[1*] Tao-Hongli

Educational data mining for student performance prediction: feature selection and model evaluation

Journal of Electrical Systems

Abstract: - Student performance is a multidimensional concept that includes academic accomplishment and cognitive development. It measures the performance of educational institutions, teaching methods and individual efforts. Engagement, motivation and support networks have a substantial impact on performance results. In this research, we intend to establish an intelligent data mining-based model for predicting student performance through a novel feature selection approach. We propose an innovative Adaptive Sea Horse Optimization (ASHO) to select the crucial features to predict the education performance level. We gathered a xAPI-Edu-Data dataset which includes students' numerous academic details, to train our suggested model. The Z-score normalization algorithm is employed to pre-process the obtained raw data, it improves the quality of the data. We utilized Independent Component Analysis (ICA) to extract relevant characteristics from the processed data. We utilize the ASH algorithm for feature selection, it dynamically adjusts search parameters, efficiently exploring the feature space to locate the ideal subset of features for improving the predictive performance. The selected features are classified by the implementation of the eXtreme Gradient Boosting (XGBoost) algorithm for predicting student performance. Our recommended approach is implemented in Python software. The finding evaluation phase examines the suggested model's prediction effectiveness with various parameters such as precision, accuracy, recall and f1 score. We performed a comparison analysis with other traditional methods to determine the effectiveness of the proposed approach. The experimental results demonstrate that the proposed prediction approach performed better than other existing approaches.

Keywords: Student Performance Prediction, Machine Learning (ML), Adaptive Sea Horse Optimization (ASHO), eXtreme Gradient Boosting (XGBoost), Education, Academic.

I. INTRODUCTION

The use of Data Mining (DM) strategies in the discipline of education has garnered a variety of interest. Collecting records is the method of DM. It is the field of using big information to extract vital discoveries or sparkling, perhaps useful facts [1]. It additionally desires to find unique developments and patterns from massive information units through the usage of diverse types of techniques [2]. The use of conventional facts mining techniques to deal with issues in education is called educational data mining, or EDM. EDM is the software of DM techniques for instructional statistics, which include scholar demographics and academic histories, taking a look at ratings, attendance and the frequency of queries from college students [3]. EDM is currently a beneficial tool for predicting instructional accomplishment, locating hidden trends in academic facts and enhancing the getting-to-know and coaching environments.

The use of EDM has given learning analytics a new perspective. Learning analytics includes gathering student data in a variety of ways, analyzing and interpreting it to get a deeper knowledge of the learning environment and identifying the most effective teachers and students [4-5]. To comprehend and improve instruction as well as the settings in which it occurs, analytics for learning is the gathering, calculating and analyzing of information regarding students and their situations. It also covers how the organizations are creating new plans [6].

Data mining methods used to examine educational data are referred as learning data mining. Large amounts of data are stored by universities to monitor students, instructors and courses [7]. This data includes syllabus, test papers, announcements and other materials, as well as academic and private information on teachers and students. Educational data mining is being used by several institutions and independent groups to enhance the quality of life for their professors and students [8]. To make their application programs consistent with their databases, these techniques are integrated into systems.

Student performance is one of the most essential requirements for every university. Students' performance could be predicted by looking at their past academic records [9]. It suggests that there could be a relationship between a student's aptitude and interests as well as performance. When educators use this type of analysis, they can offer particular focus to the pupils who most need attention [10]. One common way to assess a teacher's effectiveness is by their students' performance. Every school must assess the strength of its faculty.

[1] *Corresponding author: Tao-Hongli, ddyy19820222@163.com
Computer Engineering Technical College (Artificial Intelligence College), Guangdong Polytechnic of science and technology, Zhuhai, 519090, China

Teachers could be evaluated based on their performance, comments and other student-generated data. This kind of study could assist an organization in raising the quality of its instruction [11]. Users can assess test papers to determine the difficulty level. An institution could normalize all students' grades in multi-session exams with the use of this information [12]. Predictive models could fail to account for all of the details that affect student performance, even if they could make use of a variety of characteristics including socioeconomic status, past academic records and demographics. In this work, we provide a novel approach called ASHO to identify the critical characteristics needed to forecast the degree of academic success.

Contributions：

1. The xAPI-Edu-Data were gathered from kaggle.
2. Z-score normalization standardizes records by subtracting the mean and dividing it by the standard deviation, ensuring regular scales across features.
3. The process of dividing data into statistically independent components allows for the extraction of significant patterns from the data using ICA.
4. Inspired by the behavior of sea horses, ASHO is a revolutionary feature selection technique that maximizes feature relevance for improved accuracy in tasks such as predicting student achievement.
5. XGBoost algorithm is used for predicting the student performance.

The remaining study components could be classified: We will discuss the related works in section 2. The approaches are discussed in section 3. Section 4 presents the experiment's results. The conclusion is covered in section 5.

II. Related works

Study [13] provided instructors with an in-depth manual if they were interested in using data mining approaches to forecast student success. They have thoroughly examined all of the pertinent works, gathered the state of the art and developed a systematic strategy whereby each choice and parameter were thoroughly explained along with the reasons behind its selection. Their full capacity for use in education could be unlocked by the task. Study [14] forecasted students' achievement in a course based on their past performance in similar courses. A group of methods known as "data mining" were used to find patterns buried within enormous volumes of already-existing data. The trends could prove useful for forecasting and analysis. The gathering of data for use in the context of education was referred as educational data mining. The analysis of learners and instructor data was the focus of those programs. A study [15] suggested a novel model that forecasts undergraduate students' end test scores using machine learning algorithms with their midterm test outcomes as the input data. Those kinds of creating an environment for learning evaluation in higher education and to support decision-making, data-driven investigation was essential. Ultimately, the study determined the best machine-learning strategies and influenced the early identification of pupils who were likely to fail.

