Optimizing Student Performance Prediction Using Binary Waterwheel Plant Algorithm for Feature Selection and Machine Learning

Faris H. Rizk[1], Mahmoud Elshabrawy[2], Basant Sameh[2], Karim Mohamed[2], Ahmed Mohamed Zaki[1]

[1]Computer Science and Intelligent Systems Research Center, Blacksburg 24060, Virginia, USA
[2] Department of Communications and Electronics, Delta Higher Institute of Engineering and Technology, Mansoura, 35111, Egypt

Emails: faris.rizk@jcsis.org; CH1900052@dhiet.edu.eg; CH1900072@dhiet.edu.eg; CH1900193@dhiet.edu.eg; azaki@jcsis.org

Abstract
This paper deals with a pivotal part of educational data analytics, aiming to increase the accuracy and interpretability of student performance prediction models. The cornerstone of our method is the innovative application of binary waterwheel plant algorithm bWWPA in the feature selection. As we can see, an essential part of any model is the predicted values, which correctly define all the characteristics of this model. Practically, we begin with solid data pre-processing, which incorporates data cleaning and missing values, duplicate removal, and data transformation in order to get model input as optimally as possible. Preceding the application of bWWPA, we employ an ensemble of regression machine learning models. Set up a baseline for predictive capability, getting initial outcomes with an average Mean Squared Error (MSE) of 0.064. The following feature selection phase proceeds, showing the algorithm. Ability to recognize important elements and, as a result, improve model effectiveness and explain power. The comparative analyses after feature selection point to refined gains in the model, and the performance is reporting a lower MSE of 0.032 with the refined models. These findings, methodologically, add to student performance prediction. Accordingly, it emphasizes the decisive status of feature selection in improving models. The paper's significance extends to teachers, institutions, and researchers, giving insights into more precise and relevant student success-supporting interventions.

Keywords: feature selection; student performance prediction; Optimization; educational data analysis; regression models; Waterwheel Plant Optimization Algorithm

1. Introduction

In recent years, the educational landscape has witnessed a remarkable evolution into an approach where data is used as one of the tools to improve student learning experiences. One crucial feature of this paradigm change lies in the creation and enhancement of predictive models that are able to predict students' performance effectively. The ability to predict academic success helps teachers provide proper and necessary interventions and leads to favourable educational outcomes. With regard to student performance prediction, the application of sophisticated computational models is now obligatory. These models, mostly based on machine learning and predictive analytics, solve the complex patterns present in the educational data and predict future academic issues or successes. However, despite their popularity, existing models are faced with inherent limitations that impair performance. Overcoming these constraints is paramount for designing better and more valid prediction mechanisms and, ultimately, for creating a favourable learning context.

The models for predicting student performance in the present paradigm are plagued by their intricate nature and limitations. Predictive accuracies of these models can be limited by the plethora of factors affecting academic

DOI: https://doi.org/10.54216/JAIM.070102
Received: April 27, 2023 Revised: August 11, 2023 Accepted: January 01, 2024

success. An emphasis on nuanced, complex techniques that can manage the intricacies is required, which implies sophisticated procedures are needed to transcend the current barriers of predictive models. This paper addresses this gap by suggesting an integration of feature selection and model transformation techniques. The primary objective of this research is to enhance the accuracy and interpretability of student performance prediction models through the integration of two crucial components: Feature selection and transformation. By overcoming the shortcomings of existing approaches, we open the door to a more profound comprehension of the intricate mechanisms driving student outcomes.

To guide our investigation, we pose the following research questions:

- What is the impact of the integration of feature selection techniques on the predictive performance of student performance models?
- How does model transformation influence the interpretability and predictive accuracy of regression models in question?
- What is the contrast between these integrated approaches and traditional models with respect to predictive performance, interpretability and computational efficiency?

Furthermore, the key technical Contributions of this work can be summarized as follows:

- Data Preprocessing Phases: Use of thorough data preprocessing methods for quality assurance and integrity guarantee of the "Students Performance" dataset. We are managing missing data, removing duplicates, transforming numerical features, encoding categorical data, dealing with outliers, and performing necessary feature engineering [1].

- Model Testing Using Evaluation Metrics: Hard validation of selected regression models employing a wide range set of evaluation metrics. Model performance is evaluated using metrics like mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and the like to appreciate the predictive attributes [2].

  Feature Selection: Combining feature selection methods (bWWPA, bPSO, bWAO, bGWO and bFA) for selecting and retaining the most influential features for the student performance. Model interpretability and dimensionality reduction are enhanced to improve computational efficiency [3].

- Feature Selection Statistics: Statistical analysis application of evaluation of each feature selection method to obtain the results. Quantitative assessment of the effect of feature selection on model performance and feature significance unveiling.

- ANOVA Analysis: We are implementing an ANOVA analysis to ignore the variance in performance among the different feature selection techniques. Statistical investigation of the differences in predictive abilities provides a wholesome comprehension of the performance of each method.

- Wilcoxon Signed Rank Test: Implementation of the Wilcoxon signed rank test to compare the actual performance with alternative models. Statistically sound testing to assess the importance of found discrepancies in predictive accuracy and interpretability.

