Efficiency of Hybrid Decision Tree Algorithms in Evaluating the Academic Performance of Students

Yanxin Xie
Electrical and Information Engineering College, Jilin Agricultural Science and Technology College, Jilin 132101, Jilin, China

Abstract*—Educational institutions are anticipated to take substantial and proactive roles in guaranteeing students' successful program completion. Academic performance is conventionally employed to categorize and forecast students' future ability to confront post-graduation challenges. A student's academic accomplishments are instrumental in shaping exceptional individuals who may become future leaders. Using algorithms to assess and predict academic performance is a well-established practice in machine learning, encompassing techniques such as neural networks(NN), logistic regression(LR), decision trees(DT), and others. The goal of this project is to improve decision trees' ability to predict students' academic achievement via the use of data mining methods and meta-heuristic algorithms. Educational data mining involves the utilization of data analysis methodologies and tools to examine the extensive data generated within educational establishments as a result of students' interactions and activities throughout their academic journey. Pelican Optimization Algorithm (POA) and Runge Kutta optimization (RKO) are utilized algorithms in developing hybrid models, both of which can efficiently search for optimal or near-optimal splits by fine-tuning the hyperparameters of decision tree models. Students' final grades were predicted through training and testing models and categorized into four classes: Excellent, Good, Acceptable, and Poor. The classification capability of a single model and optimized counterparts was evaluated using Accuracy, Recall, Precision, and F1-score in separate phases for each category. Obtained results for all models revealed that POA and RKO developed Accuracy of DTC by 1.86% and 0.87%. Also, Precision and Recall metric analysis further manifest the superiority of DTPO. Prediction based on classifiers, especially workable optimized versions such as DTPO, paves the way for institutions to raise student success rates.

Keywords*—*Academic performance; decision tree; pelican optimization algorithm; runge kutta optimization

I. Introduction

Students' academic success is a fundamental educational objective, representing a key facet of any nation's educational goals. This focus on quality education as a catalyst for social change compels educational institutions to prioritize the nurturing of students who excel in academic and nonacademic assessments and acquire essential practical skills for competitiveness in the labor market. Education is at the heart of societal development, embodying collective aspirations for well-being and progress [1]. The quality of students produced by schools has thus become a prevailing concern. As highlighted by Kriegbaum et al. [2], academic achievement takes center stage, a barometer of intellectual education and a crucial prerequisite for individual and societal prosperity. In this context, Martín [3] emphasizes that academic performance extends beyond intellectual quotient (IQ), encompassing various dimensions to capture students' development's cognitive, psychomotor, and affective domains.

The primary benefit of data mining lies in its ability to thoroughly examine extensive data sets and derive rules that can capture the attention of relevant stakeholders. Furthermore, it can reveal previously undiscovered and valuable insights that greatly aid decision-making. Machine learning (ML) algorithms, specifically renowned for their effectiveness in classification tasks, are a central point of interest in numerous research endeavors [4], [5], [6]. According to Sharma, Himani, and Kumar [7], decision tree algorithms are widely recognized as effective tools for classification. Decision trees (DT) are structured models with root nodes, branches, and leaf nodes for predicting outcomes. These trees can handle numerical and categorical data, are easily understood, and are visually representable. They play a key role in identifying group characteristics and exploring relationships between variables and can be applied to predict student performance and other educational outcomes. Jorda and Raqueno [8] highlight various DT algorithms like C&R Tree, CHAID, C 5.0, and QUEST, which aid in developing classification systems.

II. Related Works Review

Numerous scholars have comprehensively investigated the multifaceted factors influencing student success across various academic levels. Several of these studies have utilized data mining (DM) techniques, particularly classification algorithms, to enhance the quality of higher education systems and predict student performance. In this section, a number of the related studies, especially those focusing on the application of the DT and classification in estimating the academic performance of the students, are presented [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31]. For instance, Qasem, Emad, and Mustafa [32] employed the CRISP framework to evaluate students' data in C++ courses, comparing classifiers such as $3, C4.5$ DT, and Naive Bayes (NB). The C4.5 DT outperformed other classifiers, shedding light on the attributes affecting student performance. Another study proposed a DT classification model to select suitable academic tracks for students, facilitating school management's decisions [33]. Nguyen and Peter [34] explored the efficiency of DT and Bayesian networks in predicting undergraduate and postgraduate student performance, revealing the superiority of DTs. Sunita and LOBO LMRJ [35] demonstrated the applicability of DM in education by using classification and clustering algorithms to predict student performance and group students. R. R. Kabra and Bichkar [36] developed classification models to identify at-risk students among first-year engineering students. S. Anupama and Vijayalakshmi [37] applied the C4.5 DT algorithm to predict MCA students' pass/fail outcomes, significantly improving results and efficiency compared to the $ID3$ algorithm. Bharadwaj and Pal [38] utilized the $ID3$ DT algorithm to predict student divisions based on various academic indicators. Surjeet and Pal [39] employed various DT algorithms to predict the performance of first-year engineering students, particularly in identifying those likely to fail. Dorina Kabakchieva [40] compared DM algorithms to predict student performance, classifying students as strong or weak, with the neural network achieving high accuracy for the strong class. Shovon and Mahfuza [41] proposed a hybrid approach combining clustering and classification to categorize students into high, medium, and low standards and make informed decisions about their academic performance, ultimately enhancing their final examination results. These studies collectively demonstrate the versatility of DM and classification algorithms in addressing various facets of student performance and academic success, aiding both educators and educational institutions in improving their educational processes and outcomes.

III. Objectives of the Current Work

As stated in the previous section, a few studies investigated the application of various decision tree classification algorithms in evaluating students' performance at different academic levels. There seems to be a significant gap in the literature related to exploiting the optimization capability of meta-heuristic algorithms in enhancing the evaluation performance of classification algorithms such as decision trees. Therefore, the current study employs Pelican and Runge Kutta optimization algorithms to develop hybrid decision tree models (DTPO and DTRK) for students' performance prediction. This innovative approach assists in detecting the optimization performance of presented algorithms in this field by comparing a single decision tree model with optimized versions using classification metrics such as Accuracy, Recall, Precision, and F1-score. In the following sections, the effect of selected input data on the outcome of models and a description of DTC and two optimizers will be presented. Results will be discussed using tables, bar charts, and confusion matrix for numerical and visual comparison between estimation models.

The study aims to enhance the predictive power of decision trees in predicting students' academic performance. This is achieved by introducing meta-heuristic algorithms, specifically the POA and RKO, to optimize decision tree models. The main research contribution lies in strategically employing these advanced algorithms to overcome limitations in conventional decision tree models. By fine-tuning hyperparameters and searching for optimal splits, the approach significantly improves the efficiency and accuracy of academic performance predictions. This contribution not only advances the understanding of ML applications in education but also provides a practical solution for educational institutions seeking to enhance student success rates.

Previous solutions for predicting students' academic success have faced persistent obstacles, necessitating the development of fresh approaches. Frequent deficiencies in previous endeavors encompass:

1) Restricted predictive precision: Numerous current models have had difficulties in attaining elevated accuracy while forecasting academic performance, frequently leading to misclassifications or imprecise categorizations of individuals.

2) Lack of adaptability: Traditional techniques may not possess the capacity to handle the dynamic nature of educational data effectively. Static models may not efficiently adapt to changes in learning settings, teaching approaches, or student behaviors.

3) Excessive dependence on traditional methods: Previous solutions may have mostly depended on conventional machine learning approaches without fully utilizing the capabilities of modern algorithms. This constraint can impede the capacity to capture complex patterns in educational data.

This study presents an innovative method that utilizes meta-heuristic algorithms, namely the POA and RKO. The application of these algorithms overcomes the constraints of conventional approaches by:

1) Improved model efficiency: The integration of meta-heuristic algorithms enhances the efficiency of the decision tree model, resulting in enhanced accuracy in forecasting students' academic achievement. The hybrid models created with the combination of POA and RKO exhibit an improved capacity to adjust to the changing characteristics of educational data, resulting in a more resilient and precise prediction mechanism.