A study [16] organized to use data mining and video learning analytics approaches to forecast students' final performance during the semester. Eight distinct classification algorithms were used to examine data from mobile apps, the educational management system and the student database. Additionally, to minimize the features, preprocessing and data conversion techniques were used. Study [17] focused on methods to employ data mining methods to forecast applicants' academic success at colleges to assist institutions with their admissions choices. The findings showed that, depending on certain pre-admission requirements, candidates' initial university performance could be predicted before admission. The results also showed that the Scholastic Success Admission Test score was the pre-admission factor that most accurately predicted future success in school. Study [18] presented a model for predicting students' academic achievement using supervised machine-learning techniques, such as logistic regression and support vector machines. When compared to logistic regression, the outcomes of several trials conducted with different technologies demonstrated that the sequential minimum optimization method operates better by obtaining higher accuracy.

Study [19] proposed to examine the efficacy of deep neural network transfer learning for the objective of predicting students' success in college. They consider the work as a significant step forward since there hasn't been much research done on the creation of predictive algorithms in the area of educational data mining using transfer learning techniques. The findings of the experiment showed that, in the majority of situations, it was possible to predict the probability of failing with a sufficient level of accuracy. Study [20] provided a thorough examination of machine learning methods for predicting students' final grades in first-semester programs by enhancing predictive accuracy efficiency. Relevant and encouraging results were shown by the suggested model, which could enhance the prediction efficiency approach for unbalanced multiple classes for student grade forecasting. Study

[21] used a deep artificial neural network to anticipate at-risk pupils and provided preventative measures for such situations based on a collection of special handcrafted elements that were derived from the traffic data of online classrooms. It was shown that students who were interested in obtaining the material from the previous classes performed better. The goal of the research was to help higher education decision-makers create the framework that was required for pedagogical assistance to promote educational sustainability.

Study [22] combined the attention process with the understanding tracer model and fully utilized the qualities of both pupil conduct and exercise to present a unique framework for the forecasting of student success. Next, to predict student performance, a fusion attention system built on a recurrent neural network structure was used. Comprehensive tests on an actual dataset demonstrate the viability and efficiency of their method. Study [23] suggested an innovative Performance Factors Analysis (PFA) strategy built on several models to improve the student performance prediction's accuracy. It concentrated on using ensemble learning techniques as a highly productive machine learning model that was used to produce several innovative approaches throughout numerous domains. The experimental findings demonstrated that the scalability XGBoost performed better than the other models that were assessed. Study [24] investigated blended learning through the use of a small private online course (SPOC) and an instructional strategy based on a flipped environment. An analysis was conducted on the effects of general online educational behavior on student performance. The findings indicated that performance could be anticipated by students based on their online activity and as the course progresses, the accuracy and stability of the predictions increase. Study [25] analyzed the Bidirectional Long Short-Term Memory (BiLSTM) network, a deep neural network model that relies on attention, to effectively forecast student achievement (grades) based on past data. Through examining presented difficulties that center on sophisticated feature categorization and prediction, they have used the most sophisticated BiLSTM in conjunction with an attention mechanism framework in the study. The ability to anticipate success early was crucial for academics, institutions and government agencies.

Study [26] examined two separate undergraduate datasets from two distinct colleges. Additionally, the goal of that effort was to forecast students' success at two points in the instructional delivery process. The research makes it possible to choose the best machine learning algorithms to apply and optimize their parameters. Experimental findings demonstrated the excellent accuracy of the suggested bagging ensemble models for the intended group on both datasets. Study [27] illustrated the significance of the behavioral characteristics of the students. To achieve that, they collected educational information from a learning management system (LMS). The included dataset was subjected to feature analysis. Following that, data preprocessing a crucial stage in the information discovery process was applied. They used typical ensemble techniques, managed to improve the outcome and showed the dependability of the suggested model. Study [28] examined the effectiveness of deep neural networks in the area of EDM, particularly in forecasting students' academic success, to identify individuals who could struggle. Undergraduate programs' "Programming" and "Data Structures" courses provided a variety of challenges for students, which explained why failing and dropout rates were high. To forecast students' academic achievement, EDM was employed to evaluate student data collected from diverse educational environments. Research [29] suggested to use an improved conditional generative adversarial network-based deep support vector machine (ICGAN-DSVM) method to forecast how well students would do while receiving supported learning through tutoring from their families and schools. Results also indicated that, when compared to either family coaching and only school education, including both types of tutoring into the forecasting model could improve achievement even more. Based on performance metrics, the suggested ICGAN-DSVM exceeded similar efforts. Study [30] provided a thorough explanation of each theory and discussed how EDM could enhance students' comprehension and implementation of the ideas in classrooms. Through the extraction of important insights from large educational datasets, EDM was an exciting area of study that used computational tools to enhance educational results. The research also showed how EDM could assist to create individualized as well as adaptable learning environments and guide the creation of successful teaching practices.

III. Methodology

We collected a xAPI-Edu-Data including a wide range of academic information about students. To pre-process the acquired raw data and enhance its quality, the Z-score normalization procedure is used. To extract essential properties from the data that is processed, we used Independent Component Analysis (ICA). The eXtreme Gradient Boosting (XGBoost) technique is used to classify the chosen characteristics to forecast student performance. We suggest a novel method called Adaptive Sea Horse Optimization (ASHO) to determine which criteria are essential for predicting the degree of education achievement. Figure 1 shows the overall flow.