- Transformed Model Performance: An introduction to transformation techniques for improved interpretability and accuracy of selected regression models. Evaluation of the efficiency of model transformation in overcoming restrictions and increasing the general performance [4].

The next section of this paper examines a detailed literature review with a description of existing models and methodologies (2), which is followed by an introduction to our methodology (3). As for our experiments, we present their results, namely, the model performance assessments, as well as the statistics in Section 4. The paper ends with a recap of findings, implications for educational institutions and directions for future studies (Section 5).

2. Literature Review

Predicting students' end-of-term results has become a more important effort in the field of education because people want to improve outcomes and support tailored interventions. In this domain, Machine Learning (ML) techniques have become very useful because they can examine vast datasets and uncover patterns that help predictive models. The literature on leveraging machine learning techniques for forecasting academic achievements is examined in this related works section. Numerous methodologies, datasets, and applications are

DOI: https://doi.org/10.54216/JAIM.070102
Received: April 27, 2023 Revised: August 11, 2023 Accepted: January 01, 2024

looked into in order to highlight the advancements made in predicting end-of-term performances and to illuminate the possible implications, challenges, and outcomes for educational practices.

The SET is a common tool used to evaluate teaching in higher learning institutions [5]. SET has an ambivalent position because it influences the jobs of individuals and influences the quality of teaching through summative and formative evaluation, respectively. A mixed method is used to untangle the coveted aspects that determine university students' willingness to give feedback using SET as an integral part of the teaching evaluation process. The results show a strong relation between commitments to feedback, positiveness of evaluative perceptions, and usefulness beliefs. Nevertheless, motivation to interact may only be sufficient if students think biases are in the picture. Students need more awareness of and a deep appreciation for the consequences of the answers to SET is an illuminating finding. [6] uses insights from earlier research and introduces two novel studies to explore the complicated link between optimistic attributional style and academic performance. According to research on the associations between optimistic attributional style, improved wellbeing, and decreased depression, positive events with stable global causes exhibit a stronger correlation with academic success than negative events with locality and instability. Academic level and test type also affect these associations. Two more studies show that an optimistic attributional style for positive events is a strong predictor of academic success across a wide range of student groups, and the benefits last for a long time. The paper ends with a discussion of potential explanations for the observed moderator effects as well as recommendations for additional research and practical implications.

In [7], the difficult dynamics of analyzing student success in the classroom are looked at. This is an area of educational research that is still being looked into. The idea is all about how important it is to find out what factors affect student success in order to make good predictive models. Even though the majority of the previous literature focuses on student-centered outcomes, this research emphasizes the significance of great teaching as a potent trigger in determining academic outcomes. A measurable indicator of change compared to predicted performance is suggested as a way to look at how teaching affects performance growth in theoretical courses. The dataset has almost 0.2 million academic records from a well-known university in India. Association mining techniques are used on it. The findings show that there is a positive correlation between better teaching and better performance. This means that better teaching encourages greater student success by giving students more confidence in their ability to get expected or better grades. A thorough investigation into the two-way longitudinal relationships between success goals and academic performance was conducted in the setting of Chinese college students by [8]. The Achievement Goal Questionnaire was used to get information from 311 Chinese college students in their first and third years of study. Their end-of-term grade point averages were used as measures of academic performance. We use continuous design and structural equation models to show that goals for performance approach are linked to better academic performance. In comparison, goals for performance avoidance are linked to worse academic performance. Academic performance, in comparison, did not exhibit a significant correlation with mastery goals. The implications of these findings are carefully examined and discussed.

Different types of learners are common, so [9] looks into the complicated patterns of online learning behavior in the setting of the 5y online learning platform. Leveraging data from 2,205 learners and their related final exam scores, the research uses factor analysis on 13 measurement indicators to look at the correlation between online learning behavior and academic performance. After that, the multiple linear regression model is used to learn more about this link. The findings demonstrate a strong correlation between learners' final exam scores and factors pertaining to basic and comprehensive questions. As a result, teachers and students who use the 5y platform should concentrate on knowledge point tests and unit tests as a way to enhance the teaching and learning experience while still staying within the allotted time. [10] discusses the complicated landscape of student performance modeling in this area, which is a fascinating and challenging research domain called education data mining (EDM). The effects of multiple factors on success are dynamic, and there are numerous education datasets available, especially for online learning. This increases research interest in the subject. There are many EDM polls in the literature, but only a few are made to look at and predict how well students will do. With a focus on predictors, identification methods, and the temporal and goal-related parts of prediction, this paper fills this gap by offering the first systematic review of EDM studies focusing on student success in classroom learning. According to a thorough study of 140 studies, the course has amazing prediction efficiency. Predicting when the course will begin should receive more attention, however.