2) Optimized hyperparameter tuning: By utilizing the techniques of POA and RKO, the hyperparameters of DT may fine-tune. This allows for a more detailed exploration of the feature space, resulting in better model performance compared to traditional decision tree models.

The overall organization of study in the next sections is as follows:

1) Dataset selection and preparation: This section outlines the process of selecting and preparing the dataset for analysis. It discusses the sources of data, data cleaning procedures, and any preprocessing steps applied to ensure the dataset's suitability for the study.

2) Decision tree and classification: Here, the focus is on explaining the decision tree algorithm and its application in classifying students' academic performance. It covers the fundamental principles of decision trees, including tree construction, node splitting criteria, and handling categorical and continuous variables.

3) Optimization algorithms: This section elaborates on the POA and RKO employed to enhance DT efficiency. It details the implementation of these algorithms to improve predictive accuracy.

4) Performance evaluation metrics: The performance evaluation metrics section discusses the metrics used to assess the effectiveness of the proposed approach. It provides insights into how accuracy, recall, precision, and F1-score are calculated and interpreted in the context of predicting students' academic performance.

5) Result: This section presents the results obtained from applying the proposed methodology to the dataset. It includes tables, graphs, or other visual aids to showcase the performance of the DT models with and without optimization algorithms.

6) Discussion: Here, the results are analyzed and interpreted in-depth. The discussion section delves into the implications of the findings and the limitations.

7) Conclusion: Finally, the conclusion section summarizes the key findings of the study and their implications for predicting students' academic performance. It reiterates the significance of the proposed approach and discusses its contributions to the field.

Fig. 1 shows the process of present study.

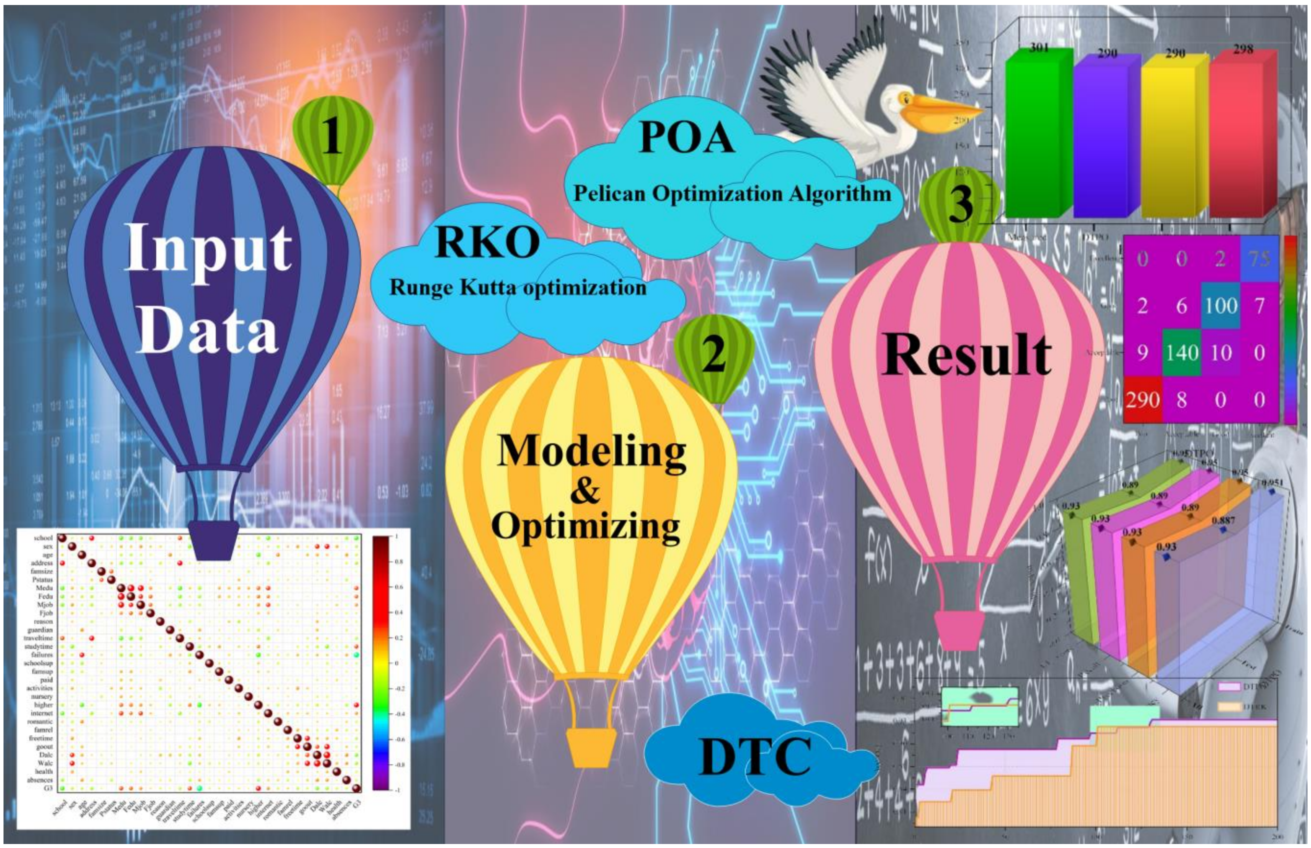

Fig. 1. Process of present study.

IV. Data Selection and Preparation

Data mining, also called database knowledge discovery, entails extracting valuable information from extensive datasets. This process employs various techniques to scrutinize vast data collections, uncovering concealed patterns and relationships that can potentially inform decision-making.

This study utilizes a comprehensive database of variables, which has been gathered from previous research articles. Included in the dataset are facts on the student's school, gender (male or female), age, domicile (rural or urban), family size (famsize), parental cohabitation status (Pstatus), and the mother's and father's educational background and employment (Medu, Fedu, Mjob, and Fjob). It also discusses the student's guardian (father, mother, or other), the reasons for choosing the school (such as proximity to home, school reputation, course preference, or others), the amount of previous class failures, participation in extracurricular activities and paid classes, attendance at nursery school, aspirations for higher education, availability of internet access at home, involvement in romantic relationships, the quality of family relationships (famrel), free time after school, socializing with friends (goout), weekday and weekend alcohol consumption (Dalc and Walc), current health status, and the A wealth of data for the research is provided by these input variables, which include a range of data kinds such as nominal, numeric, and binary. G3, which goes from 0, the lowest possible grade, to twenty, the best possible grade, indicates the final grades that pupils get according to the school. Students are placed into four different groups according to their G3 scores in order to further define and categorize the reported grades: G3 ranges for poor (range of 0– 12), acceptable (range of 12– 14), good (range of 14– 16), and excellent (range of 16– 20) are all available.

