Research Paper
Journal of Smart Systems Research (JOINSSR) 5(1),40-54,2024
Received: 8-May-2024 Accepted: 4-June-2024
DOI: 10.58769/joinssr.1480695

Assessing Student Success: The Impact of Machine Learning and XAI-BBO Approach

Cem ÖZKURT[1,2*]

[1] Department of Computer Engineering, Faculty of Technology, Sakarya University of Applied Sciences
[2] Artificial Intelligence and Data Science Application and Research Center, Sakarya University of Applied Sciences

ABSTRACT

In today's dynamic educational landscape, understanding the multifaceted factors contributing to student success is paramount. This study delves into the intricate interplay of various determinants affecting student achievement. Through comprehensive preprocessing techniques, including nuanced transformations of categorical variables such as gender, age range, and parental education level, this research endeavors to unravel the complex fabric of educational outcomes. Leveraging the Biogeography-Based Optimization (BBO) algorithm, pivotal features crucial to student success are identified and integrated into sophisticated machine learning models. Evaluation metrics encompassing Accuracy, Precision, Recall, and F1 score illuminate the efficacy of these models, with the Gradient Boosting algorithm emerging as a standout performer, attaining a notable Accuracy value of 0.7388, Precision of 0.75, Recall of 0.72, and F1 score of 0.74. Further insights into model interpretability are gleaned through the application of SHAP and LIME methodologies, shedding light on the intricate mechanisms driving predictive outcomes. Specifically, the SHAP analysis highlights the influential factors driving model predictions, while LIME provides valuable insights into individual feature contributions. This study underscores the imperative of meticulous examination in delineating the determinants of student achievement, advocating for continued inquiry to inform evidence-based educational policies. The discernments furnished herein hold promise for fostering data-driven decision-making frameworks in education and facilitating targeted interventions aimed at fostering student success.

Keywords: Biogeography-Based Optimization (BBO) algorithm, Machine Learning models, Gradient Boosting algorithm, SHAP method, LIME method

1 Introduction

This study was conducted to analyze and explain the complex socio-economic factors that determine student success. In educational systems, there are many variables that affect student achievement, with socio-economic factors being of great importance among these variables [1]. Socio-economic factors such as family income level, parental education level, and living conditions can influence a student's academic performance.

In this research, various preprocessing steps were applied to the dataset to enable more effective use of categorical variables. Specifically, categorical variables in string type were converted to integer types to enhance the processing efficiency of the models. These significant adjustments included variables such as students' gender, age range, and parents' education level. The use of the Biogeography-Based Optimization (BBO) algorithm in the analysis of the dataset proved effective in identifying complex

[1] Corresponding Author’s email: cemozkurt@subu.edu.tr

relationships and allowed for the selection of attributes according to the complexity among features. In this study, using the BBO algorithm, the 20 most influential attributes affecting student success were identified from among the 32 attributes in the dataset.

These important attributes were evaluated by including them in machine learning models and using metrics such as Accuracy, Precision, Recall, and F1-score. The Gradient Boosting algorithm, which had the best Accuracy value, was explained using interpretable artificial intelligence models such as SHAP and LIME. The results obtained from this study emphasized the need for a detailed examination of the factors influencing student success and indicated the necessity for further research to enable more effective formulation and implementation of education policies. The findings of this study could contribute to the development of data-driven decision-making processes in the field of education and to more effectively planning interventions to improve student success.

2 Literature Review

One study examined the socioeconomic factors influencing the academic achievements of middle school students in Pakistan [2]. This study revealed that the education level of parents and socioeconomic status significantly impact students' performance in mathematics and English.

Another study conducted in Australia analyzed longitudinal data to explore factors affecting student achievement [3]. It demonstrated that prior achievements and early childhood cognitive abilities are crucial determinants of student success, with limited influence from socioeconomic status.

There is a study recommending the use of a type of Biogeography-based Optimization (BBO) in feature selection problems [4]. BBO relies on mathematical equations used to model the geographical distribution of biological organisms. Researchers propose that combining BBO with Support Vector Machine Recursive Feature Elimination (SVM-RFE) could create a more effective hybrid model for feature selection problems, leading to higher performance and better results in these problems.

Another study examines the impact of using artificial neural networks (ANNs) in education [5]. Researchers evaluate how effective multi-layer perceptron artificial neural networks (MLP-ANNs) are in predicting student performance. The study shows that multi-layer perceptron artificial neural networks developed considering students' past academic achievements and other factors have a high success rate in predicting student performance.

There is also a study that uses machine learning algorithms to predict the academic performance of students in a high school in Turkey [6]. Researchers evaluate various machine learning algorithms, including K-Nearest Neighbor, Decision Trees, Random Forest, Support Vector Machines, Multi-layer Perceptron, Logistic Regression, and Naive Bayes, using a dataset of 655 students. The results indicate that the Random Forest algorithm achieves the highest success rate. This study provides an effective machine learning model for predicting the academic performance of high school students in Turkey.

A model was developed using machine learning algorithms to predict the academic achievements of students in a high school in Turkey [7]. Evaluating seven different machine learning algorithms, including K-Nearest Neighbors, Decision Trees, Random Forests, Support Vector Machines, Multi-Layer Perceptron, Logistic Regression, and Naive Bayes, using a dataset of 655 students, the study found Random Forest algorithm to have the highest success rate.