The paper [11] talks about the challenges that students face when they start higher education, which are marked by noticeable declines in performance and motivation, especially in education. Although it is clear that low-cost, broad help is required at this critical juncture, there are still questions about the optimal timing and intervention methods. Short-term motivation is a crucial element that influences success in higher education, as demonstrated by grade goals and self-efficacy. The way students' domain knowledge and interests change over time can also tell us about their general academic experience. In a first-year online mathematics course at a university in Asia that performs significant research, this study looks at the complex interplay between performance, short-term motivation, and long-term motivation. The people who took part were given different tests, and studies like

DOI: https://doi.org/10.54216/JAIM.070102
Received: April 27, 2023 Revised: August 11, 2023 Accepted: January 01, 2024

MANOVAs and a longitudinal, fully forward latent SEM model helped to show how the complex processes worked. Though there are still some unknowns, self-efficacy is a good indicator of how well someone will do. However, at first, students are not interested, but by the end of the semester, they are again. In the middle of the course, setting goals for the whole thing boosts self-efficacy. By the end of the course, it might have hurt interest. The findings have both theoretical and practical implications that can be used to improve methods for helping students. [12] provides an account of strength-based parenting (SBP) – a peculiar parenting style that reflects a strong sensitivity of the parent to the uniquely interesting abilities of his or her child. My SBP model has been proven to be better than other parenting styles in predicting various measures of adolescents' wellbeing. Based on the study of wellbeing and learning and their interconnectedness, the proposition that children brought up by a strength-based guardian are liable to complete better in their scholastics is considered. At the initial phase of the investigation, through a questionnaire, the perception of parenting style, engagement, and perseverance of 741 students from an Australian public secondary school (Mage = 13.70, SD = 1.33; 50 % female) are determined. The subsequent data that were collected three months later to assess academic results imply that SBP is associated with better wellbeing, especially in regard to adolescent engagement and resilience. Persistence, working as a moderator, has an additional effect that stems from SBP performance. The model of the study implies that teenagers whose parents emphasize their strengths gain better scores by demonstrating more resilience. These results also reveal certain parenting practices, such as highlighting the concept of SBP with emphasis on the importance of parent-student relationships and dispositional characteristics, including engagement and perseverance, in effecting positive outcomes on positive education.

The paper [13] goes into detail about how interest and self-efficacy, two important factors that affect motivation and learning that tend to drop during adolescence, work together. With the help of the newly developed Situated Expectancy-Value Theory (SEVT) by Eccles and Wigfield, the research carefully studies the interplay of competence beliefs (like self-efficacy) and value perceptions (like interest) in determining students' success in learning situations. Data from 754 German secondary school students (MAge = 13.56; SD = 1.2; 49.4% girls) are analyzed using two latent change models and a latent neighbor change model, which contain covariates like gender, age, and grades. Self-directed learning times are part of the school year at the schools that were studied, which is worth mentioning. At these times, students can follow the goals set by the curriculum more closely. Instead, they choose what they want to learn based on what they are interested in. The findings are intriguing because they demonstrate a rise in interest and self-efficacy all year long, not just during self-directed learning times. This suggests that an instructional environment with self-directed learning intervals fosters a synergistic relationship between interest and self-efficacy, demonstrating the advantageous effects of such pedagogical strategies on the developmental trajectory of students' interest and motivation components.

Using Machine Learning techniques to guess students' final grades is becoming more and more important, as shown in the literature review in this related works section. According to the research we looked at, predictive analytics can completely alter the way education is carried out, from the in-depth understanding of individual learning paths to the early identification of risk factors. As more schools use data-driven strategies to improve student outcomes, the challenges and moral concerns that come with using these types of predictive models become more apparent. However, the advancements made in this domain make it possible for students to make informed decisions, receive individualized learning experiences, and receive early education. All of these studies show not only what is happening now but also how ML techniques will continue to play a big role in changing education and student success in the years to come.

3. Proposed Methodology

The proposed Methodology is basically an elaborately designed structure that is meant to improve the accuracy and explainability of the student performance prediction models in the setting of the "Students Performance" dataset. This part depicts each phase of the Methodology in-depth and comprehensively, explaining logic, techniques, and factors that play a crucial role in the design of an efficient predictive model.

3. Dataset Description and Preprocessing

Recognizing the pivotal role of high-quality data in predictive modelling, the preprocessing stages are extensive pursuant to the preparation of the "Students Performance" dataset for effective model training and evaluation.

- Handling Missing Data: Data from the dataset is meticulously investigated to identify and deal with the existing missing values using elaborate imputation techniques for columns "Total Salary if Available" and "Transportation to the University." This, in turn, gives a complete and more detailed dataset that can be used for the following analysis.

DOI: https://doi.org/10.54216/JAIM.070102
Received: April 27, 2023 Revised: August 11, 2023 Accepted: January 01, 2024

- Dealing with Duplicate Data: A rigorous check for duplicate entries is performed with duplicates systematically removed to maintain the quality of the dataset, such that redundancy will be avoided during the information process.

- Data Transformation: Numerical features that are categorized as 'Student Age' and 'Total Salary if Available' undergo standardization and normalization. Also, Gaussian distribution is encouraged by using log transformations for skewed data, which will lessen the effect of outliers.

- Encoding Categorical Data: Categorical variables such as "Sex" and "Graduated-High-School-Type" are appropriately encoded to enable the model training process to be directionally effective, allowing these variables to contribute meaningfully to the learning process.

