LightGBM's learning curve, ROC curve, and confusion matrix in Figs. 11-12. Table 2 provides a detailed comparison. GBC demonstrated exceptional results, achieving 92.99% accuracy, 0.9255 F1-score, 93.31% precision, and 92.99% recall. Its strong Kappa (0.8898) and MCC (0.8940) values confirmed reliability and competence with imbalanced data, with a training time of 48.4370 seconds. XGBoost and tuned LightGBM also showed impressive performance, with F1 scores of 95.59% and 95% respectively, and the tuned LightGBM reaching 94.19% accuracy and 0.9390 F1-score. This demonstrates the effectiveness of GBC, XGBoost, and LightGBM in predictive modeling, each exhibiting distinct strengths across various metrics. Experiment 2's results underscore that optimal boosting algorithm selection depends on specific task requirements and performance objectives in classification tasks.

4.3. Experiment 3: Optimization-Based Feature Selection Followed by Classification

The second stage involved optimizing model performance by selecting the best possible features. Eight modern optimization techniques identified the most important features from training data for selection purposes. The methods used include MRA, AOA, JS, ALO, GWO, WOA, HBO and EO. The Feature selection process includes: *Binary Encoding* (each candidate solution is represented as a binary array where "1" indicates the feature is selected and "0" indicates exclusion), *Fitness Evaluation* (the f_measure of each candidate is computed to assess the quality of the selected feature subset) and *Data Splitting* (the dataset is segmented into training, validation, and testing sets). The training set produces candidate solutions while the validation set determines convergence, and the testing set provides final evaluation.

After optimal feature subset identification, SVM with an RBF kernel was used for classification, ensuring consistent performance comparison before and after feature selection. Among the optimization methods evaluated (metrics: accuracy, F-measure, geometric mean, sensitivity, specificity, precision), GWO and HBO showed superior performance, with GWO achieving 94.5% accuracy and strong F-measure and geometric mean scores.

5. Comparative analysis and discussion

Tables 3 and 4 present the performance of two distinct strategies for student performance prediction: Boosting-Based Models (Experiment 2) and Modern Optimization Methods (Experiment 3).

Boosting-Based Models (Experiment 2), particularly the GBC, excel in overall predictive accuracy and have been extensively validated using cross-validation. However, their computational efficiency varies, with AdaBoost offering a faster training time at a marginal cost to accuracy.

Experiment 3 (Modern optimization methods) focused on using optimization methods for feature selection, aiming to identify the most suitable feature subset from student datasets to enhance prediction capabilities. This process involved comparing the proposed method against contemporary optimization algorithms (MRA, AOA, JS, ALO, GWO, WOA, HBO, EO); implementing binary encoding for feature selection, where

"ones" denote included features and "zeros" represent excluded ones; evaluating individual fitness within optimization methods using the f_measure; segmenting the dataset into training, validation, and testing subsets, with optimization methods searching for the best feature subset on the training set and validating on the validation set until termination criteria are met; classifying the identified optimal subset using an SVM with an RBF kernel to ensure fair comparison of performance gains post-feature selection.

Table 3. Comparison of the outcome of different optimization algorithms

Method	Accuracy	F_Measure	Gmean	German_Features
Or	74.67	0.73	75.79	360
MRA	91.25	91.32	91.25	177
AOA	60.5	62.02	60.37	255
JS	62	65.77	61.02	275
ALO	93.5	93.33	93.47	181
GWO	94.5	94.27	94.42	159
WOA	89.25	89.49	89.22	266
HBO	94	93.78	93.93	186
EO	91.25	91.23	91.25	188

Table 4. Comparison of outcome of different optimization algorithms based on sensitivity, specificity and precision

Method	Sensitivity	Specificity	Precision
Or	76.92	74.67	0.36
MRA	92	90.5	90.64
AOA	64.5	56.5	59.72
JS	73	51	59.84
ALO	91	96	95.79
GWO	90.5	98.5	98.37
WOA	91.5	87	87.56
HBO	90.5	97.5	97.31
EO	91	91.5	91.46