A framework was proposed to provide career counseling to students using machine learning and artificial intelligence techniques [8]. The study examined how white-box and black-box models could analyze educational data to offer career advice to students, with Naive Bayes identified as the most effective model.

A systematic review was conducted on interpretable student performance prediction models between 2015 and 2020 [9]. The review examined various studies in the literature and discussed how interpretable student performance prediction models could be developed, summarizing different research methods and findings and providing recommendations for future research.

A study aimed to improve decision support systems using local interpretable machine learning modeling to predict student attrition [10]. By developing a model using a large dataset to predict students' academic performance and enrollment status, the study enhanced the accuracy of decision support systems and facilitated the identification of more effective interventions.

An investigation was described, employing artificial intelligence methods to assess academic achievement in general high schools in a European Union country [11]. The study developed a model using artificial intelligence algorithms to measure the academic achievements of general high school students, considering various factors to predict students' academic performance and provide a criterion for determining their success.

3 Materials And Methodology

In this section, a dataset encompassing socio-economic and personal factors influencing students' academic achievements was employed, comprising parameters such as gender, age, parental education levels, family structure, residence type, commute time to school, study duration, social activities, alcohol consumption, health status, and prior semester grades. Biogeography-Based Optimization (BBO) algorithm was utilized for feature selection, inspired by natural migration and distribution processes of species. Within the realm of machine learning, Artificial Neural Networks (ANN), XGBoost, LightGBM, Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Gradient Boosting algorithms were deployed, serving as data-driven modeling techniques to predict student success.

3.1 Dataset

In this study, a dataset containing socio-economic and personal factors affecting students' academic achievements has been utilized. The dataset includes parameters such as the student's gender, age, parents' education levels, family structure, type of residence, travel time to school, study duration, social activities, alcohol consumption, health status, and previous semester grades. An example of the dataset is provided in Figure 1.

	school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob	...	famrel	freetime	goout	Dalc	Walc	health	absences
0	GP	F	18	U	GT3	A	4	4	at_home	teacher	...	4	3	4	1	1	3	6
1	GP	F	17	U	GT3	T	1	1	at_home	other	...	5	3	3	1	1	3	4
2	GP	F	15	U	LE3	T	1	1	at_home	other	...	4	3	2	2	3	3	10
3	GP	F	15	U	GT3	T	4	2	health	services	...	3	2	2	1	1	5	2
4	GP	F	16	U	GT3	T	3	3	other	other	...	4	3	2	1	2	5	4
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

Figure 1: Example of raw data set used in the study.

3.2 Biogeography-Based Optimization (BBO)

The Biogeography-Based Optimization (BBO) algorithm is defined as an optimization algorithm developed by drawing inspiration from nature, modeling the migration and distribution processes of species to solve optimization problems. In the context of feature selection, BBO allows for the identification of the most important features by evaluating their contribution to the desired outcome. By simulating the natural migration process of species, BBO effectively explores the feature space and selects the features that have the greatest impact on the desired outcome [12].

3.3 Machine Learning

Machine learning is considered as a branch of artificial intelligence where computer systems are enabled to learn from data-driven experiences to solve complex problems. This method relies on algorithms having the capability to learn automatically from data in order to accomplish a specific task. Algorithms typically acquire knowledge by identifying patterns using large amounts of data and can make predictions or decisions based on these patterns. Machine learning intersects the fields of artificial intelligence and data science, and finds applications in various domains including image recognition, natural language processing, medical diagnosis, financial forecasting, and automation [13]. In this study, the Artificial Neural Networks (ANN), XGBoost, LightGBM, Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Gradient Boosting algorithms were utilized.

3.3.1 Artificial Neural Networks (ANN)

Artificial Neural Networks (ANN) are artificial intelligence models designed by drawing inspiration from biological neural networks in computer systems. These networks are based on structures that mimic the functioning of neurons in the human brain and are typically multi-layered. They are used to perform complex tasks such as data analysis and pattern recognition. ANNs can perceive and learn patterns in datasets, allowing them to make predictions or classifications. During the training process, the network is presented with large amounts of data, and it learns by adjusting its internal weights and structures based on this data. As a result, ANNs are widely used in various fields such as image recognition, natural language processing, speech recognition, and autonomous driving [14]. The formula equation 1, which describes the working structure of Artificial Neural Networks, is provided.

$$z = \sum_{i=1}^{n} w_i x_i + b \tag{1}$$

In equation 1, z represents the output of the neuron, which signifies the weighted sum of inputs. w_i denotes the weight of input i, indicating the importance assigned to that input. x_i represents the value of input i. b signifies the bias value, a constant added to the output of the neuron. n indicates the total number of inputs. These equations are used to compute the weighted sum of inputs for a neuron in artificial neural networks, determining the neuron's output.