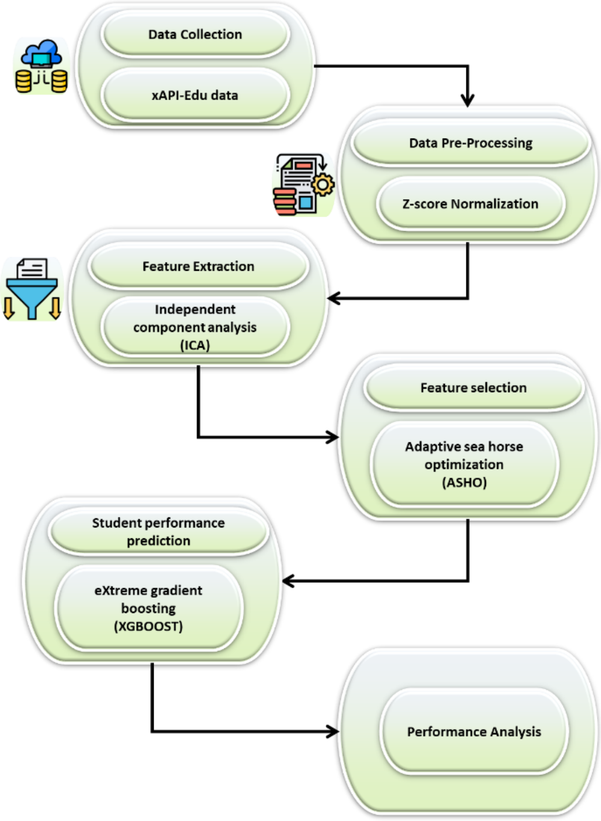
Data Collection
xAPI-Edu data
Data Pre-Processing
Z-score Normalization
Feature Extraction
Independent component analysis (ICA)
Feature selection
Adaptive sea horse optimization (ASHO)
Student performance prediction
eXtreme gradient boosting (XGBOOST)
Performance Analysis

Figure 1: overall flow

A. Dataset

The xAPI- Edu-Data were gathered from Kaggle [31]. There are 16 characteristics and 480 student entries in the dataset. Three main categories are used to classify the features, (1) Two examples of societal characteristics are ethnicity and gender. (2) Features of the academic background such as the grade level, academic stage and division. (3) Acts like as raising one's hand during class, using the assets available to them and showing gratitude for the school.

There are 175 female and 305 male participants in the sample. The students are from a variety of countries, Kuwait is home to 179 people, Jordan to 172, Palestine to 28, Iraq to 22 people, Lebanon to 17 people, Tunis to 12 people, Saudi Arabia to 11 people, Egypt to 7 people, Syria to 6 people, Libya, Iran, and the US are home to 4 people, Morocco to 1 person, and Venezuela to 22 people.

Two academic semesters are used to gather the dataset the 1st semester gathers 245 student records, whereas the 2nd semester gathers 235. Additionally, there is data on school attendance in the data set, with kids split into two categories. According to the number of absence days they have missed: 289 students have fewer absence days than seven and 191 students have more than seven. Table 1 shows the selected features.

Table 1: shows the selected features

Feature	Description
Gender	Students gender (Male or Female)
Place of birth	student's birth place (Lebanon, Kuwait, Saudi Arabia, Egypt, Jordan, USA, Iran, Venezuela, Tunis, Morocco, Tunis, Palestine, Syria, Lybia, Iraq,)
Nationality	student's nationality (Lebanon, Egypt, Kuwait, Jordan, Venezuela, USA, Saudi Arabia, Morocco, Syria, Iraq, Iran, Palestine, Lybia, Tunis,)
Grade Levels	Students grade ($G-01$, $G-02$, ……,$G-12$)
Educational Stages	educational level student (Middle School, High School, lower level)
Section ID	classroom student belongs (A, B, C)
Semester	school year semester (1^{st} , 2^{nd})
Topic	Topics (Spanish, English, IT, French, Arabic, Math, Biology, Chemistry, History, Geology, Science)
Hand raise	How many times in the classroom does the kid raise his or her hand (number:0-100)
Student Absence Days	the total number of days that a pupil has missed (under-7, above-7)
Resources visited	how often a student accesses a course's material (number:0-100)

B. Data preprocessing

Z score normalization is an important step in records preprocessing for predicting student overall performance. By making use of this approach, we standardize the distribution of records, ensuring that each feature contributes similarly to the prediction model. This normalization system involves calculating the mean and standard deviation of each function, then transforming the information to have a mean of 0 and a standard deviation of 1.

1) Z score normalization

The average and standard deviation of every feature across a set of training data serves as the basis for the normalization of all the features in the input data. Considering every characteristic, we calculate the mean and standard deviation. The equation w' which displays the normalized information for the input variable w_i, its average value μ_i and its standard deviation σ_i, reflects the equality used in the procedure.

$$w' = \frac{w_i - \mu_i}{\sigma_i} \tag{1}$$

Through the use of zero mean and one normal deviation, this method normalizes each characteristic in the dataset. Normalization is the first step in the process for the feature vectors in the dataset. The average and variance of every feature are computed using the training information and these values are retained as parameters in the final system design.

C. Feature extraction

The purpose of independent component analysis (ICA), a potent data analysis method, is to divide a multivariate signal into additive, independent components. ICA can extract useful characteristics from a variety of educational data, including test results, attendance records and study habits then apply them to the prediction of student success. ICA could uncover hidden patterns and linkages that influence a student's success or difficulties by finding underlying independent components within the data.

1) ICA

Using higher-order statistical parameters, the ICA is a multi-variant statistical technique for removing non-gaussian independent components (ICs) from the data.

$W = [W_1, W_2, \ldots.. W_m]^S$, an industrial multi-variate data collection with $W \epsilon\, \Re^{n\times m}$, is the sum of $l(\leq n)$unidentified ICs. The mathematical representation of the ICA model is as follows:

$$\mathrm{W} = \mathrm{BT} + \mathrm{E} \tag{2}$$

Where $B = [b_1 \ldots.b_l]^S \epsilon \Re^{n\times l}$ is a randomized mixing matrix with $T = [t_1, t_2 \ldots.t_m]^S \epsilon\, \Re^{l\times m}$. The resulting matrix is $E\epsilon\, \Re^{n\times m}$and the matrix containing ICs, where l denotes the ICs. The primary goal of ICA is to identify an independent matrix X assuming that the reconstruction matrix is provided by:

$$\hat{T} = XW \tag{3}$$

1. $Y = RW_d$is produced using $R = \Delta^{-1}A^S, Y$, where Y is the whitening matrices, Δ is the diagonal matrices and A is an eigenvalue matrix derived from the correlation of W_d. This process whitens the normalized data W_d. After the whitening step, the transformation is represented as $Y = RW_d = RBT = UT$