- Handling Outliers: Methods that are resilient and resistant to outliers are applied so that these data points with undue influence are not able to abnormally bias the learning process, thus leading to a robust model.

- Feature Engineering: The feature extraction step follows a complicated engineering process that includes selecting and removing the irrelevant features. The objective of this step is to strengthen the prediction power of the dataset by adding striking attributes.

- Data Splitting: The dataset is purposely divided into training and testing partitions, using a stratified method to ensure a balanced representation of the outcomes. This enables rigorous model evaluation while maintaining the data's statistical properties.

- Handling Imbalanced Data: Techniques are used to sort out operational issues around feature categories such as "Scholarship Type" and "Regular Artistic or Sports Activity" to make sure that there are no classes that have any advantage.

- Text Data Preprocessing: For textual data, preprocessing involves tokenization, removal of stop words and stemming/lemmatization. Re: These steps are of the utmost importance to working with and deriving useful knowledge from the textual characteristics.

- Time Series Data: For those features that concern time, careful consideration is given, and relevant resampling is performed to keep the temporal coherence, taking into account and dealing with the temporal dependence within the data as well.

- Normalization of Images: Image data being part of the dataset, scaling pixel values, and data augmentation are absolutely important issues if one is speeding up with image features.

Figure 1 exposes the joint distribution between the study hours per week and the Cumulative Grade Point Average (CumGPA), hence, providing a view of how the two factors work hand in hand in influencing student's performance. This scatter plot is a source of clarification on the patterns displayed by the data, explaining how different study hours per week relate to the cumulative GPA, thus revealing some factors that determine academic success.

DOI: https://doi.org/10.54216/JAIM.070102
Received: April 27, 2023 Revised: August 11, 2023 Accepted: January 01, 2024

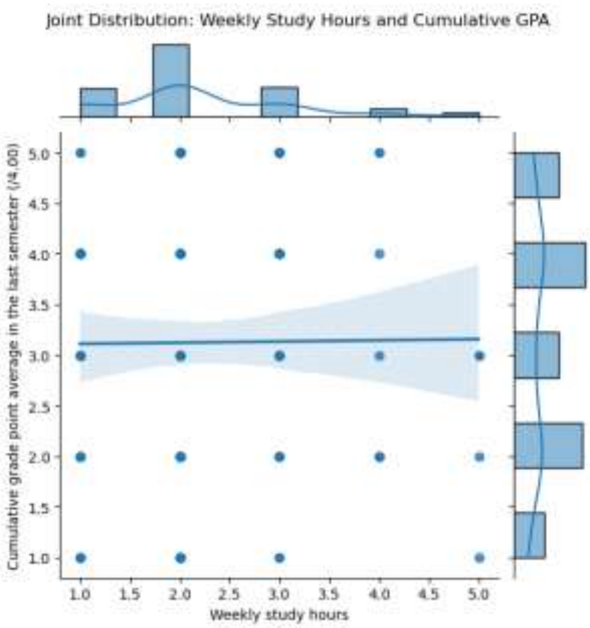

Figure 1: Joint Distribution for weekly Study Hours and Cumulative GPA

Figure 2 clarifies how the students' GPA relate to their mother's education level. This insightful visualization is aimed at finding out possible connection between mother’s education level and child’s academic performance. This figure, through plotting the data, also presents a nuanced explanation of how a mother's educational background may be correlated with a student's overall GPA and thus adds to the idea of other general socio-economic factors having an influence on student performance.

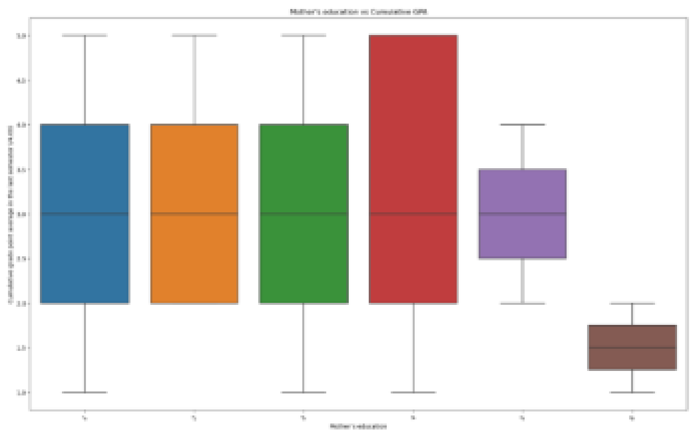

Figure 2: Mother's Education vs Cumulative GPA

In Figure 3, we show the correlation matrix based on the dataset of student performance. This visualization directs with the interconnections between variables presented in the dataset. Each cell in the matrix discloses the value of strength and direction of the correlations; hence, the matrix serves as a complete review of what factors could be affecting students' performances.