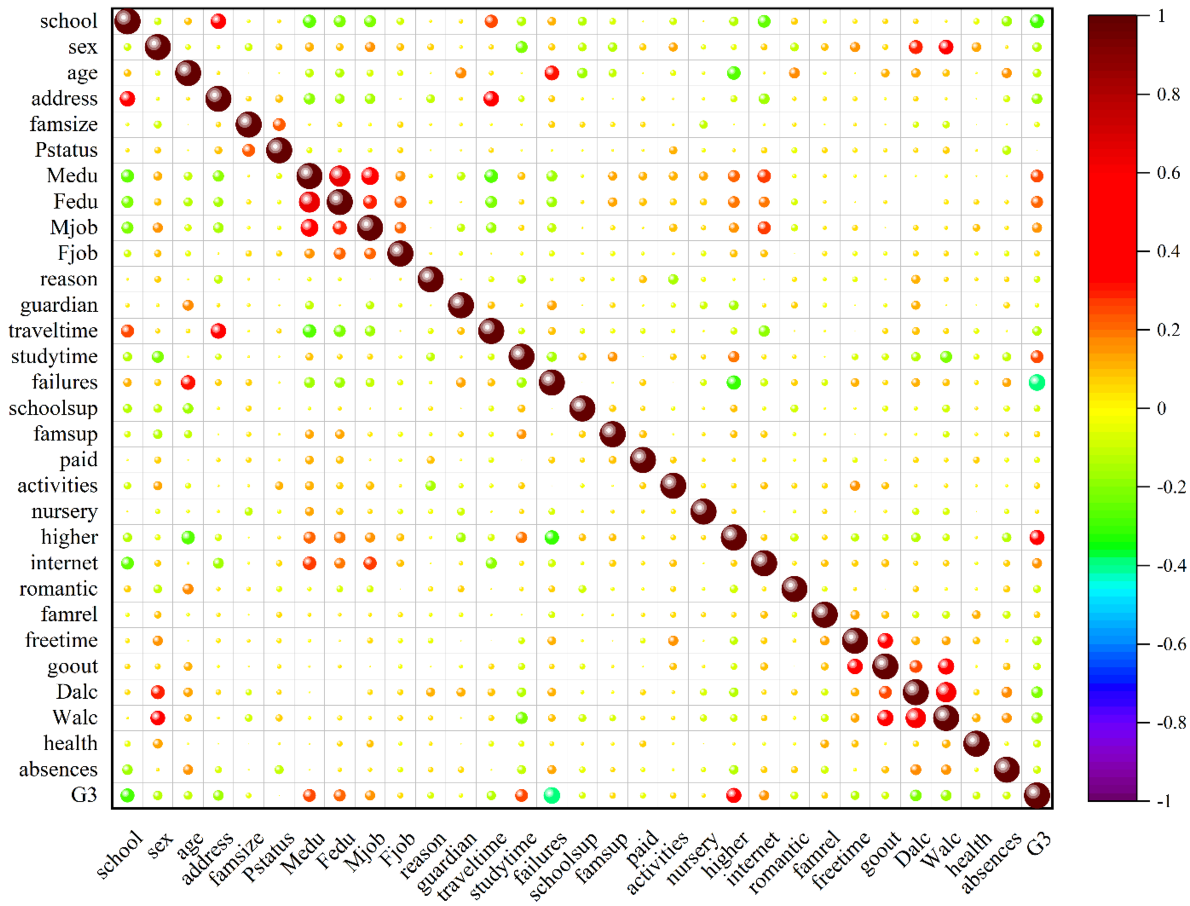

Fig. 2. Correlation matrix for the input and output variables.

In Fig. 2, the correlation matrix illustrates the relationships among all the input and output variables under examination. Parental education has the most substantial positive influence on a student's final grade, with the mother's education exerting a more pronounced effect than the father's. As anticipated, increased study time is positively associated with better academic outcomes, while the number of past failures by a student negatively impacts their grades. The presence of internet access and a student's aspiration for higher education both positively contribute to academic performance, while the adverse effects of alcohol consumption are evident. The most important variables affecting the quantity of absences from school were found to be daily and weekly alcohol intake, past failures, and student age.

V. Decision Tree and Classification

In a DT, which is shaped like a tree and looks like a flowchart, each internal node represents a test that is based on an attribute, each branch denotes the result of that test, and each leaf node also called a terminal node represents a particular class label. In order to use a DT for prediction, a route from the tree's root to a leaf node which holds the predicted class label for that data point is used to evaluate the attribute values of a particular data point ($tuple$). A benefit of decision trees is their ease of conversion into categorization rules. In DT learning, they function as predictive models that allow observations about an item to be translated into judgments about its desired value. Classification trees are a particular kind of these models that handle finite class values, and they are used in statistics, data mining, and machine learning. When compared to other categorization techniques, DT creation is often thought to be a speedy procedure [42].

The DT operates with three crucial parameters:

1) D (Data Partition): The first dataset, denoted by D, consists of training instances and the class labels that go with them.

2) Attribute list: In essence, this parameter is a set of properties that characterize the characteristics of the data.

3) Attribute selection method: The method for selecting the best suitable attribute to form branches or divisions in the decision tree is specified by this option. This usually entails using an attribute selection metric such as the Gini index or knowledge gain.

This is an explanation of the algorithm's operation:

- It starts by creating a node, which can be called "A".
- Should every example in the present dataset belong to the same class, "A" will be designated as a leaf node and assigned the common class label.

- Node "A" is once again designated as a leaf node and given the class that occurs most often in the data samples when the attribute list is empty.
- Next, the algorithm determines which characteristic will be used to divide the data into the cleanest subsets possible.
- This chosen property is given to Node "A" as the decision criteria.
- The selected attribute is eliminated from the list of characteristics if it is discrete.
- Based on the results of the chosen property, subsets of the data are created.
- A leaf node is connected to node "A" and labeled with the majority class of the original dataset if any of these subsets are empty.
- The procedure is repeated recursively for non-empty subsets, beginning with the creation of a new node, and the method continues until all data partitions have been handled.
- The method finally yields the decision tree structure that results.

This approach is a basic procedure for creating decision trees and is often used to data analysis and machine learning applications requiring data categorization and predictive modeling.

VI. Optimization Algorithms

A. Pelican Optimization Algorithm (POA)

Dehghani and Trojovský [43] introduced a novel metaheuristic optimization method, the POA, to address optimization issues based on swarm intelligence. The hunting habits of pelicans, who often hunt in groups and display a variety of clever tactics, served as the model for the algorithm. As an example, pelicans identify their prey's location ahead of time and move quickly to get close to it. They then hunt by swatting at a distance of 10 to 20 meters. The phases in the POA algorithm provide a comprehensive structure [44].

1) Initialization of the population: Usually, pelicans look for food in a specific area of the search habitat. As a result, each pelican inside this range is randomly assigned a starting location by the POA algorithm, which therefore initializes the population. A random number generator is used for this random initialization, which gives the pelicans their starting places.

$$P_i = S_{min} + rand.(S_{max} - S_{min}) \quad i = 1, \dots, I \tag{1}$$

Here, P_i signifies the original spatial location of the $i-th$ one in the populace. The maximum number of pelicans in the population is indicated by parameter I. The search space's exploration area's $lower$ and $upper$ bounds are defined, respectively, by the variables $Smin$ and $Smax$. Every pelican in the designated search area is given a random location based on a random integer generated by the function $rand\ \cdot()$.

2) Exploration phase: The pelicans' main goal at this point is to discover and pinpoint the location of their prey while simultaneously changing postures in anticipation of an assault. In order to do this, each pelican in the population has its geographic coordinates updated using Eq. (2):

$$P_{i,j}^1 = \begin{cases} P_{i,j} + rand.\left(P_{pj} - U.P_{i,j}\right), & f_p < f_i \\ P_{i,j} + rand.\left(P_{i,j} - P_{pj}\right), & else \end{cases} \tag{2}$$

Here, P_{ij}^1 represents the updated position of the $i-th$ pelican in the $j-th$ dimension, while P_{pj} represents the position of the prey. The prey's and the pelican's performance measures are shown by f_p and f_i, respectively. The update equation that is used to modify the pelican's location is determined by the parameter U, which is a random integer that may have a value of either 1 or 2.

3) Exploitation phase: The pelicans are ready to attack at this point as they have found their victim. The pelicans use acrobatic manoeuvres above the water as part of their predatory strategy to force the fish into their throat pouches. The following is a mathematical representation of this strategy:

$$P_{i,j}^2 = P_{i,j} + z.\left(1 - \frac{t}{T}\right).(2.rand - 1).P_{i,j} \tag{3}$$

Here, $P_{i,j}^2$ signifies the $j-th$ dimension's revised position for the $i-th$ one. The pelican's location is changed throughout exploitation by adjusting the parameter z, which is a random integer that might have a value of 0 or 2. T stands for both the current iteration of the algorithm and the maximum number of iterations.