2. Iteration $(j = 1{,}2\ldots.n)$ is calculated as follows:

$$u_j = \arg\max\big(I(z)\big) \tag{4}$$

According to

$$u_j, F(zz^S) = J \tag{5}$$

$$I(Z) \approx [F\{H(z)\} - F\{H(a)\}]^2 \tag{6}$$

Where,

$$z = u_j^S y \tag{7}$$

For extracting the ICs, the negentropy approximation that maximizes non-gaussianity is the favored method. The negentropy function is represented by $I(Z)$ in Equation (6), the gaussian variables a have a zero mean and unit variance, then the non-gaussian function H is utilized to calculate the independent components. Following n iterations$U = [u_1 \ldots.u_n]\epsilon\Re^{n\times n}$ and the separation matrix $X = U^S R$ are found.

3. The best ICs are found using the cumulative percentage variance (CPV) approach, which is explained. The problem indications are built in the following manner for training data.

$$J_c^2 = W_d^S W_l^S X_l W_d \tag{8}$$

$$J_c^2 = W_d^S W_{n-l}^S X_{n-l} W_d \tag{9}$$

$$TOF = f.f^S \tag{10}$$

Where f is the remaining error with $f = W_d(j) - \widehat{W}_d$ with $\widehat{W}_d = R^{-1}U_l X_l W_d$, $X_l \epsilon \mathfrak{R}^{l\times n}$Provides the matrix with maintained ICs and $X_{n-l}\epsilon\mathfrak{R}^{(n-l)\times n}$provides the matrix with disregarded ICs.

4. Next, for the problem signals, thresholds $Sg_1, Sg_2, and\ Sg_3$are constructed.

D. *Feature selection*

Predictive modelling requires careful consideration of features, particularly when forecasting student achievement. Using the ASHO algorithm is one method that shows possibilities. ASHO effectively crosses the search space to find the most relevant elements for prediction, drawing inspiration from the movements of sea horses in the ocean. ASHO improves student performance prediction models' accuracy and efficiency by choosing the most useful attributes.

1) *Adaptive Sea Horse Optimization*

During the navigation phase, the initial SHO algorithm has certain flaws, particularly in its inability to achieve a harmonic equilibrium among global and local search behaviors. This problem originates from the search strategy's arbitrary number selection, which is based on an arbitrary number (i.e., a spiral or braided motion strategy). Furthermore, the optimization procedure could be hampered by the fixed values given to variables v and u, which determine the stem lengths. This could result in an inability of the algorithm to direct solutions to new locations. This work presents an enhanced version of SHO, called ASHO, to improve the algorithm's performance and resolve its primary weaknesses.

We examine the suggested ASHO approach in this part, which significantly alters the motivation behavior phase. As opposed to the conventional technique, the ASHO methodology incorporates a total of three discrete steps:

Neighborhood-based local search strategies, non-neighborhood-based global search strategies, and wandering-around-based search strategies are the three types of search strategies. A surroundings-based local search technique makes use of a person's neighborhood to improve the level of exploitation in that area. In particular, one neighbor, designated as d_{global}, is chosen at random from beyond the specified region but the lowest fitness function value, while another neighbor, designated asd_{local}, is chosen at random from inside the person's local area. Following that, the person modifies its location to match that of the d_{local}, as determined by equation (11), if the fitness value of thed_{local} is identified as being less than thed_{local}.

$$W_j(s+1) = W_j(s) + q_j \times ek_j(s) \times \left(n_{local}(s) - W_j(s)\right) \tag{11}$$

Where$n_{local}(s)$the concealing location of d is local for repetition$s$, q_j is an unknown number in an interval of $[0, 1]$, and $ek_j(s)$ is the person's flight duration in repetition s.

$$W_{ji}(s+1) = q_j \times ek_j(s) \times \left(n_{global\ i}(s) - W_{ji}(s)\right) \tag{12}$$

Where: $n_{global\ i}(s)$ the concealing location of $d$is global for repetition s and dimensions i, and i is an integer value.

A validation phase is included in the neighborhood-based localization method in addition to a non-neighborhood-based worldwide search method to confirm the new spot is in the specified range of the issue space. If not, the approach randomly modifies the elements that have strayed outside of this range to bring them back within the bounds of the issue space.

When the first two search strategies are unable to improve a person's health value, roaming around-based search approach is used. It functions by evaluating the surroundings and assisting the individual in adopting an approach that could be more beneficial while having a lower fitness rating. This new location is determined by equation (13), where $W_{ji}(s)$ refers to a randomly gathered person in the i^{th} dimension, and $n_{gbesti}(s)$ indicates the best concealing place among the whole population for category i.

$$W_{ji}(s+1) = n_{global\ i}(s) + q_j \times ek_j(s) \times \left(W_{qi}(s) - W_{ji}(s)\right) \tag{13}$$

Equation (13) is used to adjust the individual's position if their level of fitness at the fresh location is equal to or higher than the old one.

E. *XGBoost for predicting the student performance*

The XGBoost algorithm is based on the Gradient-Boosted Decision Trees (GBDT) structure. The XGBoost objective feature, unlike GBDT, has a phrase for regularization to prevent overfitting. Following is an equation (14) description of the primary objective function:

$$U=\sum_{j=1}^{n} S(y_j,\left(F(x_j)\right)+\sum_{h=1}^{t} G\,(f_h)+D \tag{14}$$

D is a constant that can be selectively eliminated, and $G(f_h)$ denotes the regularization term at iteration h. $G(f_h)$is a regularization term is written as equation (15):

$$G(f_h)=\alpha\mathrm{T}+\frac{1}{2}\eta\sum_{\mathrm{i}=1}^{\mathrm{H}} \mathrm{w}_{\mathrm{i}}^{2} \tag{16}$$

Where α is the leaf complexity, H stands for the leaf count, η is the consequence variable and $w_i$is the end product for each side node. In contrast to the leaf node, this represents a tree node that cannot be separated, leaves represent, based on classification criteria, the anticipated categories.