3.3.2 XGBOOST

XGBoost is a widely used algorithm in recent years in machine learning, often preferred for solving classification and regression problems. This algorithm is a variant of the Gradient Boosting framework and is an optimized tree learning algorithm that provides fast and effective learning. XGBoost offers the advantage of combining many base models to achieve powerful predictive power. Additionally, when used in conjunction with techniques such as feature selection and model tuning, it typically delivers high accuracy and performance. Therefore, XGBoost has a wide range of applications, from industrial applications to competitions, and is preferred for obtaining effective results even with large datasets [15]. The most general formula used in the XGBoost algorithm is given in equation 2.

$$F(x) = L(\theta) + \Omega(\theta) \tag{2}$$

In Equation 2, (x) represents the general function predicted by the model, $L(\theta)$ denotes the loss function on the data, and $\Omega(\theta)$ signifies the regularization term controlling the complexity of the model.

$$L(\theta) = \sum_{i=1}^{n} -(y_i \, log \, log \, (\hat{y}_i) \, + (1 - \, y_i) \, log \, log \, (\hat{y}_i) \,) \tag{3}$$

In Equation 3, it represents the cross-entropy loss used in a classification model. (θ) denotes the loss function calculated based on the model's parameters. y_i represents the true class label, while $\hat{y}_i$represents the probability value predicted by the model. The cross-entropy loss measures the difference between the true class label and the probability distribution predicted by the model for each data point. If the true class label is y_i, the loss function is computed as $(\hat{y}_i)$; otherwise, it is calculated as $(1 - \, y_i \,)$ These losses gauge how far off the model's prediction is from the true class label. The total loss is computed as the sum of these error terms across all data points.

$$\Omega(\theta) = \, \gamma T + \, \frac{\lambda}{2} \sum_{j=1}^{T} w_j^2 \tag{4}$$

In Equation 4, $\Omega(\theta)$ represents the regularization term, where γ and λ are regularization parameters. T denotes the number of trees in the ensemble, and w_j represents the weight assigned to the jth tree. The regularization term consists of two parts: The first part, γT, penalizes the complexity of the model by multiplying the number of trees by a regularization parameter γ. The second part, $\frac{\lambda}{2} \sum_{j=1}^{T} \quad w_j^2$, applies L2 regularization to the weights of the trees, where λ controls the strength of regularization. This regularization term helps prevent overfitting by discouraging overly complex models and promoting smoother solutions.

3.3.3 LightGBM

LightGBM is a machine learning algorithm that has gained popularity recently and is often preferred for fast and high-performance modeling on large datasets. This algorithm is a variant of the Gradient Boosting framework and has a structure that enables faster training times and lower memory usage. LightGBM is designed to model complex relationships using tree-based learning methods and is typically successful in dealing with large-scale datasets and high-dimensional feature spaces. Additionally, its ability to directly support categorical features and its parallel computation capabilities have made it a preferred choice for large-scale machine learning problems. Therefore, LightGBM is widely used in various applications, especially in industrial applications and large-scale data analysis projects [16].

$$F_m = \, F_{m-1}(x) + \, \eta \, \cdot \, h_m(x) \tag{5}$$

Equation 5 represents the update rule used in the LightGBM (Light Gradient Boosting Machine) algorithm. F_m represents the prediction of the model at the mmth iteration, while $F_{m-1}(x)$ is the prediction of the model from the previous iteration. η is the learning rate, and $h_m(x)$ is the candidate tree. In LightGBM, in each iteration, the weighted version of the candidate tree $h_m(x)$ is added to the prediction of the current model. This candidate tree is trained to focus on further reducing the residuals in the dataset, aiming to improve the performance of the model.

3.3.4 Random forest (RF)

Random Forest (RF) is a widely used algorithm in machine learning, often preferred for solving classification and regression problems. This algorithm is an ensemble of multiple decision trees that come together. Each decision tree is trained on randomly selected subsets of features and data points, resulting in independent predictions. These predictions are then combined by averaging them or by selecting the most common class or value. Random Forest is resilient to overfitting and typically provides high accuracy. Additionally, it can effectively scale to handle large-scale datasets due to its parallel computation capabilities. Therefore, Random Forest has a broad range of applications and is successfully used in many fields, from industrial problems to computational biology [17].

$$f(x) = \frac{1}{N}\sum_{i=1}^{N} f_i(x) \tag{6}$$

Equation 6 represents the fundamental prediction function used in the Random Forest (RF) algorithm. $f(x)$ denotes the prediction made for the input x. For each tree i, $f_i(x)$ represents the prediction of that tree. The overall prediction is obtained by averaging the predictions of all trees. In other words, $f(x)$ is the average of predictions from all trees for the data point x. This approach is a fundamental feature of the Random Forest algorithm, allowing for stronger and more generalizable predictions by combining predictions from multiple decision trees.