The optimization methods were assessed based on accuracy, F-measure, geometric mean, sensitivity, specificity, and precision. The Grey Wolf Optimizer (GWO) and Heap-Based Optimizer (HBO) demonstrated superior

performance in identifying optimal feature subsets, significantly boosting model performance. GWO achieved 94.5% accuracy along with exceptional F-measure (94.27) and geometric mean (94.42) scores. HBO reached 94% accuracy with outstanding F-measure and geometric mean values. Both methods also effectively minimized false positives, with HBO showing 90.5% sensitivity and 97.5% specificity, and GWO demonstrating 98.5% specificity.

This research highlights that while boosting algorithms like Gradient Boosting and AdaBoost offer high accuracy and precision for overall prediction, optimization methods like GWO and HBO provide unique approaches to feature selection that significantly enhance model performance across different scenarios. The integration of a unified classifier (SVM with an RBF kernel) post-feature selection ensures a fair comparison of performance gains attributable to feature optimization. Ultimately, the choice of an appropriate method depends on the specific task requirements and performance objectives.

This study systematically evaluated machine learning and optimization strategies across three experiments to predict student achievement. Experiment 1 assessed baseline machine learning algorithms (LR, DT, RF, Multi-Layer Perceptron), revealing 12th-grade percentage, CGPA, and gender as key performance predictors through cross-validation and feature importance analysis. Experiment 2 utilized boosting algorithms (GBC, XGBoost, LightGBM). GBC achieved particularly impressive classification metrics, while XGBoost and tuned LightGBM also demonstrated excellent accuracy and F1 scores, underscoring the importance of task-specific boosting method selection. Experiment 3 applied modern optimization methods for feature selection, notably GWO and HBO, which proved highly effective in identifying crucial features. These methods delivered superior accuracy and F-measure when combined with an SVM (RBF kernel) classifier after feature selection, significantly enhancing model predictions. Overall, the experiments demonstrate the successful application of both machine learning algorithms and optimization techniques. While boosting algorithms deliver enhanced predictive accuracy, optimization-based feature selection provides additional gains by identifying the most influential features. Based on experimental findings, the most optimal strategy to

predict student performance is a combination of boosting algorithms for precise predictions and optimization-based feature selection methods. This integrated approach empowers researchers to build stronger, more accurate predictive models for student outcomes.

6. Implications for Administrative Practice and Institutional Policy

The XGBoost and GWO-enhanced predictive models are valuable strategic tools for higher education. They enable a shift from reactive to proactive, data-driven student support, improving retention and institutional effectiveness.

The models function as a ***data-driven triage system***, accurately identifying "at-risk" students and allowing for the proactive allocation of limited resources like advising and tutoring. This system circumvents the issue of students being unlikely to seek help themselves. By using objective data, the models also promote equity, flagging struggling students based on need rather than social or cultural factors.

Beyond individual student support, these models act as a ***diagnostic tool for the institution***. Aggregated “at-risk” data can reveal systemic issues, like “hot spots” in specific courses or programs. This empirical evidence allows administrators to make fact-based decisions about curricular reform, faculty development, and policy changes, fostering a dynamic feedback loop for continuous improvement.

Finally, the models serve as a ***catalyst for coordinated intervention***. An "at-risk" flag can trigger a multi-departmental workflow, breaking down institutional silos. This creates a holistic, wraparound support network where academic advisors, financial aid officers, and student services can work together to address a student's needs simultaneously, creating a more responsive and adequate infrastructure for student success.

7. Conclusion

Hybrid machine learning solutions are crucial for accurate student performance prediction. Our research, spanning multiple experimental phases, found that XGBoost and similar boosting algorithms deliver superior accuracy, further optimized by intelligent feature selection techniques like

GWO and HBO for effective data dimension reduction. The consistent use of SVM as a classifier facilitated fair performance comparisons. These findings offer practical insights for educational institutions to identify and support vulnerable students proactively. Future predictive model advancements will require integrating real-time student data and ensemble-based optimization to foster data-driven education.