The following pseudo-code contains the full statement of the POA algorithm:

```
Start POA
Input the optimization problem information.
Determine the POA population size (N) and the number of iterations (T
Initialization of the position of pelicans and calculation of the objecti
For t = 1:T
Generate the position of the prey at random.
For I = 1:N
Phase 1: Moving towards prey (exploration phase).
For j = 1:m
Calculate the new status of the jth dimension using Eq.(2).
End.
Update the ith population member.
Phase 2: Winging on the water surface (exploitation phase).
For j = 1:m.
Calculate the new status of the jth dimension using Eq.(3).
End.
Update the ith population member.
End.
Update best candidate solution.
End.
Output best candidate solution obtained by POA.
End POA.
```

B. Runge-Kutta Optimization (RUN)

Ahmadianfar et al. introduced the Runge-Kutta optimizer (RUN) [45], a population-based algorithm inspired by the

Runge-Kutta method for solving differential equations. RUN comprise two main stages: an initial search procedure influenced by Runge−Kutta principles and a subsequent phase called enhanced solution quality (ESQ) to improve solution quality. This study details the core principles supporting the RUN algorithm.

1) First stage: The algorithm of RUN utilizes a search mechanism (SM) that depends on the Runge-Kutta method to update the current solution's position in each iteration.

Algorithm 1: The RUN algorithm employs a SM to update the present solution's position.

$if\ rand\ <\ 0.5\ then$
$(exploration\ phase)$
$X_{\ n+1} = (X_c + r \times SF \times g \times x_c)\ + SF \times SM + \mu\ \times (randn\ \times\ (x_m - x_c))$
$else$
$(exploration\ phase)$
$X_{\ n+1} = (X_m + r \times SF \times g \times x_m) + SF \times SM + \mu\ \times (randn \times (x_{r1} - x_{r2}))$
$end\ if$

m indicates a random numerical value. g is allocated a stochastic value within the range of [0, 2]. r is a numeric value, which can take on either 1 or -1, which serves to increase the diversity within the range.

The determination of the adaptive factor SF involves calculations that include Eq. (4):

$$SF = 2 \times (0.5 - rand) \times f \tag{4}$$

$$F = a \times exp(-b \times rand \times (\frac{i}{Max_i})) \tag{5}$$

Max_i denotes the upper limit for the number of iterations.

x_c and x_m is computed through the utilization of Eq. (6) and Eq. (7):

$$x_c = \emptyset \times x_n + (1 - \emptyset) \times x_{r1} \tag{6}$$

$$x_m = \emptyset \times x_{best} + (1 - \emptyset) \times x_{lbest} \tag{7}$$

x_{best} represents the currently best available solution. x_{lbest} denotes the optimal position achieved during each iteration. f indicates a random value within the interval of (0, 1).

2) Second stage: To enhance solution quality and mitigate the risk of becoming stuck in local optima during each iteration, the RUN algorithm employs a technique referred to as Enhanced Solution Quality (ESQ).

Algorithm 2 delineates the steps involved in generating the solution (x_{new2}) through ESQ.

Algorithm 2: mathematical presentation of the second stage

if $rand < 0.5$ then
if $w < 1$ then
$$x_{new2}\ =\ x_{new1}\ +\ r.w.\left|(x_{new1}\ -\ x_{avg})\ +\ randn\right|$$
else
$$x_{new2}\ =\ (x_{new1}\ -\ x_{avg}) +\ r.w.\left|(u.x_{new1}\ -\ x_{avg})\ +\ randn\right|$$
end if
end if

$$w = rand(0,2).exp(-c(\frac{t}{Max_i})),\quad c = 5\ \times\ rand \tag{8}$$

$$x_{avg}\ = \frac{x_{r1} + x_{r2} + x_{r3}}{3} \tag{9}$$

$$x_{new1} = \beta \times x_{avg} + (1 - \beta) \times x_{best} \tag{10}$$

b denotes a number generated at random within the range of [0,1].

x_{best} denotes the best solution identified up to the current stage of exploration. r can take on any of the following values: 1.0 or -1. $rand$ indicates a parameter that is generated randomly.

The answer x_{new2} might not consistently exhibit better fitness when compared to existing solutions under these circumstances, the RUN algorithm presents an additional opportunity to boost fitness through utilization x_{new3}. Algorithm 3 outlines the procedure's sequential steps.

Algorithm 3: Improving the novel solution xnew3

if $rand\ <\ w$ then
$x_{new3} = (x_{new2} - rand x_{new2}) + SF.(rand.x_{RK} + (v.x_b - x_{new2}))$
end if

v is randomly generated and equals double the value of $rand$.

Algorithm 4 offers the pseudo-code for the main stages of the RUN optimization procedure.

Algorithm 4: Pseudo-Code of RUN Optimization

$Stage\ 1. Initialization$
$Start\ a, b$
$Create\ the\ RUN\ population\ X_n\ (n = 1,2,\ldots,N)$
$Compute\ the\ objective\ function\ of\ each\ member\ of\ the\ population$
$Determine\ the\ solutions\ x_w, x_b, and\ x_{best}$
$Stage\ 2. RUN\ operators$
$for\ i\ = 1:\ Maxi\ do$
$for\ n\ = 1:\ N\ do$
$for\ l\ = 1:\ D\ do$
$Updating\ solutions$
$Compute\ position\ x_{n+1,l}$
$end\ for$
$Enhance\ the\ solution\ quality$
$if\ rand\ < 0.5\ then$
$Compute\ position\ x_{new2}$
$if\ f(x_n) < f(x_{new2})\ then$
$if\ rand\ < w\ then$
$Compute\ position\ x_{new3}$
$end\ if$
$end\ if$
$end\ if$
$Update\ positions\ x_w\ and\ x_b$
$end\ for$
$Update\ positions\ x_{best}$
$i\ =\ i\ +\ 1$
$end\ for$
$Stage\ 3. return\ x_{best}$

VII. Performance Evaluation Metrics

In this section, the effectiveness of the proposed approach and models is systematically assessed by employing a set of well-established performance evaluation metrics. These metrics serve as quantitative measures to gauge the accuracy, precision, recall, and overall efficacy of the decision tree models, particularly those optimized using the POA and RKO.

The accuracy of the model is determined by dividing the number of accurately predicted occurrences by the total number of instances. It offers a broad summary of the model's performance in producing precise classifications for every class.

Precision measures the proportion of accurately anticipated positive occurrences to all expected positive instances, with an emphasis on the accuracy of positive forecasts. The model's capacity to reduce false positives is shown by a high accuracy score.

The capacity of the model to accurately identify all relevant occurrences is measured by recall, which is sometimes referred to as sensitivity or true positive rate. The ratio of accurately anticipated positive cases to the total number of actual positive instances is used to compute it.

The F1-score is the harmonic mean of precision and recall. It provides a balanced measure, especially in situations where there is an imbalance between the classes. A higher F1-score indicates a model that performs well in both precision and recall.

Statistical metrics for evaluating the classification capability of developed models are presented as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{11}$$

$$Precision = \frac{TP}{TP + FP} \tag{12}$$

$$Recall = TPR = \frac{TP}{P} = \frac{TP}{TP + FN} \tag{13}$$

$$F1\_score = \frac{2 \times Recall \times Precision}{Recall + Precision} \tag{14}$$

where,

- TP (True positives): the instances where the model's predictions were accurate.
- TN (True negatives): the instances that were correctly predicted.
- FP (False positives): the instances that were inaccurately forecasted
- FN (False negatives): the instances that were wrongly predicted.