3.3.5 Support Vector Machine (SVM)

Support Vector Machine (SVM) is a powerful classification and regression algorithm widely used in the field of machine learning. This algorithm aims to find an optimized hyperplane to separate the data. SVM utilizes this hyperplane to determine the class of a particular data point, thereby achieving the widest possible margin between classes. Additionally, SVM provides flexibility through kernel functions, enabling it to handle both linearly separable and non-linearly separable datasets. Consequently, SVM is recognized as an effective tool for classifying complex datasets and often performs well in high-dimensional data spaces and with small datasets [17]-[18].

$$f(x) = w^T x + b \tag{7}$$

Equation 7 represents a linear decision boundary. $f(x)$denotes the predicted class for input feature vector x. w and b are the parameters to be learned by the model; w represents the weights associated with the feature vectors, and b is the bias term of the model.

$$margin = \frac{2}{\|w\|} \tag{8}$$

Equation 8 calculates the marginal distance around the decision boundary. $\|w\|$ represents the norm of the normal vector of the decision boundary (the magnitude of w).

$$minimize \;\; \frac{1}{2}\|w\|^2 \tag{9}$$

Equation 9 represents the loss function used during the training phase of SVM. The goal is to optimize the weights w used to determine the decision boundary. This equation is applicable for linearly separable classes.

$$K(x_i, x_j) = \phi(x_i)^T \phi(x_j) \tag{10}$$

Equation 10 represents the usage of kernel functions in SVM to enhance its flexibility. It denotes the inner product between two input vectors x_i and x_j Kernel functions transform data points from the input space to higher-dimensional spaces, making them linearly separable, especially when they are not linearly separable in the original input space.

3.3.6 K-Nearest Neighbors (KNN)

K-Nearest Neighbors (KNN) is a popular algorithm, especially for classification and regression tasks. This algorithm operates by calculating the distance between the input data point and neighboring points in the feature space. Subsequently, KNN determines the class or value of the input data point based on the classes or values of its nearest neighbors. The "k" in KNN represents the number of neighbors considered for classification or regression. KNN is a non-parametric algorithm as it does not make any assumptions about the underlying data distribution. Additionally, it is a versatile algorithm capable of handling both numerical and categorical data effectively. However, despite its simplicity and interpretability, KNN may face computational inefficiency with large datasets [18].

$$d(x_i, x_j) = \sqrt{\sum_{p=1}^{n} (x_{i,p} - x_{j,p})^2} \tag{11}$$

Equation 11 calculates the Euclidean distance between two data points x_i and x_j. $x_{i,p}$ and $x_{j,p}$ represent the feature values of the pth dimension of the data points.

$$y = mode(y_1, y_1, \dots, y_k) \tag{12}$$

Equation 12 represents the fundamental rule of the KNN algorithm for classification problems. To determine the class of a data point, the classes of its k nearest neighbors are collected, and the most frequently occurring class is assigned to the data point.

$$y = \frac{1}{k}\sum_{i=1}^{k} y_i \tag{13}$$

Equation 13 represents the fundamental rule of the KNN algorithm for regression problems. To determine the value of a data point, the values of its k nearest neighbors are collected, and the average of these values is used as the estimated value for this data point.

3.3.7 GRADİENT BOOSTİNG

Gradient Boosting is an effective ensemble learning technique widely used in machine learning. This technique enables the creation of a strong predictor by combining base models called weak learners. Gradient Boosting trains each base model to correct the errors of the previous model. This way, each successive model is improved to address the weaknesses of the previous models, thereby enhancing the overall prediction power. This process is optimized by carefully tuning hyperparameters such as the learning rate. Gradient Boosting yields successful results in various tasks, including classification and regression problems, and is commonly implemented with popular libraries such as XGBoost and LightGBM [19].

$$F_0(x) = 0 \tag{14}$$

Equation 14 represents the initial prediction of the Gradient Boosting algorithm. Initially, predictions start with zero.

$$F_m(x) = F_{m-1}(x) \ + \ \rho * h_m(x) \tag{15}$$

Equation 15 represents the fundamental update rule of the Gradient Boosting algorithm. $F_m(x)$ denotes the prediction at the mmth iteration, $F_{m-1}(x)$ represents the prediction from the previous iteration, $h_m(x)$ specifies the prediction function of the new model added in the mth iteration, and ρ indicates the contribution rate of the new model.

3.4 Explainable AI (XAI)

Explainable AI (XAI) is an approach aimed at making the decisions and inferences of artificial intelligence systems understandable and traceable by humans. XAI develops methods to explain the inner workings of complex AI models, allowing us to understand why decisions are made, how input data is processed, and how results are obtained. As a result, users can confidently accept the decisions of AI systems and question the logic behind them when necessary. XAI is recognized as an important tool for developing reliable and transparent AI systems in various fields, from medical diagnostics to financial risk analysis, autonomous vehicles, and security systems [20].

3.4.1 SHAP

SHAP (SHapley Additive exPlanations) is a technique used to explain predictions of complex artificial intelligence models. It particularly aids in understanding the inner workings of models like deep learning models, often referred to as black boxes, which are not easily interpretable. At the core of SHAP lies the concept of Shapley values from cooperative game theory. These values quantitatively measure the contribution of each feature to a prediction outcome. SHAP is utilized to understand the impact of each feature on a prediction outcome and visually elucidate the internal structure of the model. Consequently, it enhances the transparency of AI models' decisions and improves the model's reliability [21].

$$\phi_i(f) = \frac{1}{N!} \sum_{\pi} \left[f\left(x_{\pi(i)}\right) - f(x_\pi) \right] \tag{16}$$

Equation 16 represents a formula used in the SHAP (SHapley Additive exPlanations) algorithm. Here, $\phi_i(f)$ denotes the SHAP value of the ith feature of the model f. f represents the model's prediction function, N denotes the number of data points, x_π represents a permutation of π, and $x_{\pi(i)}$ represents the value of the ith feature in a permutation of π. This equation calculates the contribution of each feature to a prediction outcome as the SHAP value. This contribution is determined by averaging over all permutations. SHAP values quantify the impact of each feature on the model's predictions and make the model's decisions more interpretable.