The main function can be expressed as follows equation (16) if mean square error (MSE) is the loss function:

$$U=\sum_{j=1}^{n}\left[p_j\omega_{p(y_j)}+\frac{1}{2}\left(q_j\omega_{q(y_j)}^{2}\right)\right]+\alpha T+\frac{1}{2}\eta\sum_{\mathrm{i}=1}^{\mathrm{T}}\omega_i^2 \tag{16}$$

Where g_j and T_j stand for 1st and 2nd derivatives of the loss function, respectively, $q(y_j)$is a procedure that changes data points into leaves.

The total of the ultimate loss values is determined by the loss values. The DT selections provide the leaf node's absolute loss value because they condense the leaf node loss values. Consequently, an equation could be used to represent the basic function. (17).

$$U=\sum_{i=1}^{H}[p_i\omega_i]+\frac{1}{2}(Q_{\,i}+\eta)\omega_i^2+\alpha T \tag{17}$$

Where $p_i=\sum_{j\in I_i}p_j$, $Q_{\,i}=\sum_{j\in I_i}q_j$, and $I_i$are the overall sample count in leaf node i.

The difficulty of maximizing the primary function is simplified to locating a quadratic function's minimum. Regularization phenomena have been added, giving XGBoost a stronger capacity to prevent overfitting.

IV. Experimental results

Python 3.10.1 is used to implement the suggested solution on a Windows 10 laptop equipped with an Intel i7 core CPU and 8GB of RAM. Use tools like TensorFlow/Keras or Scikit-Learn to train our recommended model using the training data. The proposed technique is adaptive Sea Horse Optimization- eXtreme Gradient Boosting (ASHO-XGBoost) compared to existing methods such as Genetic algorithm feature selection- Decision tree (GAFS-DT) [32], Genetic algorithm feature selection- Naive Bayes (GAFS-NB) [32], Genetic algorithm feature selection- Random forest (GAFS-RF) [32], Boruta feature selection- linear regression (BFS-LR) [33], Boruta feature selection- Support vector regressor (BFS-SVR) [33], Boruta feature selection- Random forest regressor (BFS-RFR) [33]. Accuracy, recall, precision, F measure, Root means square error (RMSE), and Mean Absolute Error (MAE), are used to evaluate these approaches' performance.

When it comes to predicting students' instructional accomplishments, accuracy is the degree of correctness. Based on several input variables, consisting of earlier grades, attendance information, demographic records and other pertinent characteristics, it assesses how correctly a predictive version can accurately discover or estimate a overall performance.

The accuracy performance is shown in Table 2 and Figure 2. The accuracy of the suggested ASHO-XGBoost system is 92.3%, which is higher than that of the GAFS-DT, GAFS-NB, and GAFS-RF, which are already in use and have accuracy rates of 74.58%, 75.42%, and 82.29%, respectively. As a result, compared to the existing methods for predicting student performance, the proposed approach is more accurate.

Table 2: values for accuracy, precision, recall, F measure

Method	Accuracy (%)	Precision (%)	Recall (%)	F measure (%)
GAFS-DT	74.58	75.39	75.15	75.26
GAFS-NB	75.42	76.26	76.25	76.25
GAFS-RF	82.29	82.7	82.81	82.75
ASHO-XGBoost [Proposed]	92.3	90.4	93.5	91.2

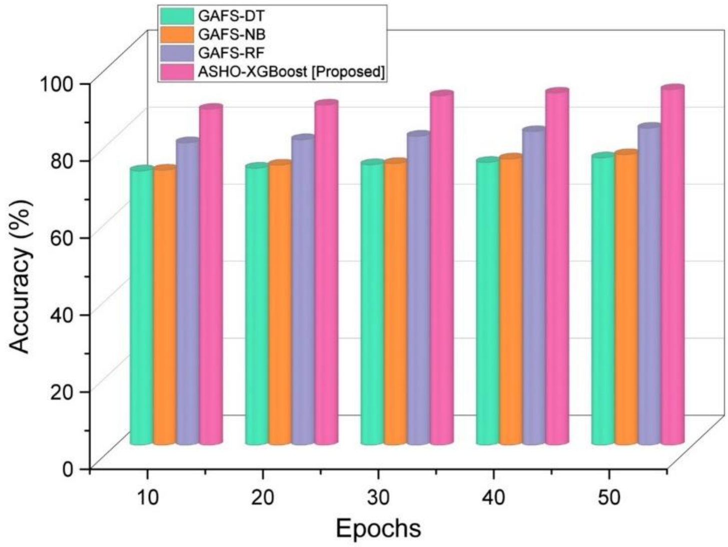

Figure 2: Accuracy performance

When discussing student performance, precision pertains to the degree of reliability and exactness with the model predicts a student's academic results. It calculates the percentage of all situations that the model correctly predicts as positive.

Table 2 and Figure 3 display the precision performance. The proposed ASHO-XGBoost system has an accuracy of 90.4%, higher than the existing GAFS-DT, GAFS-NB, and GAFS-RF systems, which have values of 75.39%, 76.26%, and 82.7%, respectively. Therefore, the suggested strategy for forecasting student performance is more precise than existing approaches.