Acknowledgements

This paper is financed by the European Union-NextGenerationEU, through the National Recovery and Resilience Plan of the Republic of Bulgaria, project № BG-RRP-2.004-0001-C01.

REFERENCES

Ajibade, S. S. M., Ahmad, N. B., & Shamsuddin, S. (2019). An heuristic feature selection algorithm to evaluate academic performance of students. In 2019 *IEEE 10th Control and System Graduate Research Colloquium*, 110 – 114.

Apriyadi, M., & Rini, D. (2023). Hyperparameter optimization of support vector regression algorithm using metaheuristic algorithm for student performance prediction. *International Journal of Advanced Computer Science and Applications*, *14*(2), 144 – 150.

Chen, J., Hsieh, H., & Do, Q. (2014). Predicting student academic performance: A comparison of two meta-heuristic algorithms inspired by cuckoo birds for training neural networks. *Algorithms*, *7*(4), 538 – 553.

Cheng, B., Liu, Y., & Jia, Y. (2024). Evaluation of students' performance during the academic period using the XG-Boost Classifier-Enhanced AEO hybrid model. *Expert Systems with Applications*, *238*, 122136.

Farissi, A., et al. (2022). High Accuracy Feature Selection Using Metaheuristic Algorithm for Classification of Student Academic Performance Prediction. *Int. Conf. of Advanced Computing and Informatics*, Cham: Springer International Publishing, 399 – 409.

Goran, R., et al. (2024). Identifying and understanding student dropouts using metaheuristic optimized classifiers and explainable artificial intelligence techniques. *IEEE Access, 12*, 122377 – 122400.

Hai, Q., & Wang, C. (2024). Optimizing Student Performance Prediction: A Data Mining Approach with MLPC Model and Metaheuristic Algorithm. *International Journal of Advanced Computer Science & Applications*, *15*(4), 55 – 71.

Hasheminejad, S., & Sarvmili, M. (2019). S3PSO: Students' performance prediction based on particle swarm optimization. *Journal of AI and Data Mining*, *7*(1), 77 – 96.

Hussain, K., Talpur, N., & Aftab, M. (2020). A Novel Metaheuristic Approach to Optimization of Neuro-Fuzzy System for Students' Performance Prediction. *Journal of Soft Computing and Data Mining*, *1*(1), 1 – 9.

Kukkar, A., Mohana, R., Sharma, A., & Nayyar, A. (2024). A novel methodology using RNN+ LSTM+ ML for predicting student's academic performance. *Education and Information Technologies*, 1 – 37.

Kukkar, A., Sharma, A., Singh, P. K., & Kumar, Y. (2023). Predicting Students Final Academic Performance Using Deep Learning Techniques. In *IoT, Big Data and AI for Improving Quality of Everyday Life: Present and Future Challenges: IOT, Data Science and Artificial Intelligence Technologies*, Cham: Springer International Publishing, 219 – 241.

Li, Y., & He, M. (2024). Elevating Student Performance Prediction using Extra-Trees Classifier and Meta-Heuristic Optimization Algorithms. *International Journal of Advanced Computer Science & Applications*, *15*(2), 434 – 450.

Ma, C. (2024). Improving the Prediction of Student Performance by Integrating a Random Forest Classifier with Meta-Heuristic Optimization Algorithms. *International Journal of Advanced Computer Science & Applications*, 15(6), 1032 – 1044.

Punitha, S., & Devaki, K. (2024). A high ranking-based ensemble network for student's performance prediction using improved meta-heuristic-aided feature selection and adaptive GAN for recommender system. *Kybernetes*. https://doi.org/10.1108/K-03-2024-0824

Ren, Z., & He, M. (2024). Meta-Model Classification Based on the Naïve Bias Technique Auto-Regulated via Novel Metaheuristic Methods to Define Optimal Attributes of Student Performance. *IJACSA*, *15*(1), 1049 – 1060.