3.4.2 LIME

LIME (Local Interpretable Model-agnostic Explanations) is a technique used to explain the decisions of complex artificial intelligence models. It particularly aids in understanding the behaviors of models termed as black boxes, whose internal workings are not fully understood. LIME is designed to interpret the predictions of these models locally. Its working principle is as follows: First, an example for which the prediction of the model to be explained is given. For instance, consider wanting to explain the prediction made by an image classification model for a specific image. LIME generates randomly perturbed examples around this example and evaluates the predictions of the model on these examples. Then, using the effects of these random examples on the model, LIME attempts to determine which features or variables determine the prediction. For example, to understand which pixels are most influential in the classification of an image, LIME analyzes the effects of the modified versions of pixels on the model prediction. Finally, the output of LIME is usually presented as a simple and understandable model. This model represents the local behavior of the original complex model and is easier to interpret. LIME is an effective tool used especially to make complex artificial intelligence models more understandable [22].

$$e(x) = argmin_{g \in ς}(f, g, \pi_x) + \Omega(g) \tag{17}$$

Equation 17 represents a formula used in the LIME (Local Interpretable Model-agnostic Explanations) algorithm. Here, $e(x)$ denotes the best interpretable model for a specific data point x. The $argmin$ operator finds the minimum among all models g in a certain set $ς$, considering the smallest loss (f, g, π_x) and the complexity term $\Omega(g)$. This equation represents the process of finding the most suitable interpretable model for a specific data point in the LIME algorithm. LIME focuses on a small region around the data point to explain how the data point is influenced by the complex model, thereby enabling local interpretation of the model's decisions.

4 Results

In this study, various preprocessing steps have been applied to the dataset, with the main aim of enabling more effective utilization of categorical variables. Particularly, categorical variables in string format in the dataset have been converted to integer types to enhance the processing efficiency of models. This transformation process has included variables such as students' gender, age range, and parents' education level. These adjustments have facilitated better understanding of the dataset and yielded more accurate results. Additionally, the transformation has enabled algorithms working with numerical values to operate more effectively, leading to more reliable analysis outcomes. An example of the dataset used in the analysis is provided in Figure 2.

	school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob	...	romantic	famrel	freetime	goout	Dalc	Walc	health	absences	Lecture_Notes	Course_Pass_Status
0	0	0	18	0	1	1	4	4	0	4	...	0	4	3	4	1	1	3	6	28.333333	0
1	0	0	17	0	1	0	1	1	0	2	...	0	5	3	3	1	1	3	4	26.666667	0
2	0	0	15	0	0	0	1	1	0	2	...	0	4	3	2	2	3	3	10	41.666667	0
3	0	0	15	0	1	0	4	2	1	3	...	1	3	2	2	1	1	5	2	73.333333	1
4	0	0	16	0	1	0	3	3	2	2	...	0	4	3	2	1	2	5	4	43.333333	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1039	1	0	19	1	1	0	2	3	3	2	...	0	5	4	2	1	2	5	4	51.666667	1
1040	1	0	18	0	0	0	3	1	4	3	...	0	4	3	4	1	1	1	4	76.666667	1
1041	1	0	18	0	1	0	1	1	2	2	...	0	1	1	1	1	1	5	6	53.333333	1
1042	1	1	17	0	0	0	3	1	3	3	...	0	2	4	5	3	4	2	6	50.000000	0
1043	1	1	18	1	0	0	3	2	3	2	...	0	4	4	1	3	4	5	4	53.333333	1

Figure 2 Dataset Example

The Biogeography-Based Optimization (BBO) algorithm has been effective in identifying complex relationships within the dataset and allowing the selection of attributes that are appropriate for the complexity among features. Therefore, the BBO algorithm is considered a suitable option in determining student success. In this study, using the BBO algorithm, the 20 attributes that most influence student success were identified from among the 32 attributes in the dataset.

With the determination of the most important 20 attributes using the BBo algorithm, these attributes were included in machine learning models, and metrics such as Accuracy, Precision, Recall, F1-score were used for their evaluation. The results of the evaluation conducted with the most important 20 attributes identified by the BBo algorithm are presented in Table 1 below.

Table 1: *The results of the evaluation conducted with the 20 most important attributes identified by the BBo algorithm*

Model	**Accuracy**	**Precision**	**Recall**	**F1-score**
Gradient Boosting*	0.7388	0.7311	0.7388	0.7234
Random Forest (RF)	0.7252	0.7414	0.7452	0.7290
XGBoost	0.7188	0.7317	0.7388	0.7313
LightGBM	0.7124	0.7244	0.7324	0.7220
K-Nearest Neighbors (KNN)	0.6815	0.6646	0.6815	0.6650
Support Vector Machine (SVM)	0.6514	0.6637	0.6624	0.5722
Artificial Neural Networks (ANN)	0.6624	0.6478	0.6624	0.6504

The confusion matrix of the Gradient Boosting algorithm with the highest Accuracy value is presented in Figure 3 based on the evaluation results obtained with the BBo algorithm using the most important 20 features.