DOI: https://doi.org/10.54216/JAIM.070102
Received: April 27, 2023 Revised: August 11, 2023 Accepted: January 01, 2024

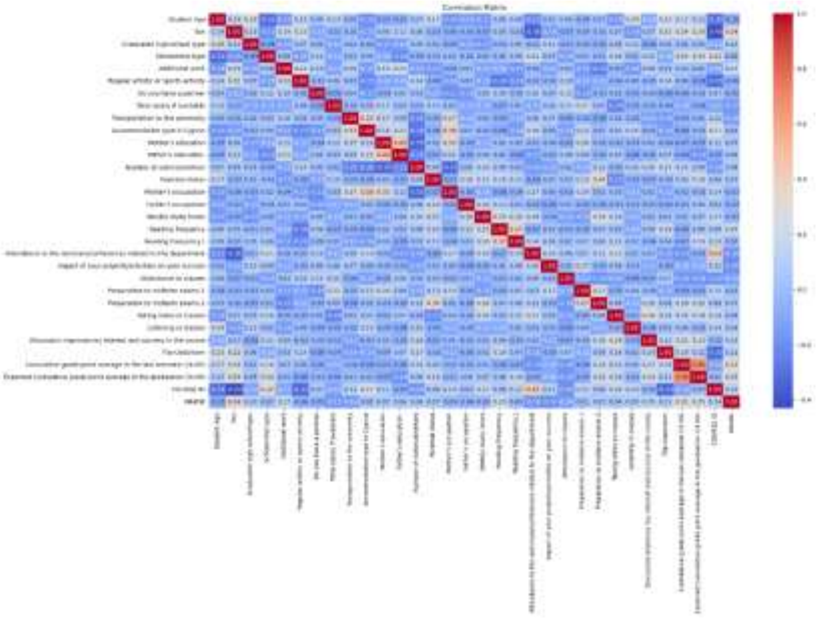

Figure 3: Correlation Matrix of "Student Performance" Dataset

This meticulous preprocessing phase ensures that the dataset is refined, representative, and poised for subsequent model development.

3.2 Model Selection

Model selection is a critical component of the Methodology | We seek variety in regression models in order to account for the complex relationships within the "Students Performance" dataset.

- **GradientBoostingRegressor:** Thanks to the ensemble learning capability, Gradient Boosting is selected for its strength in capturing complex relationships within the data [14]. The iterative refinement of weak learners of the model explains its power to capture subtle patterns. The prediction is given by:
- $\hat{y}_i = \phi(x_i) = \sum_{k=1}^{K} f_k(x_i)$, where f_k is the k-th weak learner.

- **XGBoost:** The XGBoost has a great reputation for handling sparse data and for deploying strong regularization techniques; it provides the robustness of the ensemble [15]. This is because it has good accuracy and scalability. The final prediction is given by:
- $\hat{y}_i = \phi(x_i) = \sum_{k=1}^{K} f_k(x_i)$, where f_k is the k-th weak learner.

- **DecisionTreeRegressor:** Decision Trees, selected due to their interpretability, are good at capturing nonlinear relationships manifested in the data [16]. Tree structure gives glimpses into the feature's importance on predictions.

- **RandomForestRegressor:** The Random Forest model as an ensemble model avoids overfitting and hence it ensures good generalization of the dataset [17]. Combining predictions from several decision trees increases predictive accuracy and stability.

- **MLPRegressor:** MLPs, one type of artificial neural network, are taken into account thanks to their capability to learn complex patterns, mostly among different datasets [18]. The capacity of the network to capture nonlinear relationships for improved prediction is harnessed. The prediction is given by:
- $\hat{y}_i = \phi(x_i) = f^{(L)}(f^{(L-1)}(\ldots f^{(1)}(x_i; \theta^{(1)}) \ldots ; \theta^{(L-1)}); \theta^{(L)})$, where $f^{(l)}$ is the activation function in layer l and $\theta^{(l)}$ are the weights.

DOI: https://doi.org/10.54216/JAIM.070102
Received: April 27, 2023 Revised: August 11, 2023 Accepted: January 01, 2024

- **ExtraTreesRegressor:** Like Random Forest, Extra Trees models introduce more randomness, which promotes diversity in the ensemble [19]. This randomness improves the model's generalization ability to unseen data.

- **Linear regression:** A baseline linear model is provided as a reference to evaluate the performance of more complex models in comparison with a simpler, linear solution [20]. Linear regression is thus very useful for judging the predictive power of more sophisticated models. The prediction is given by:
- $\hat{y}_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \cdots + b_n x_{in}$, where $b_0, b_1, \ldots, b_n$ are the coefficients.

- **CatBoost:** Selected for its capability to manage categorical features well, CatBoost increases the ensemble's diversity [21]. Embedded in it is support for categorical variables, which streamline the preprocessing steps and improve model performance. The prediction is given by:
- $\hat{y}_i = \phi(x_i) = \sum_{k=1}^{K} f_k(x_i)$, where f_k is the k-th weak learner.

- SVR (Support Vector Regressor): Support Vector Regressors, which perform well at capturing complex relationships, are used [22]. Through feature transformation to a high-dimensional space, SVR shines in capturing intricate patterns. The prediction is given by:
- $\hat{y}_i = \sum_{j=1}^{n} (\alpha_j - \alpha_j')K(x_i, x_j) + b$, where K is the kernel function, α and α' are the Lagrange multipliers, and b is a bias term.