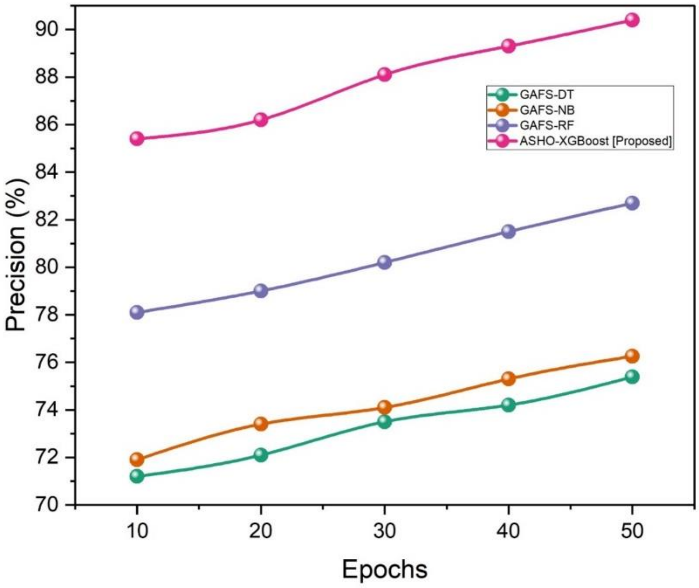

Figure 3: Precision performance

A predictive analytics approach known as "student performance prediction" makes use of data sources, consisting of demographics, educational history, and behavioral tendencies, to expect a character's potential educational fulfillment. This approach seems to earlier performance similarly to other pertinent variables like attendance and engagement to expect a student will do in next exams or guides.

Table 2 and Figure 4 present the recall performance. The recommended ASHO-XGBoost system has values of 93.5%, higher than the existing GAFS-DT, GAFS-NB, and GAFS-RF systems, which have 75.15%, 76.25%, and 82.81%, respectively. Therefore, the suggested methodology for forecasting student performance is better than the existing techniques.

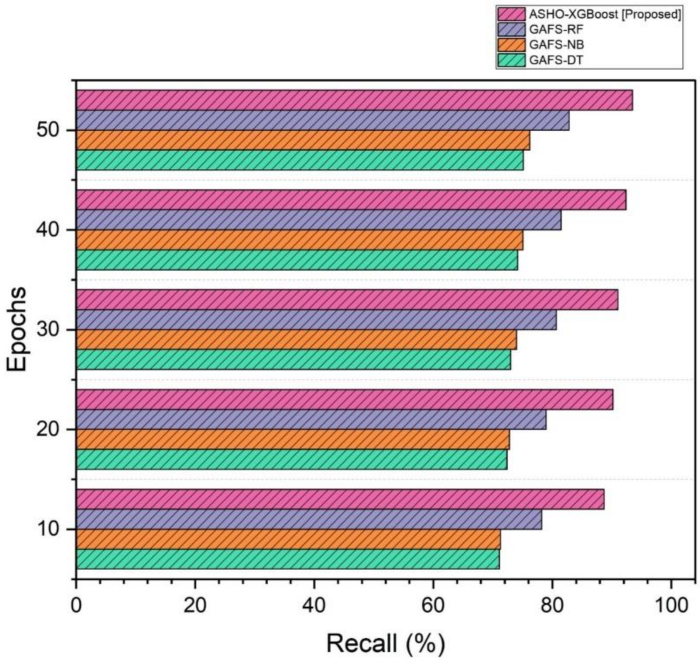

Figure 4: Recall performance

The statistic known as the "F-measure" is used when predicting pupil achievement. This metric generates a single cost by integrating recall and precision. The percent of expected effective instances amongst all forecasted positive effects is referred to precision, whereas the percentage of efficaciously anticipated high-quality cases amongst all actual advantageous instances is called as recall.

Table 2 and Figure 5 display the F measure performance. The proposed ASHO-XGBoost system has values of 91.2%, higher than the existing GAFS-DT, GAFS-NB, and GAFS-RF systems, which have values of 75.26%, 76.25%, and 82.75%, respectively. Therefore, the suggested strategy for forecasting student performance is more reliable than the existing approaches.

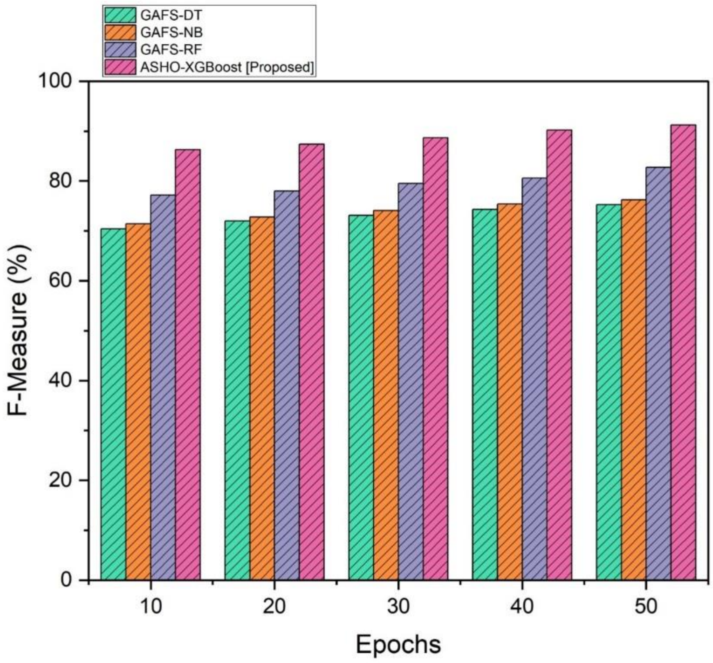

Figure 5: F measures performance

One statistic used in forecasting to evaluate a model's prediction accuracy is MAE. The average absolute difference between students' actual and anticipated performance scores is specifically measured in the context of educational outcome prediction.

Figure 6 and Table 3 display the MAE performance. Compared to the BFS-LR, BFS-SVR, and BFS-RFRwhich are existing and have values of 20.3, 16.68, and 13.52, respectivelythe recommended ASHO-XGBoost system has a lesser value of 11.55. This makes the suggested methodology better than the current techniques for forecasting student success.