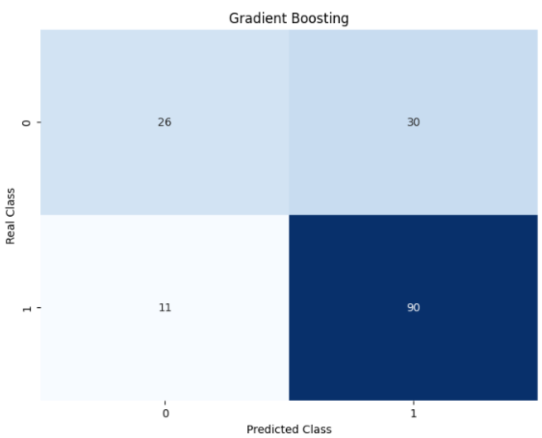

Figure 3 The output of the Gradient Boosting Confusion Matrix

- True Positive (TP): Represents the case where the model accurately predicts the positive class.
  TP = 90
- True Negative (TN): Represents the case where the model accurately predicts the negative class.
  TN = 11
- False Positive (FP): Represents the case where the model incorrectly predicts the positive class.
  FP = 30
- False Negative (FN): Represents the case where the model incorrectly predicts the negative class.
  FN = 26

The accuracy metric is a performance measure that evaluates the ratio of correctly predicted samples by a classification model. This metric is calculated based on True Positives (TP) and True Negatives (TN) values. A high accuracy value indicates that the model has a high ability to make correct predictions with few incorrect predictions. The accuracy metric is effective when all classes are correctly classified and can be used in balanced class distributions. However, in cases of imbalanced class distributions (for example, situations where there are rare classes), the accuracy metric may be inadequate because incorrect predictions of rare classes can misleadingly inflate the overall accuracy value. The calculation of the accuracy metric is shown in Equation 18.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{18}$$

The precision metric is a performance measure that evaluates how many of the samples predicted as positive by a classification model are actually positive. This metric is calculated based on the True Positive (TP) and False Positive (FP) values. A high precision value indicates that most of the samples predicted as positive are indeed positive, with few false positive predictions. The calculation of the precision metric is shown in Equation 19.

$$Precision = \frac{TP}{TP + FP} \tag{19}$$

The recall metric is a performance measure that evaluates how many of the true positives a classification model correctly predicts. This metric is calculated based on the True Positive (TP) and False Negative (FN) values. A high recall value indicates that the model correctly predicts most of the true positives, with few false negative predictions. The recall metric is particularly important in areas of vital importance, such as medical diagnosis, and in situations where missing true positives is significant. The calculation of the recall metric is shown in Equation 20.

$$Recall = \frac{TP}{TP + FN} \tag{20}$$

The F1 metric is the harmonic mean of a classification model's precision and recall metrics, aiming to balance the impact of both false positives and false negatives. This metric is calculated based on True Positive (TP), False Positive (FP), and False Negative (FN) values. A high F1 value indicates that both precision and recall metrics are high, and the model reduces both false positives and false negatives. The F1 metric is particularly useful in cases of imbalanced class distributions (for example, situations where there are rare classes), as precision and recall metrics may be inadequate in these scenarios. The calculation of the F1 metric is shown in Equation 21.

$$F1 = \frac{Precision \times Recall}{Precision + Recall} \tag{21}$$

The Gradient Boosting algorithm, which yielded the highest accuracy value, has been explained using interpretable artificial intelligence models SHAP and LIME. The explanations are provided in Figures 3 and 4.

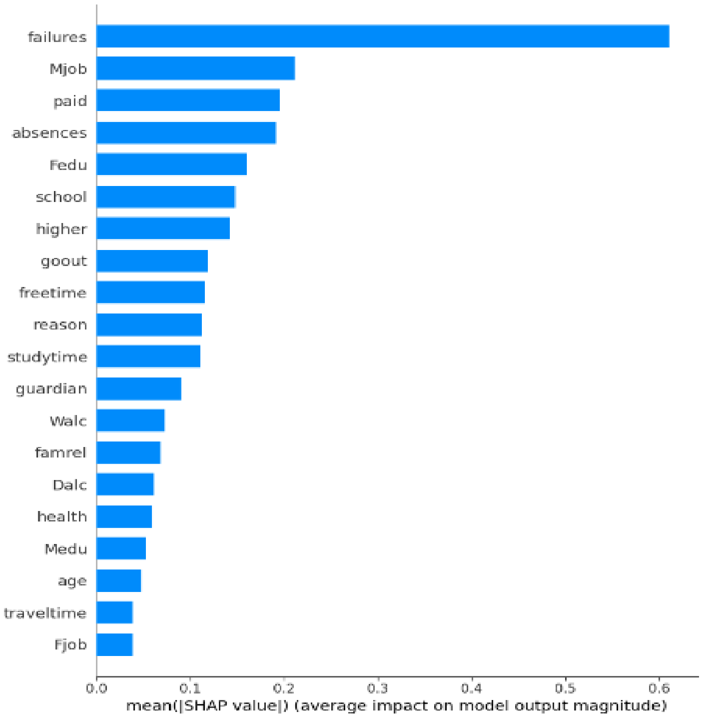

Figure 4 SHAP output, which is an explainable artificial intelligence model.