- **KNeighborsRegressor:** The simplicity and adaptability make K-Nearest Neighbors a welcome addition to the variety of models [23]. The approach is based on nearby interactions, thus increasing its potential for representing local structures. The prediction is given by:
- $\hat{y}_i = \frac{1}{k}\sum_{j=1}^{k} y_{N_j}$, where N_j is the j-th nearest neighbor.

This painstaking selection guarantees that the models, in total, are able to explore the range of relationships found in the dataset, thus supplying an all-around portrait of its internal workings.

3.3 Performance Metrics

The choice and usage of evaluation metrics is of critical importance when evaluating models in our research. The metrics determine the predictive power of the model. Table 1 shows that the metrics used were varied to include several dimensions related to model accuracy, correctness, and generalization.

Table 1: Metrics for reporting result of regression.

Metric	Formula
RMSE	$\sqrt{\frac{1}{N}\sum_{n=1}^{N}(\hat{V}_n - V_n)^2}$
RRMSE	$\frac{RMSE}{\sum_{n=1}^{N}\hat{V}_n} \times 100$
MAE	$\frac{1}{N}\sum_{n=1}^{N}\lvert\hat{V}_n - V_n\rvert$
MBE	$\frac{1}{N}\sum_{n=1}^{N}(\hat{V}_n - V_n)$
NSE	$1 - \frac{\sum_{n=1}^{N}(V_n - \hat{V}_n)^2}{\sum_{n=1}^{N}(V_n - \hat{V}_n)^2}$
WI	$1 - \frac{\sum_{n=1}^{N}\lvert\hat{V}_n - V_n\rvert}{\sum_{n=1}^{N}\lvert V_n - \bar{V}_n\rvert + \lvert\hat{V}_n - \bar{V}_n\rvert}$
R^2	$1 - \frac{\sum_{n=1}^{N}(V_n - \hat{V}_n)^2}{\sum_{n=1}^{N}(\sum_{n=1}^{N}V_n) - V_n)^2}$

DOI: https://doi.org/10.54216/JAIM.070102
Received: April 27, 2023 Revised: August 11, 2023 Accepted: January 01, 2024

r	$\frac{\sum_{n=1}^{N}(\hat{V}_n - \hat{V}_n)(V_n - \bar{V}_n)}{\sqrt{\left(\sum_{n=1}^{N}\left(\hat{V}_n - \bar{\hat{V}}_n\right)^2\right)\left(\sum_{n=1}^{N}(V_n - \bar{V}_n)^2\right)}}$

Some metrics matter more within our holistic review process. Hence, ignoring the facility to manage the risk of preterm birth's is highly irresponsible. The key metrics include:

1. **Mean Squared Error (MSE):** This calculates the mean square difference between actual and expected outcomes, with a smaller MSE indicating a superior model.

2. **Root Mean Squared Error (RMSE):** The square root of MSE, reflecting the average size of errors. A lower RMSE is preferable, analogous to MSE.

3. **Mean Absolute Error (MAE):** This determines the average absolute difference between actual and anticipated values, providing insight into the extent of average discrepancies.

4. **Mean Bias Error (MBE):** The average discrepancy between presumed and observed values. Positive values indicate overestimation, while negative values suggest underestimation.

5. **R (Correlation Coefficient):** Indicates the strength and direction of a linear relationship between two factors, with a value close to 1 signifying a very strong positive relationship.

6. **R^2 (Coefficient of Determination):** Reveals the extent to which variations in a dependent variable can be explained by certain variables, with a higher R^2 indicating a better model fit.

7. **RRMSE (Relative Root Mean Squared Error):** Represents the relationship between the smallest and largest values, providing normalized measures of precision in estimates.

8. **NSE (Nash-Sutcliffe Efficiency):** Tests the accuracy of model forecasts in comparison to observed annual average metrics. Values closer to 1 indicate better model performance.

9. **WI (Willmott Index):** Determines the extent to which actual and estimated figures correspond. A WI value of 1 signifies perfect agreement, while zero indicates a need for improvement.

10. **Fitted Time:** Represents the time required to train or work on the data, with smaller values considered faster.

3.4 Feature Selection Methods

Feature selection is an important step in both performance enhancement and model interpretability. Five different feature selection methods are strategically applied, each one with specific properties that are molded to the "Students Performance" dataset.

• bWWPA (Binary Waterwheel Plant Optimization Algorithm): This approach efficiently searches through the solution space to pick and keep the most informative features. Being inspired by the natural behavior of whales, it optimizes feature weights to highlight the most important variables [24].

• bPSO (Binary Particle Swarm Optimization): Mimicking the collective behavior of particles, bPSO navigates the feature space jointly to select informative features for model training. The swarm intelligence approach ensures an extensive search of significant variables [25].

• bWAO (Binary Whale Optimization Algorithm): Like bWWPA, bWAO employs binary coding for easy navigation optimization landscape filtering out the key features. The application of this binary optimization method brings about its computational advantages but ensures effectiveness [26].

• bGWO (Binary Grey Wolf Optimization): It is based on the cooperative hunting strategy of grey wolves, which bGWO performs well in extracting significant features that contribute to the performance of predictive modeling. The binary coding ensures a multitude of features of different combinations to be explored [27].