Table 3: values for MAE, RMSE

Method	MAE	RMSE
BFS-LR	20.3	24.59
BFS-SVR	16.68	21.93
BFS-RFR	13.52	18.46
ASHO-XGBoost [Proposed]	11.55	17.25

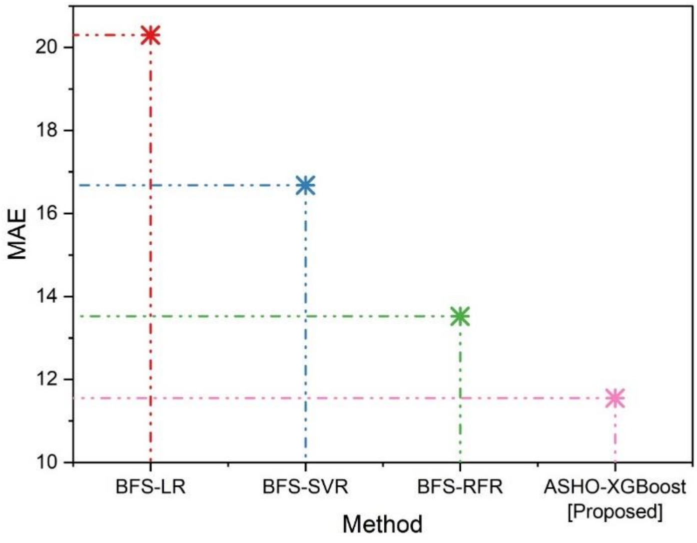

Figure 6: MAE performance

A metric called Root Mean Square Error (RMSE) is used to calculate how much a dataset's actual values depart from its anticipated values. It's very prevalent in domains like machine learning and statistics.

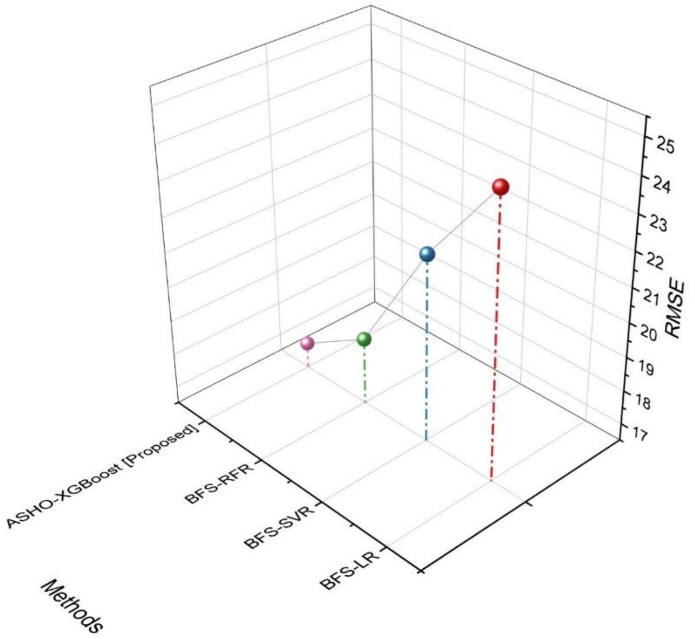

Figure 7: RMSE performance

Table 3 and Figure 7 present the RMSE performance. The recommended ASHO-XGBoost system has 17.25, lower than the existing BFS-LR, BFS-SVR, and BFS-RFR systems, which have values of 24.59, 21.93, and 18.46, respectively. Thus, the suggested methodology for forecasting student performance is better than the existing techniques.

V. Conclusion

The concept of student performance has several aspects and encompasses both cognitive growth and academic achievement. It evaluates student work, instructional strategies, and the effectiveness of educational institutions. Through the use of an innovative feature selection method, we desire to develop an intelligent data mining-based model in this study that will predict student performance. To choose the essential features to forecast the degree of education success, we provide a novel method called Adaptive Sea Horse Optimization (ASHO). To train our proposed model, we first collected a dataset called xAPI-Edu-Data, which contains a wealth of academic information about students. The Z-score normalization procedure is utilized to preprocess the acquired raw data, hence enhancing the data's quality. To extract pertinent characteristics from the data processed, we employed Independent Component Analysis (ICA). The eXtreme Gradient Boosting (XGBoost) technique is used to classify the chosen features to forecast student performance. The proposed feature selection method is compared to the existing method in terms of accuracy (92.3%), precision (90.4%), recall (93.5%), F measure (91.2%), MAE (11.55), RMSE (17.25). It can be difficult for the study to fully capture the intricacy of the factors that affect student success which could result in simplistic models. Deep learning and natural language processing are two innovative methods that improve prediction accuracy and discover hidden trends in student performance data.

Acknowledgment

This research received no external funding.

References

[1] Kumar, A.D., Selvam, R.P. and Palanisamy, V., 2021, March. Hybrid classification algorithms for predicting student performance. In 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS) (pp. 1074-1079). IEEE.

[2] Saa, A.A., Al-Emran, M. and Shaalan, K., 2020. Mining student information system records to predict students' academic performance. In The International Conference on Advanced Machine Learning Technologies and Applications (AMLTA2019) 4 (pp. 229-239). Springer International Publishing.

[3] Arun, D.K., Namratha, V., Ramyashree, B.V., Jain, Y.P. and Choudhury, A.R., 2021, January. Student academic performance prediction using educational data mining. In 2021 International Conference on Computer Communication and Informatics (ICCCI) (pp. 1-9). IEEE.

[4] Ashraf, M., Zaman, M. and Ahmed, M., 2020. An intelligent prediction system for educational data mining based on ensemble and filtering approaches. Procedia Computer Science, 167, pp.1471-1483.

[5] Hussain, S. and Khan, M.Q., 2023. Student-performulator: Predicting students' academic performance at secondary and intermediate level using machine learning. Annals of data science, 10(3), pp.637-655.

[6] Arcinas, M.M., 2022. Design of machine learning-based model to predict students' academic performance. ECS Transactions, 107(1), p.3207.

[7] Hung, H.C., Liu, I.F., Liang, C.T. and Su, Y.S., 2020. Applying educational data mining to explore students' learning patterns in the flipped learning approach for coding education. Symmetry, 12(2), p.213.