In Figure 4, a graph is presented illustrating the average effect size of a model's output and representing an explanation method that measures the individual contribution of each feature to the model's prediction using SHAP (SHapley Additive exPlanations) values. The graph displays the average absolute SHAP value on the x-axis and the average effect size on the y-axis. The features depicted in the graph include the number of 'failures', representing the instances of student failure, 'Mjob', indicating the mother's occupation, 'paid', denoting the student's employment status, and 'absences', representing the number of days the student has been absent from school. Among the conclusions that can be drawn from the graph, it is observed that the number of student failures and the mother's occupation have the greatest impact on the model's output, while the effects of other features are comparatively smaller. However, it is important to note that the graph may not fully represent the performance of the model for a specific student, and it is crucial to acknowledge that the effects of each feature are not linear and may not account for interactions between features.

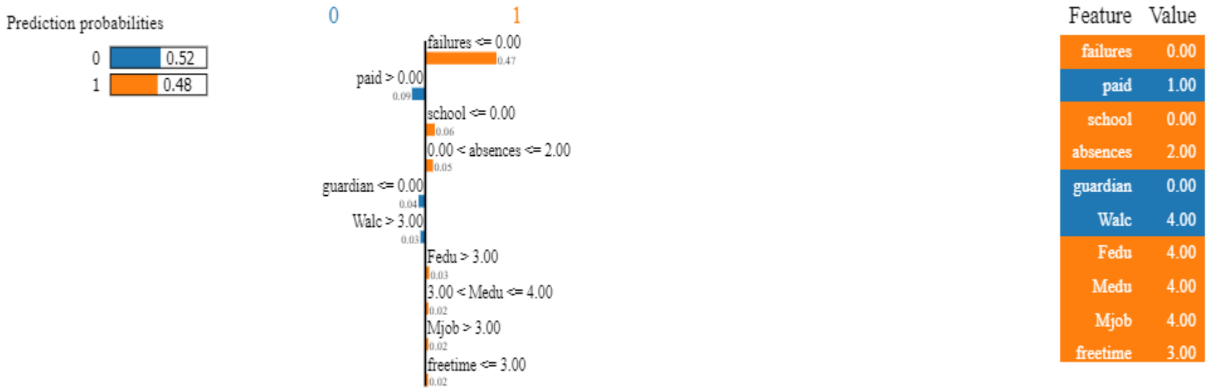

Figure 5 LIME output, which is an explainable artificial intelligence model.

Figure 5 contains information regarding student data and predictions made based on this data. The features include factors such as the number of failures, amount paid, type of school, number of absences, guardian status, test scores, and parental occupation. The effects of these features on the model prediction are explained using the LIME method, which demonstrates how the prediction changes with each feature value alteration. The graph highlights that the Walc, Fedu, and Medu test scores have the most significant impact on the model's prediction. This information indicates that these factors increase the likelihood of student success. However, it is also noted that other features contribute to the model's prediction, providing insights into how the prediction is generated. Thus, the LIME analysis serves as a valuable tool for educators and other stakeholders by making the model predictions more understandable.

5 Conclusions

The analyses conducted emphasize the necessity of examining the factors influencing student success in detail. Specifically, the discussion of how these factors impact learning outcomes is supported by theoretical frameworks. This study applied various preprocessing steps to the dataset, aiming to enable more effective utilization of categorical variables. Particularly, categorical variables in string format were converted to integer types, including variables such as students' gender, age range, and parents' education level. These adjustments facilitated better understanding of the dataset and yielded more accurate results, enhancing the processing efficiency of the models.

Furthermore, the evaluation of algorithms and interpretable artificial intelligence models used in determining student success demonstrates how data-driven decision-making processes in education can be shaped. The Biogeography-Based Optimization (BBO) algorithm was particularly effective in identifying complex relationships within the dataset, selecting the 20 most influential attributes from among the 32 attributes. This selection improved the performance of machine learning models, as evidenced by metrics such as Accuracy, Precision, Recall, and F1-score.

The results indicate that models like Gradient Boosting and Random Forest performed well, with Gradient Boosting achieving an accuracy of 0.7388 and Random Forest achieving an accuracy of 0.7252. These findings highlight the potential of advanced algorithms to provide reliable insights into student success.

Additionally, methods like SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) were used to interpret the model predictions. The SHAP analysis revealed that features such as the number of failures and the mother's occupation had the greatest impact on the model's output. Similarly, LIME analysis indicated that factors like Walc, Fedu, and Medu test scores significantly influenced the model's predictions.

In light of these findings, there is a need for further research to create and implement education policies more effectively. Future studies should focus on deeper exploration of the factors influencing student success and better understanding the interactions among these factors. Additionally, there is a necessity to enhance the role of artificial intelligence techniques in education and integrate these techniques more effectively into educational practices.

The results of this study could contribute to the development of data-driven decision-making processes in education and more effectively planned interventions to improve student success. Moreover, this study serves as an important resource for educators and policymakers, aiding in the more sustainable and inclusive enhancement of the education system.

6 Declarations

6.1 Study Limitations

None.

6.2 Acknowledgement

The author would like to thank all the data sets, materials, information sharing and support used in the assembly of this article.