VIII. Convergence Assessment

Throughout this investigation, we employed two metaheuristic optimization algorithms, namely the POA and RKO, to augment the DTC, leading to the development of hybrid models denominated as DTPO and DTRK. A thorough examination of the convergence dynamics of these optimized models was conducted, utilizing a convergence curve illustrated in Fig. 3.

The convergence curve, derived from Accuracy measurements spanning 200 iterations, serves as an instructive visual representation of the optimization procedure. Fig. 3 distinctly delineates the convergence trajectories of DTRK and DTPO. Noteworthy is the similarity in convergence rates exhibited by both models, particularly up to the midpoint of the optimization process.

At approximately the 115th iteration, DTRK and DTPO achieved a comparable peak Accuracy, both recording an impressive 0.92. However, as the iterations progressed, an intriguing divergence in their convergence patterns surfaced. DTRK showcased exceptional stability, sustaining its peak accuracy throughout the remaining iterations. Conversely, DTPO experienced a substantial surge in accuracy around the 130th iteration, ultimately surpassing DTRK in the final phases of the optimization process.

These subtle fluctuations in convergence behavior provide insights into the dynamic nature of the optimization algorithms and their influence on the performance of the hybrid models. The observed intricacies underscore the delicate interplay between the metaheuristic algorithms, DT optimization, and resulting accuracy levels. This nuanced analysis offers valuable perspectives on the strengths and limitations inherent in each approach throughout the iterative optimization process.

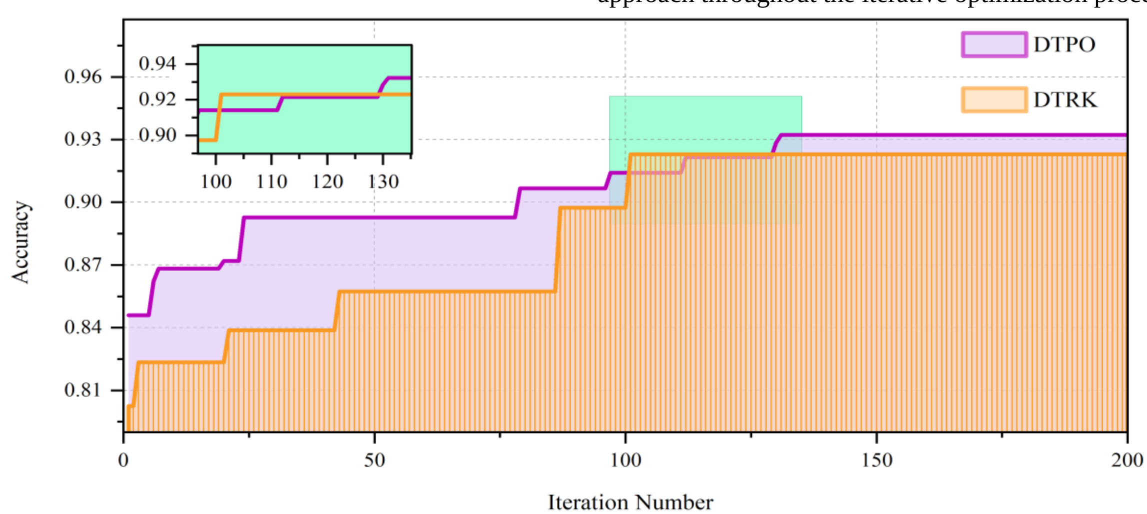

Fig. 3. Convergence of hybrid models.

IX. Results

With the purpose of predicting students' academic performance and methodically improving their future grades, this research presented three prediction models that use a categorization methodology. Table I shows the results of presented models. One of these models was a Decision Tree Classifier (DTC), while the other two were created by using Runge Kutta optimization (RKO) and the Pelican Optimization Algorithm (POA) to optimize the DTC. A portion of the dataset 70% for train and 30% for test the model was kept aside. For each model, training and testing stages provide metrics like accuracy, precision, recall, and F1-score; all results are shown in Fig. 4. Metric values were notably higher for all models during the train period than during the test phase. DTPO achieved the highest values across all metrics ($Accuracy = 0.932$, $Precision = 0.930$, $Recall = 0.930$, and $F1 - score = 0.930$), while DTC scored lower by approximately 1%. DTRK, in most cases, has metrics values slightly higher than the single model or the same values.

TABLE I. Result of Presented Models

Model	**Phase**	***Index values***			
		Accuracy	***Precision***	***Recall***	***F1 _score***
DTC	*Train*	0.915	0.910	0.910	0.910
	Test	0.892	0.890	0.890	0.890
	All	0.915	0.920	0.920	0.910
DTPO	*Train*	0.952	0.950	0.950	0.950
	Test	0.887	0.890	0.890	0.890
	All	0.932	0.930	0.930	0.930
DTRK	*Train*	0.923	0.920	0.920	0.920
	Test	0.903	0.900	0.900	0.900
	All	0.923	0.920	0.920	0.920

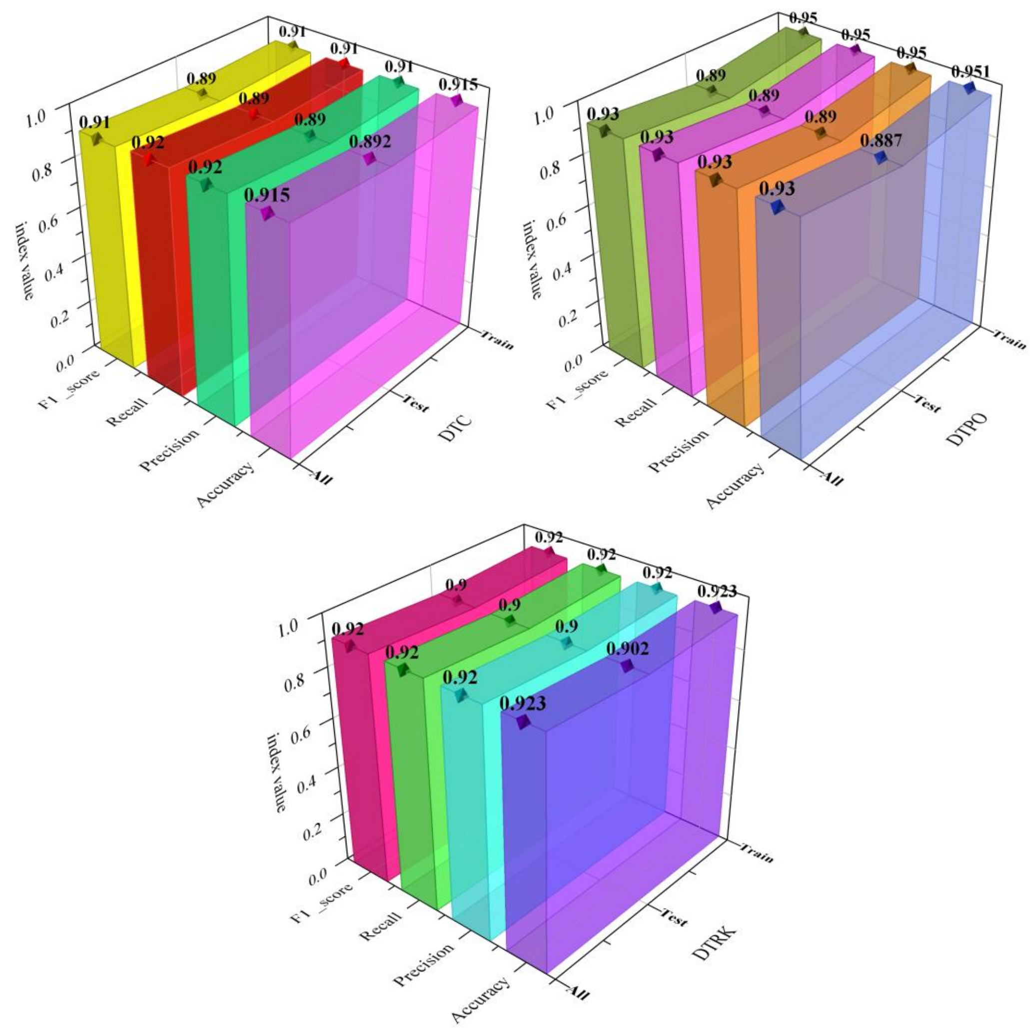

Fig. 4. Metrics performance of developed models.