[8] Dien, T.T., Luu, S.H., Thanh-Hai, N. and Thai-Nghe, N., 2020. Deep learning with data transformation and factor analysis for student performance prediction. International Journal of Advanced Computer Science and Applications, 11(8).

[9] Neha, K., 2021. A study on prediction of student academic performance based on expert systems. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12(7), pp.1483-1488.

[10] Fayoumi, A.G. and Hajjar, A.F., 2020. Advanced learning analytics in academic education: Academic performance forecasting based on an artificial neural network. International Journal on Semantic Web and Information Systems (IJSWIS), 16(3), pp.70-87.

[11] Zafari, M., Sadeghi-Niaraki, A., Choi, S.M. and Esmaeily, A., 2021. A practical model for the evaluation of high school student performance based on machine learning. Applied Sciences, 11(23), p.11534.

[12] Abou Naaj, M., Mehdi, R., Mohamed, E.A. and Nachouki, M., 2023. Analysis of the factors affecting student performance using a neuro-fuzzy approach. Education Sciences, 13(3), p.313.

[13] Alam, A. and Mohanty, A., 2022, December. Predicting Students' Performance Employing Educational Data Mining Techniques, Machine Learning, and Learning Analytics. In International Conference on Communication, Networks and Computing (pp. 166-177). Cham: Springer Nature Switzerland.

[14] Pallathadka, H., Wenda, A., Ramirez-Asís, E., Asís-López, M., Flores-Albornoz, J. and Phasinam, K., 2023. Classification and prediction of student performance data using various machine learning algorithms. Materials today: proceedings, 80, pp.3782-3785.

[15] Yağcı, M., 2022. Educational data mining: prediction of students' academic performance using machine learning algorithms. Smart Learning Environments, 9(1), p.11.

[16] Hasan, R., Palaniappan, S., Mahmood, S., Abbas, A., Sarker, K.U. and Sattar, M.U., 2020. Predicting student performance in higher educational institutions using video learning analytics and data mining techniques. Applied Sciences, 10(11), p.3894.

[17] Mengash, H.A., 2020. Using data mining techniques to predict student performance to support decision-making in university admission systems. Ieee Access, 8, pp.55462-55470.

[18] Bhutto, E.S., Siddiqui, I.F., Arain, Q.A. and Anwar, M., 2020, February. Predicting students' academic performance through supervised machine learning. In 2020 International Conference on Information Science and Communication Technology (ICISCT) (pp. 1-6). IEEE.

[19] Tsiakmaki, M., Kostopoulos, G., Kotsiantis, S. and Ragos, O., 2020. Transfer learning from deep neural networks for predicting student performance. Applied Sciences, 10(6), p.2145.

[20] Bujang, S.D.A., Selamat, A., Ibrahim, R., Krejcar, O., Herrera-Viedma, E., Fujita, H. and Ghani, N.A.M., 2021. Multiclass prediction model for student grade prediction using machine learning. IEEE Access, 9, pp.95608-95621.

[21] Waheed, H., Hassan, S.U., Aljohani, N.R., Hardman, J., Alelyani, S. and Nawaz, R., 2020. Predicting the academic performance of students from VLE big data using deep learning models. Computers in Human behavior, 104, p.106189.

[22] Liu, D., Zhang, Y., Zhang, J.U.N., Li, Q., Zhang, C. and Yin, Y.U., 2020. Multiple features fusion attention mechanism enhanced deep knowledge tracing for student performance prediction. IEEE Access, 8, pp.194894-194903.

[23] Asselman, A., Khaldi, M. and Aammou, S., 2023. Enhancing the prediction of student performance based on the machine learning XGBoost algorithm. Interactive Learning Environments, 31(6), pp.3360-3379.

[24] Xu, Z., Yuan, H. and Liu, Q., 2020. Student performance prediction based on blended learning. IEEE Transactions on Education, 64(1), pp.66-73.

[25] Yousafzai, B.K., Khan, S.A., Rahman, T., Khan, I., Ullah, I., Ur Rehman, A., Baz, M., Hamam, H. and Cheikhrouhou, O., 2021. Student-performulator: student academic performance using hybrid deep neural network. Sustainability, 13(17), p.9775.

[26] Injadat, M., Moubayed, A., Nassif, A.B. and Shami, A., 2020. Multi-split optimized bagging ensemble model selection for multi-class educational data mining. Applied Intelligence, 50(12), pp.4506-4528.

[27] Ajibade, S.S.M., Dayupay, J., Ngo-Hoang, D.L., Oyebode, O.J. and Sasan, J.M., 2022. Utilization of ensemble techniques for prediction of the academic performance of students. Journal of Optoelectronics Laser, 41(6), pp.48-54.

[28] Nabil, A., Seyam, M. and Abou-Elfetouh, A., 2021. Prediction of students' academic performance based on courses' grades using deep neural networks. IEEE Access, 9, pp.140731-140746.

[29] Chui, K.T., Liu, R.W., Zhao, M. and De Pablos, P.O., 2020. Predicting students' performance with school and family tutoring using generative adversarial network-based deep support vector machine. IEEE Access, 8, pp.86745-86752.

[30] Alam, A., 2023, May. Improving Learning Outcomes through Predictive Analytics: Enhancing Teaching and Learning with Educational Data Mining. In 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 249-257). IEEE.

[31] https://www.kaggle.com/datasets/aljarah/xAPI-Edu-Data

[32] Farissi, A., Dahlan, H.M. and Samsuryadi, 2020. Genetic algorithm based feature selection for predicting student's academic performance. In Emerging Trends in Intelligent Computing and Informatics: Data Science, Intelligent Information Systems and Smart Computing 4 (pp. 110-117). Springer International Publishing.

[33] Syed Mustapha, S.M.F.D., 2023. Predictive analysis of students' learning performance using data mining techniques: A comparative study of feature selection methods. Applied System Innovation, 6(5), p.86.