Shekhar, S., Kartikey, K., & Arya, A. (2020). Integrating decision trees with metaheuristic search optimization algorithm for a student's performance prediction. *2020 IEEE SSCI*, 655 – 661.

Song, X. (2024). Student performance prediction employing k-Nearest Neighbor Classification model and meta-heuristic algorithms. *Multiscale and Multidisciplinary Modelling, Experiments and Design*, 1 – 16.

Thaher, T., et al. (2021). An enhanced evolutionary student performance prediction model using whale optimization algorithm boosted with sine-cosine mechanism. *Applied Sciences*, *11*(21), 10237.

Turabieh, H., et al. (2021). Enhanced Harris Hawks optimization as a feature selection for the prediction of student performance. *Computing*, *103*(7), 1417 – 1438.

Xu, H., & Kim, M. (2024). Combination prediction method of students' performance based on ant colony algorithm. *Plos one*, *19*(3), e0300010.

✉ **Mr. Ranjit Paul**
ORCID iD: 0009-0007-3046-7532
Centre for Computer Science and Applications,
Dibrugarh University, Dibrugarh 786004, India
E-mail: rs_ranjitpaul@dibru.ac.in

✉ **Dr. Asmaa Mohamed**
ORCID iD: 0000-0003-4314-463X
Department of Mathematics, Faculty of Sciences,
Al-Azhar University (girls branch), Egypt
E-mail: asmaamohamed89@azhar.edu.eg

✉ **Dr. Peren Jerfi Canatalay**
ORCID iD: 0000-0002-0702-2179
Department of Computer Engineering, Faculty of Engineering and Natural Sciences,
Istinye University, 34396
Istanbul, Turkey
Email: peren.canatalay@istinye.edu.tr

✉ Dr. Ashima Kukkar
ORCID iD: 0000-0001-7664-1005
Chitkara University Institute of Engineering and Technology,
Chitkara University, Punjab, India
E-mail: ashima@chitkara.edu.in

✉ Dr. Sadiq Hussain
ORCID iD: 0000-0002-9840-4796
Centre for Computer Science and Applications,
Dibrugarh University, Dibrugarh 786004, India
E-mail:sadiq@dibru.ac.in

✉ Prof. Arun K. Baruah
ORCID iD: 0009-0003-2115-9977
Department of Mathematics,
Dibrugarh University, Dibrugarh, Assam
E-mail: arun@dibru.ac.in

✉ Prof. Jiten Hazarika
ORCID iD: 0009-0001-3464-1071
Department of Statistics,
Dibrugarh University, Dibrugarh, Assam
E-mail: jiten_stats@dibru.ac.in

✉ Dr. Silvia Gaftandzhieva, Assoc. Prof.
ORCID iD: 0000-0002-0569-9776
University of Plovdiv Paisii Hilendarski, Plovdiv (Bulgaria)
E-mail: sissiy88@uni-plovdiv.bg

✉ Dr. Esraa A. Mahareek
ORCID iD: 0000-0002-9042-248X
Department of Mathematics, Faculty of Sciences,
Al-Azhar University (girls branch), Egypt
E-mail: esraa.mahareek@azhar.edu.eg

✉ Prof. Dr. Abeer S. Desuky
ORCID iD: 0000-0003-1661-9134
Department of Mathematics, Faculty of Sciences,
Al-Azhar University (girls branch)
E-mail: abeerdesuky@azhar.edu.eg

✉ Prof. Dr. Rositsa Doneva
ORCID iD: 0000-0003-0296-1297
University of Plovdiv Paisii Hilendarski, Plovdiv (Bulgaria)
E-mail: rosi@uni-plovdiv.bg