After processing the data and evaluating the models' classification performance in both training and testing phases, the 649 pupils' $G3$ test scores were used to divide them into four groups: Poor (G3: 0-12), Acceptable (G3: 12-14), Good (G3: 14-16), and Excellent (G3: 16-20). The distribution revealed that the majority of students (46.38%) fell into the Poor category, with 23.73% in Acceptable, 17.26% in Good, and 12.63% in excellent categories.

To evaluate how well the created models performed in terms of categorization across various student groups, Table II shows the values for the Precision, Recall, and F1-score indices. In the analysis that follows:

- Comparing Precision values, in the excellent group, DTC and DTPO had an identical performance with 0.97, while DTRC was less precise with 0.95. In the Good and Poor groups, DTPO outperformed two other models, and finally, in the Acceptable group, the DTC single model performed superior to others. Considering all these results, it is not possible to introduce an absolute optimal model based on the Precision metric.
- Variation of the Recall metric was the same as Precision. All models performed better in the Poor category, with higher recall values of 0.96 for optimized versions and 0.99 for single models.
- The F1-score, a comprehensive metric, provides a nuanced basis for comparison. Higher F1-scores (nearest to 1) indicate superior model performance by balancing accurate identification of positive cases ($Precision$) and capturing all genuine positive cases ($Recall$). Across all student grades, DTPO demonstrated the highest accuracy with F1-scores of 0.94, 0.88, 0.89, and 0.97 for Excellent, Good, Acceptable, and Poor students, respectively. So, based on F1-score, DTPO came in the first ranking, followed by DTRK and DTC.

According to Fig. 5, there were really $301, 154, 112$, and 82 pupils in the Poor, Acceptable, Good, and Excellent categories. The frequency of students in each category based on classification models' outcomes is illustrated in the form of a bar chart for visual comparison. Comparing the two optimized models, they perform similar performance in Poor and Good classes, with 290 and 100 students correctly positioned in this group, but in two other groups, DTPO correctly classified three students higher than DTRK. The classification performance of the single model in the Poor and Good classes is better than hybrid models, especially in the Poor category. However, in Acceptable and Excellent groups, DTC succeeds inappropriately classifying a lower number of students than hybrid versions.

The confusion matrix presented in Fig. 6 clearly represents the accurate assignment of students to their respective grade categories and those who were misclassified. The numbers in diagonal raw represent the number of successfully organized models, and all numbers out of these squares are related to incorrect classification. For the DTPO model, 605 students were correctly categorized into Excellent, Good, Acceptable, and Poor classes, with only 44 misclassified. In the case of DTRK and DTC, 50 and 55 students were misclassified. The highest value of misclassification occurred in the case of DTC (20 students). Therefore, DTPO and DTC were the best and worst models for estimating students' academic performance.

TABLE II. PERFORMANCE EVALUATION INDICES FOR THE DEVELOPED MODELS BASED ON GRADES

Model	Grade	$Index\ values$		
		$Precision$	$Recall$	$F1-score$
DTC	$Excellent$	0.970	0.830	0.890
	$Good$	0.820	0.900	0.860
	$Acceptable$	0.950	0.820	0.880
	$Poor$	0.930	0.990	0.960
DTPO	$Excellent$	0.97	0.91	0.94
	$Good$	0.87	0.89	0.88
	$Acceptable$	0.88	0.91	0.89
	$Poor$	0.97	0.96	0.97
DTRK	$Excellent$	0.95	0.88	0.91
	$Good$	0.85	0.89	0.87
	$Acceptable$	0.91	0.89	0.9
	$Poor$	0.95	0.96	0.96

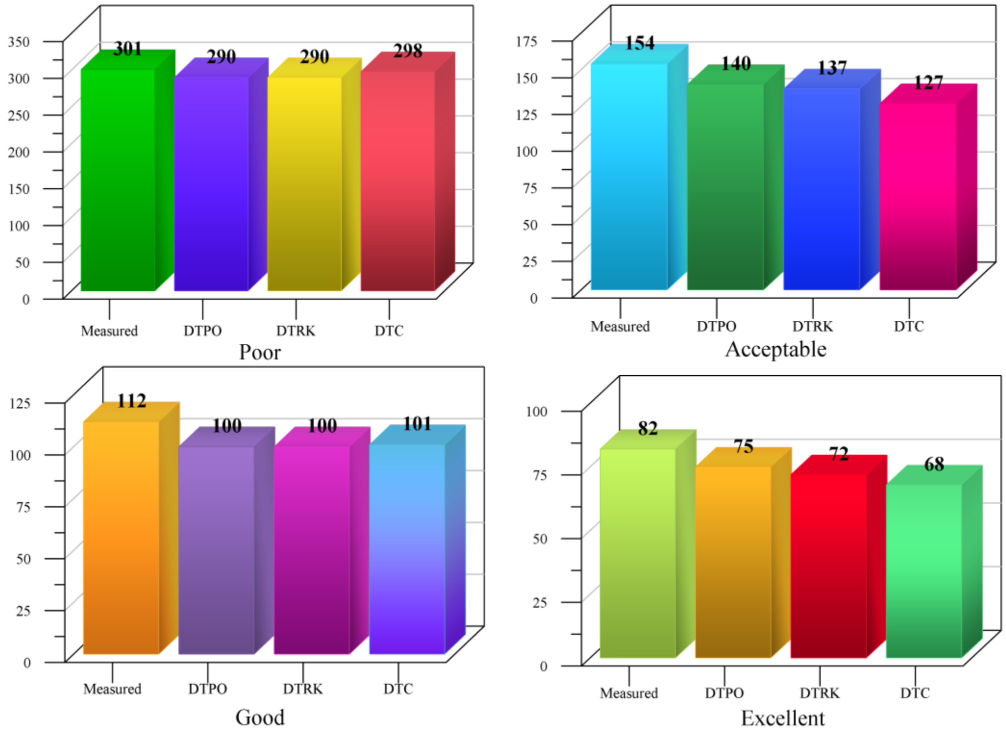

Fig. 5. Bar chart for the measured and estimated classification of students in four categories.