6.3 Funding source

The author gratefully acknowledge the financial support of Sakarya University of Applied Sciences AI And Data Science Research And Application Center.

6.4 Competing Interests

There is no conflict of interest in this study.

6.5 Author contributions

Cem Özkurt wrote and reviewed the article.

References

[1] Z. Akhtar, "Socio-economic status factors effecting the students achievement: a predictive study," *International Journal of Social Sciences and Education*, vol. 2, no. 1, pp. 281-287, 2012.

[2] Lakhan, G. R., Soomro, B. A., & Channa, A. (2021). INVESTIGATION OF THE SOCIO-ECONOMIC FACTORS THAT INFLUENCE YOUNG LEARNERS ACADEMIC SUCCESS: A CASE STUDY OF SECONDARY SCHOOLS OF SINDH, PAKISTAN. *New Horizons (1992-4399)*, 15(1).

[3] Marks, G. N. (2016). The relative effects of socio-economic, demographic, non-cognitive and cognitive influences on student achievement in Australia. *Learning and Individual Differences*, 49, 1-10.

[4] Singh, P., & Choudhary, G. (2015). Impact of socio-economic status on academic achievement of school students: An investigation. *International journal of applied research*, 1(4), 266-272.

[5] Albashish, D., Hammouri, A. I., Braik, M., Atwan, J., & Sahran, S. (2021). Binary biogeography-based optimization based SVM-RFE for feature selection. *Applied Soft Computing*, 101, 107026.

[6] Lau, E. T., Sun, L., & Yang, Q. (2019). Modelling, prediction and classification of student academic performance using artificial neural networks. *SN Applied Sciences*, 1(9), 982.

[7] Şahin, S., & Erol, Ç. (2024). Prediction of Secondary School Students' Academic Achievements with Machine Learning Methods and a Sample System. *Cybernetics and Systems*, 55(4), 940-960.

[8] Guleria, P., & Sood, M. (2023). Explainable AI and machine learning: performance evaluation and explainability of classifiers on educational data mining inspired career counseling. *Education and Information Technologies*, 28(1), 1081-1116.

[9] Alamri, R., & Alharbi, B. (2021). Explainable student performance prediction models: a systematic review. *IEEE Access*, 9, 33132-33143.

[10] Delen, D., Davazdahemami, B., & Rasouli Dezfouli, E. (2023). Predicting and mitigating freshmen student attrition: A local-explainable machine learning framework. *Information Systems Frontiers*, 1-22.

[11] Cruz-Jesus, F., Castelli, M., Oliveira, T., Mendes, R., Nunes, C., Sa-Velho, M., & Rosa-Louro, A. (2020). Using artificial intelligence methods to assess academic achievement in public high schools of a European Union country. *Heliyon*, 6(6).

[12] R. Hans and H. Kaur, "Hybrid Biogeography-Based Optimization and Genetic Algorithm for Feature Selection in Mammographic Breast Density Classification," *International Journal of Image and Graphics*, vol. 22, no. 03, p. 2140007, 2022.

[13] K. Bakshi and K. Bakshi, "Considerations for artificial intelligence and machine learning: Approaches and use cases," in *2018 IEEE Aerospace Conference*, 2018, pp. 1-9.

[14] Z. Zhang, "A gentle introduction to artificial neural networks," *Annals of translational medicine*, vol. 4, no. 19, 2016.

[15] X. Y. Liew, N. Hameed, and J. Clos, "An investigation of XGBoost-based algorithm for breast cancer classification," *Machine Learning with Applications*, vol. 6, p. 100154, 2021.

[16] A. Villar and C. R. V. de Andrade, "Supervised machine learning algorithms for predicting student dropout and academic success: a comparative study," *Discover Artificial Intelligence*, vol. 4, no. 1, pp. 1-24, 2024.

[17] 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," in *Proceedings of the 2020 3rd International Conference on Education Technology Management*, December 2020, pp. 100-107.

[18] H. Al-Shehri, A. Al-Qarni, L. Al-Saati, A. Batoaq, H. Badukhen, S. Alrashed, J. Alhiyafi, and S. O. Olatunji, "Student performance prediction using Support Vector Machine and K-Nearest Neighbor," in *2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE)*, 2017, pp. 1-4. doi: 10.1109/CCECE.2017.7946847.

[19] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, "A comparative analysis of gradient boosting algorithms," *Artificial Intelligence Review*, vol. 54, pp. 1937-1967, 2021.

[20] A. Adadi and M. Berrada, "Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52138-52160, 2018. doi: 10.1109/ACCESS.2018.2870052.

[21] I. U. Ekanayake, D. P. P. Meddage, and U. Rathnayake, "A novel approach to explain the black-box nature of machine learning in compressive strength predictions of concrete using Shapley additive explanations (SHAP)," *Case Studies in Construction Materials*, vol. 16, p. e01059, 2022.

[22] K. R. Chowdhury, A. Sil, and S. R. Shukla, "Explaining a black-box sentiment analysis model with local interpretable model diagnostics explanation (LIME)," in *Advances in Computing and Data Sciences: 5th International Conference, ICACDS 2021, Nashik, India, April 23–24, 2021, Revised Selected Papers, Part I*, vol. 5, pp. 90-101, Springer International Publishing, 2021.