DOI: https://doi.org/10.54216/JAIM.070102
Received: April 27, 2023 Revised: August 11, 2023 Accepted: January 01, 2024

• bFA (Binary Firefly Algorithm): Inhaling the firefly's flickering behavior, bFA traverses the feature space, stressing feature dominance. While this nature-inspired algorithm performs feature selection, it also preserves computational efficiency [28].

Cumulatively, these feature selection methodologies contribute to variable filtering and enhance a model process that is more oriented and interpretable. Through the selective preservation of the informative features, the models are able to grasp the idea of the dataset more precisely.

3.5 Evaluation Metrics for Feature Selection Algorithm

In the evaluation of the Feature Selection Algorithm's performance, diverse metrics are employed to gauge its effectiveness. Let's define key notations: M represents the number of repetitions of an optimizer's runs for the feature selection problem; g∗j denotes the best solution at the j-th run; and N signifies the number of tested points.

1. **Average Error (AvgError):**
   The average Error metric serves as a cornerstone in the determination of the accuracy of the classifier in representing the chosen feature set. It is calculated using the formula: He returned to Morningside with *a wicker laundry bag* in his hand.

$$\text{AvgError} = 1 - \frac{1}{M}\sum_{j=1}^{M}\frac{1}{N}\sum_{i=1}^{N}\text{Match}(C_i, L_i)$$

   Where C_i is the label of the classifier output for point i, L_i is the label for point i and the Match function scores the match between the inputs.

2. **Average Fitness (AvgSelectSize):**
   The Average Fitness suggests mean size of the selected features against the overall number of features in the dataset (D). Its calculation is expressed as: Person who informed about this situation.

$$\text{AvgSelectSize} = \frac{1}{M}\sum_{j=1}^{M}\frac{\text{size}(g_j^*)}{D}$$

   Here, size(g∗j) represents the size of g∗j vector.

3. **Mean:**
   The Mean represents the average of the solutions obtained from running an optimizer M times and is given by:

$$\text{Mean} = \frac{1}{M}\sum_{j=1}^{M} g_j^*$$

4. **Best Fitness (BestFn):**
   Best Fitness corresponds to the minimum fitness function achieved by an optimizer after M runs:

$$\text{BestFn} = \min_{j=1}^{M} g_j^*$$

5. **Worst Fitness (WorstFn):**
   Worst Fitness signifies the poorest solution found by an optimizer after M runs:

$$\text{WorstFn} = \max_{j=1}^{M} g_j^*$$

6. **Standard Deviation (SD):**
   The Standard Deviation (SD) is a measure of the variation of the obtained best solutions, giving an estimate of the stability and robustness of the optimiser. A smaller SD suggests better convergence to the same solution, and it is calculated as:

DOI: https://doi.org/10.54216/JAIM.070102
Received: April 27, 2023 Revised: August 11, 2023 Accepted: January 01, 2024

$$\mathrm{SD} = \sqrt{\frac{1}{M-1}\sum_{j=1}^{M}(g_j^* - \mathrm{Mean})^2}$$

Altogether, the evaluation metrics give an extensive evaluation of the Feature Selection. Algorithm's performance on tackling feature selection problems.

3.6 Experimental Setup

The experimental design is the most important determination of the reliability and replicates of the study results. Overall, the implemented approach is positively evaluated.

• Data Splitting: The datasets are divided into training and testing sets judiciously using a stratified approach which ensures that the outcome distribution represents the distribution of all outcomes. This enables the models to be evaluated on diverse instances which boost their generalizability.

• Model Training: Selection of some regression models is associated with trainings like, the use of cross-validation, model parameters optimization and seeing to their generalization. The models are trained on the training set to extract underlying structure from the data.

• Model Testing: This model is evaluated using the metrics such as MSE, RMSE, MAE and other metrics which are applied in a systematic manner. Validation on a disjoint dataset sample ensures unbiased estimation of predictive accuracy.

• Feature Selection Analysis: Meticulous statistical analyses and significance tests testify the influence of feature selection on model performance. The study compares the models with and without selected features determining the contribution of the variables to predictive accuracy.

• ANOVA and Wilcoxon Signed Rank Test: The statistical tests of ANOVA with Wilcoxon signed rank test being used to determine statistical significance of the observed differences in model performance. This test gives statistically strong inference about how well the proposed Methodology works.

• Transformed Model Evaluation: Complexity of model transformation in the aspect of predictive accuracy and interpretability is completed evaluated. This is a comprehensive study of the performance of the transformed models and their original counterparts with respect to the success of applying the transformations in emphasizing the difference.

Every step identified is purposefully executed in order to produce a complete and articulate overview of the integrated strategy. It provides tools which enable us to navigate into the effectiveness of feature selection, model transfer and overall joint results in raising the scores of the student performance prediction models of the "Students Performance" data set. The methodological rigor improves the accuracy and range of application of the results thus positioning the study as a weighty input in the framework development for predictive purposes in the educational field.