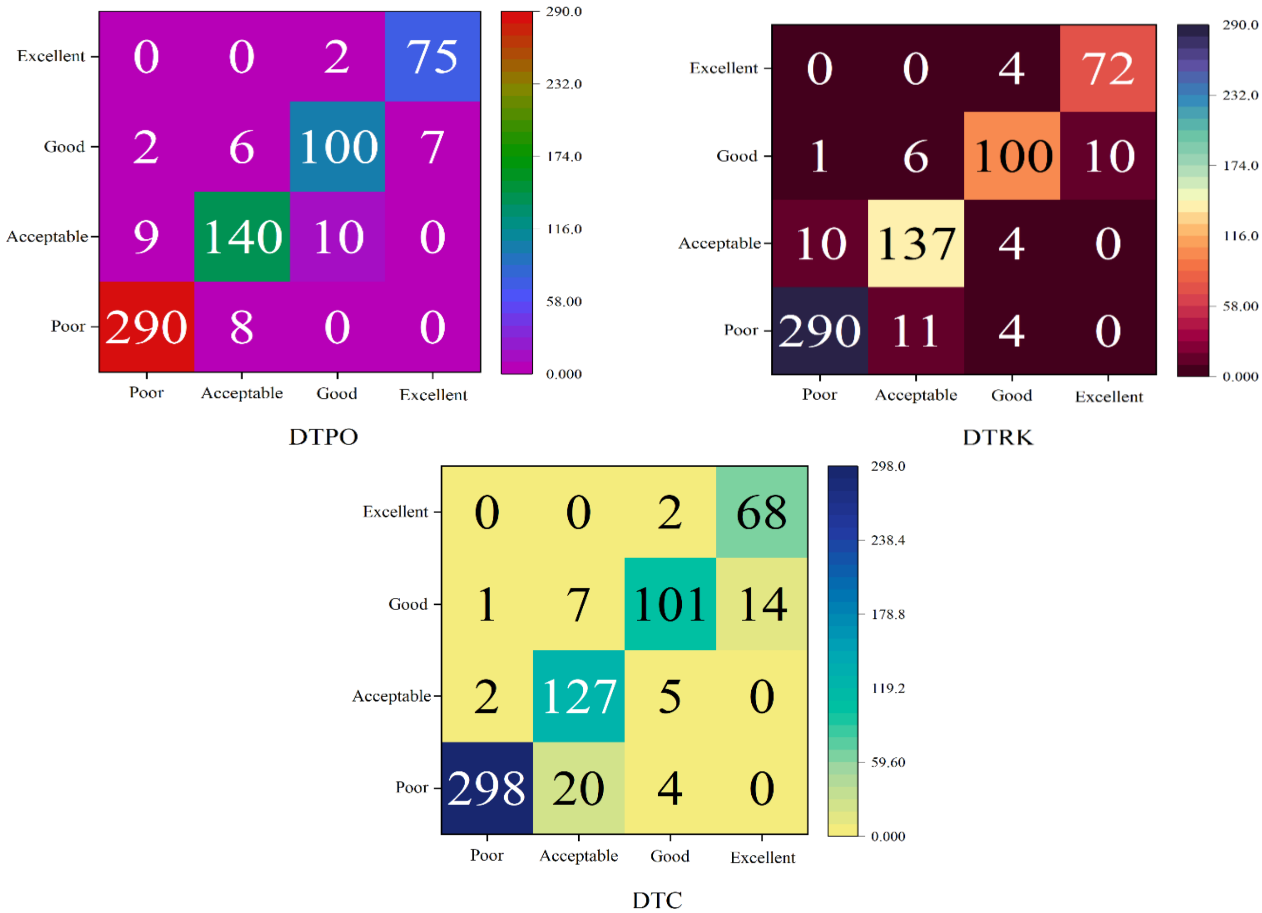

Fig. 6. Confusion matrix for each model's classification accuracy.

X. Discussion

A. Limitations

Ensuring data quality is vital for accurate predictive models, necessitating thorough preprocessing to address issues like missing values. Generalizing findings to diverse settings requires additional validation, emphasizing collaborations and diverse datasets. Meta-heuristic algorithm sensitivity underscores the importance of exploring robustness under varied conditions and conducting stability analyses. While these algorithms enhance accuracy, their reduced interpretability can be addressed through interpretable techniques, promoting transparency and trust in educational settings. Addressing data quality, generalizability, algorithm sensitivity, and interpretability collectively contributes to reliable, applicable, and transparent predictive models, facilitating improvements in student outcomes and educational practices.

B. Application of Study

The study's application in education encompasses the implementation of optimized decision tree models using meta-heuristic algorithms. These models facilitate early intervention for at-risk students, personalized learning plans, and strategic resource allocation. Insights from the study inform curriculum adaptation, student guidance, and institutional planning. The continuous improvement aspect involves refining models based on comparative analyses, while metrics evaluation ensures quality assurance. Overall, the study's practical implications extend to various facets of education, contributing to enhanced student success and institutional effectiveness.

XI. Conclusion

In pursuing academic excellence and improving education, this research underscores the pivotal role of data mining and classification algorithms, particularly decision tree models, in understanding and predicting student performance. It builds on a substantial body of related studies by introducing an innovative approach that leverages meta-heuristic optimization algorithms, specifically the Pelican and Runge Kutta optimizers (RKO and POA), to enhance the precision and accuracy of student performance models. The comprehensive evaluation employing key metrics like Accuracy, Precision, Recall, and F1-score highlights the potential of these meta-heuristic algorithms in optimizing classification outcomes. RKO enhanced the Accuracy and Precision of DTC by about 1 to 2 percent. POA performed weaker with the same Precision as DTC and lower than 1% enhancement in Accuracy. Furthermore, the categorization of 649 students based on their final grades revealed the superior performance of the DTPO in enhancing classification accuracy as it demonstrated a remarkable ability to correctly classify the majority of students (605 out of 649), while DTRK and DTC had more false classifications. This study not only added to the existing knowledge in the field but also provided valuable insights for educators and institutions striving to enhance educational processes and foster academic success, thus contributing to the broader goal of societal development and progress. Future studies should focus on improving the validity and real-world applicability of proposed predictive models for academic performance. This includes validating models across diverse educational institutions, assessing their long-term predictive power post-graduation, and conducting a thorough analysis of feature importance to guide targeted interventions. Exploring techniques for enhancing model interpretability without sacrificing accuracy is crucial for building trust among stakeholders. Additionally, comparative analyses with other advanced machine learning models in educational data mining can offer a comprehensive understanding of the proposed models' effectiveness. These recommendations aim to strengthen the reliability and practical utility of predictive models in predicting academic performance.

XII. Funding

Exploration of continuous improvement of experimental teaching under the mixed mode of "five in one"With practice Jilin Province higher education teaching reform research topic (2021) JLJY202168647754.

References

[1] L. M. Mphale and M. B. Mhlauli, "An Investigation on students' academic performance for junior secondary schools in Botswana," European Journal of Educational Research, vol. 3, no. 3, pp. 111–127, 2014.

[2] K. Kriegbaum, N. Becker, and B. Spinath, "The relative importance of intelligence and motivation as predictors of school achievement: A meta-analysis," Educ Res Rev, vol. 25, pp. 120–148, 2018.

[3] N. Martin Sanz, I. Rodrigo, C. Izquierdo GarcÃa, and P. Ajenjo Pastrana, "Exploring Academic Performance: Looking beyond Numerical Grades.," Universal Journal of Educational Research, vol. 5, no. 7, pp. 1105–1112, 2017.

[4] I. D. Mienye, Y. Sun, and Z. Wang, "Prediction performance of improved decision tree-based algorithms: a review," Procedia Manuf, vol. 35, pp. 698–703, 2019.

[5] F. Masoumi, S. Najjar-Ghabel, A. Safarzadeh, and B. Sadaghat, "Automatic calibration of the groundwater simulation model with high parameter dimensionality using sequential uncertainty fitting approach," Water Supply, vol. 20, no. 8, pp. 3487–3501, Dec. 2020, doi: 10.2166/ws.2020.241.

[6] Behnam Sedaghat, G. G. Tejani, and S. Kumar, "Predict the Maximum Dry Density of soil based on Individual and Hybrid Methods of Machine Learning," Advances in Engineering and Intelligence Systems, vol. 002, no. 03, 2023, doi: 10.22034/aeis.2023.414188.1129.

[7] H. Sharma and S. Kumar, "A survey on decision tree algorithms of classification in data mining," International Journal of Science and Research (IJSR), vol. 5, no. 4, pp. 2094–2097, 2016.

[8] E. R. Jorda and A. R. Raqueno, "Predictive model for the academic performance of the engineering students using CHAID and C 5.0 algorithm," Int. J. Eng. Res. Technol, pp. 917–928, 2019.

[9] R. Hasan, S. Palaniappan, A. R. A. Raziff, S. Mahmood, and K. U. Sarker, "Student academic performance prediction by using decision tree algorithm," in 2018 4th international conference on computer and information sciences (ICCOINS), IEEE, 2018, pp. 1–5.

[10] A. Hamoud, "Selection of best decision tree algorithm for prediction and classification of students' action," American International Journal of Research in Science, Technology, Engineering & Mathematics, vol. 16, no. 1, pp. 26–32, 2016.

[11] S. Sivakumar and R. Selvaraj, "Predictive modeling of students performance through the enhanced decision tree," in Advances in Electronics, Communication and Computing: ETAEERE-2016, Springer, 2018, pp. 21–36.

[12] A. K. Srivastava, A. Chaudhary, A. Gautam, D. P. Singh, and R. Khan, "Prediction of students performance using KNN and decision tree-a machine learning approach," Strad, vol. 7, no. 9, pp. 119–125, 2020.

[13] F. Chiheb, F. Boumahdi, H. Bouarfa, and D. Boukraa, "Predicting students' performance using decision trees: Case of an Algerian

University," in 2017 International Conference on Mathematics and Information Technology (ICMIT), IEEE, 2017, pp. 113–121.

[14] P. Guleria, N. Thakur, and M. Sood, "Predicting student performance using decision tree classifiers and information gain," in 2014 International conference on parallel, distributed and grid computing, IEEE, 2014, pp. 126–129.

[15] A. Hamoud, A. S. Hashim, and W. A. Awadh, "Predicting student performance in higher education institutions using decision tree analysis," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 5, pp. 26–31, 2018.

[16] G.-H. Wang, J. Zhang, and G.-S. Fu, "Predicting student behaviors and performance in online learning using decision tree," in 2018 seventh international conference of educational innovation through technology (EITT), IEEE, 2018, pp. 214–219.

[17] P. Cheewaprakobkit, "Predicting student academic achievement by using the decision tree and neural network techniques," Human Behavior, Development And Society, vol. 12, no. 2, pp. 34–43, 2015.

[18] A. B. Adeyemo and G. Kuye, "Mining students' academic performance using decision tree algorithms," Journal of Information Technology Impact, vol. 6, no. 3, pp. 161–170, 2006.

[19] P. Strecht, J. Mendes-Moreira, and C. Soares, "Merging Decision Trees: a case study in predicting student performance," in Advanced Data Mining and Applications: 10th International Conference, ADMA 2014, Guilin, China, December 19-21, 2014. Proceedings 10, Springer, 2014, pp. 535–548.

[20] L. D. Yulianto, A. Triayudi, and I. D. Sholihati, "Implementation Educational Data Mining For Analysis of Student Performance Prediction with Comparison of K-Nearest Neighbor Data Mining Method and Decision Tree C4. 5: Implementation Educational Data Mining For Analysis of Student Performance Prediction with Comparison of K-Nearest Neighbor Data Mining Method and Decision Tree C4. 5," Jurnal Mantik, vol. 4, no. 1, pp. 441–451, 2020.

[21] V. Matzavela and E. Alepis, "Decision tree learning through a predictive model for student academic performance in intelligent m-learning environments," Computers and Education: Artificial Intelligence, vol. 2, p. 100035, 2021.

[22] S. Wiyono, T. Abidin, D. S. Wibowo, M. F. Hidayatullah, and D. Dairoh, "Comparative study of machine learning knn, svm, and decision tree algorithm to predict students performance," International Journal of Research-Granthaalayah, vol. 7, no. 1, pp. 190–196, 2019.

[23] S. Wiyono, D. S. Wibowo, M. F. Hidayatullah, and D. Dairoh, "Comparative study of KNN, SVM and decision tree algorithm for student's performance prediction," (IJCSAM) International Journal of Computing Science and Applied Mathematics, vol. 6, no. 2, pp. 50–53, 2020.

[24] T. M. Ogwoka, W. Cheruiyot, and G. Okeyo, "A model for predicting students' academic performance using a hybrid of K-means and decision tree algorithms," International Journal of Computer Applications Technology and Research, vol. 4, no. 9, pp. 693–697, 2015.

[25] D. K. Kolo and S. A. Adepoju, "A decision tree approach for predicting students academic performance," 2015.

[26] M. Pandey and V. K. Sharma, "A decision tree algorithm pertaining to the student performance analysis and prediction," Int J Comput Appl, vol. 61, no. 13, pp. 1–5, 2013.

[27] Y. S. Alsalman, N. K. A. Halemah, E. S. AlNagi, and W. Salameh, "Using decision tree and artificial neural network to predict students academic performance," in 2019 10th international conference on information and communication systems (ICICS), IEEE, 2019, pp. 104–109.

[28] M. Apolinar-Gotardo, "Using decision tree algorithm to predict student performance," Indian J Sci Technol, vol. 12, p. 5, 2019.

[29] A. B. Raut and M. A. A. Nichat, "Students performance prediction using decision tree," International Journal of Computational Intelligence Research, vol. 13, no. 7, pp. 1735–1741, 2017.

[30] M. H. I. Shovon and M. Haque, "An Approach of Improving Students Academic Performance by using k means clustering algorithm and Decision tree," arXiv preprint arXiv:1211.6340, 2012.

[31] S. A. Kumar, "Efficiency of decision trees in predicting student's academic performance," 2011.

[32] Q. A. Al-Radaideh, E. M. Al-Shawakfa, and M. I. Al-Najjar, "Mining student data using decision trees," in International Arab Conference on Information Technology (ACIT'2006), Yarmouk University, Jordan, 2006.

[33] Q. A. Al-Radaideh, A. Al Ananbeh, and E. Al-Shawakfa, "A classification model for predicting the suitable study track for school students," Int. J. Res. Rev. Appl. Sci, vol. 8, no. 2, pp. 247–252, 2011.

[34] N. T. Nghe, P. Janecek, and P. Haddawy, "A comparative analysis of techniques for predicting academic performance," in 2007 37th annual frontiers in education conference-global engineering: knowledge without borders, opportunities without passports, IEEE, 2007, pp. T2G-7.

[35] S. B. Aher and L. Lobo, "Data mining in educational system using weka," in International conference on emerging technology trends (ICETT), 2011, pp. 20–25.

[36] R. R. Kabra and R. S. Bichkar, "Performance prediction of engineering students using decision trees," Int J Comput Appl, vol. 36, no. 11, pp. 8–12, 2011.

[37] S. A. Kumar, "Efficiency of decision trees in predicting student's academic performance," 2011.

[38] B. K. Baradwaj and S. Pal, "Mining educational data to analyze students' performance," arXiv preprint arXiv:1201.3417, 2012.

[39] S. K. Yadav and S. Pal, "Data mining: A prediction for performance improvement of engineering students using classification," arXiv preprint arXiv:1203.3832, 2012.

[40] D. Kabakchieva, "Student performance prediction by using data mining classification algorithms," International journal of computer science and management research, vol. 1, no. 4, pp. 686–690, 2012.

[41] M. H. I. Shovon and M. Haque, "An Approach of Improving Students Academic Performance by using k means clustering algorithm and Decision tree," arXiv preprint arXiv:1211.6340, 2012.

[42] H. Sharma and S. Kumar, "A survey on decision tree algorithms of classification in data mining," International Journal of Science and Research (IJSR), vol. 5, no. 4, pp. 2094–2097, 2016.

[43] P. Trojovský and M. Dehghani, "Pelican optimization algorithm: A novel nature-inspired algorithm for engineering applications," Sensors, vol. 22, no. 3, p. 855, 2022.

[44] X. Mei, Z. Cui, Q. Sheng, J. Zhou, and C. Li, "Application of the improved POA-RF model in predicting the strength and energy absorption property of a novel aseismic rubber-concrete material," Materials, vol. 16, no. 3, p. 1286, 2023.

[45] I. Ahmadianfar, A. A. Heidari, A. H. Gandomi, X. Chu, and H. Chen, "RUN beyond the metaphor: An efficient optimization algorithm based on Runge Kutta method," Expert Syst Appl, vol. 181, p. 115079, 2021.