4. Results

This critical part presents the outcomes of our suggested methodology implementation on the "Students Performance" dataset. Our exhaustive review covers a multipronged perspective, assessing the predictive power of regression models, the efficiency of feature selection techniques, the transformative effect of model tuning, and the statistical significance of detected changes.

4.1 Machine Learning Before Feature Selection

We analyzed the predicting capability of ten regression models which we adopted in our paper as the major question of this work. Every model is taken through a rigorous reviewing process that is done by employing a variety of performance metrics. Table 2 illustrates our observations, showing many performance metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Bias Error (MBE), Coefficient of Determination (R^2), Relative RMSE (RRMSE), Nash-Sutcliffe.

DOI: https://doi.org/10.54216/JAIM.070102
Received: April 27, 2023 Revised: August 11, 2023 Accepted: January 01, 2024

Table 2: Models Performance Metrics

Models	mse	Rmse	Mae	mbe	r	R2	RRMSE	NSE	WI	Fitted Time
GradientBoostingRegressor	0.0475	0.2180	0.1753	0.0637	0.5419	0.2937	38.4638	0.2131	0.5618	0.0111
XGBoost	0.0511	0.2261	0.1755	0.0889	0.5517	0.3044	39.9036	0.1531	0.5612	32.3940
DecisionTreeRegressor	0.0580	0.2409	0.2001	0.0305	0.5700	0.3248	42.5172	0.0385	0.4998	0.2328
RandomForestRegressor	0.0601	0.2451	0.1849	0.1167	0.5462	0.2983	43.2481	0.0051	0.5378	0.4771
MLPRegressor	0.0644	0.2538	0.2062	0.1107	0.4167	0.1736	44.7827	0.0667	0.4844	5.5786
ExtraTreesRegressor	0.0653	0.2556	0.1972	0.1153	0.4896	0.2397	45.1070	0.0822	0.5069	0.0106
LinearRegression	0.0679	0.2606	0.2068	0.1016	0.4533	0.2055	45.9808	0.1246	0.4829	0.0141
CatBoost	0.0693	0.2632	0.1986	0.1186	0.3858	0.1489	46.4422	0.1472	0.5036	17.5646
SVR	0.0727	0.2697	0.2148	0.1506	0.4710	0.2218	47.5969	0.2050	0.4631	0.3140
KNeighborsRegressor	0.0868	0.2946	0.2400	0.1733	0.4055	0.1644	51.9959	0.4380	0.4000	0.1424
Pipeline	0.2798	0.5289	0.4407	0.1536	0.2845	0.0809	93.3440	0.6345	0.1017	0.2394

These results provide a complex insight into multifarious showings of the regression models of ours. Models that GradientBoostingRegressor and XGBoost display lower MSE and RMSE, indicating stronger predictive accuracy. On the other hand, models such as MLPRegressor and SVR have higher error metrics which suggests the need for improvement in their predictive power.

Figure 4 is a residual plot illustrating residuals from the model's predictions. This is visual tool performes a crucial function of the model's capability to make correct pattern observations and infer traits from data. Horizontal line deviations show those model regions which either could lead to failure or give inaccurate predictions. Researchers employ such deviations to enhance their predictive models and explore the source of predicted value bias.

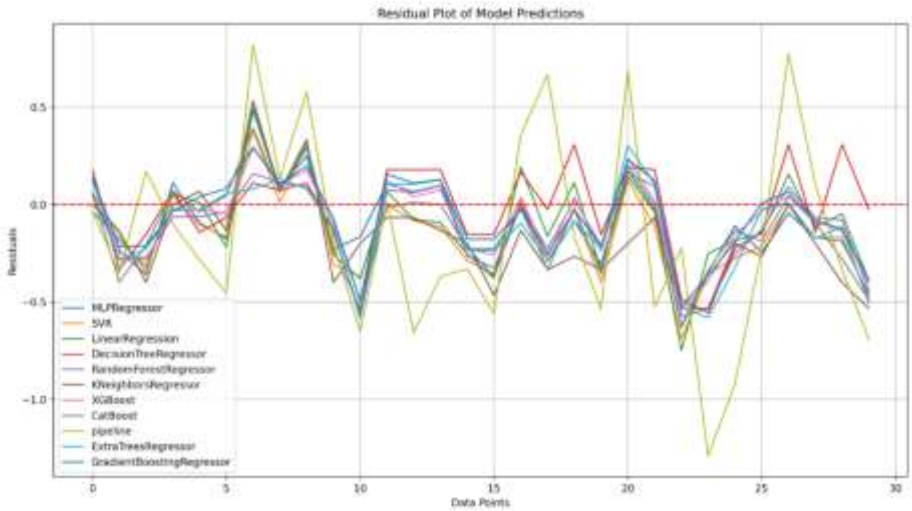

Figure 4: Residual Plot of Model Predictions

The correlation heatmap in Figure 5 gives in general the picture of the relationship among predicted models. This Figure helps readers cluster a model prediction accuracy against models as well as to understand whether different approaches cannot give the same result.

DOI: https://doi.org/10.54216/JAIM.070102
Received: April 27, 2023 Revised: August 11, 2023 Accepted: January 01, 2024

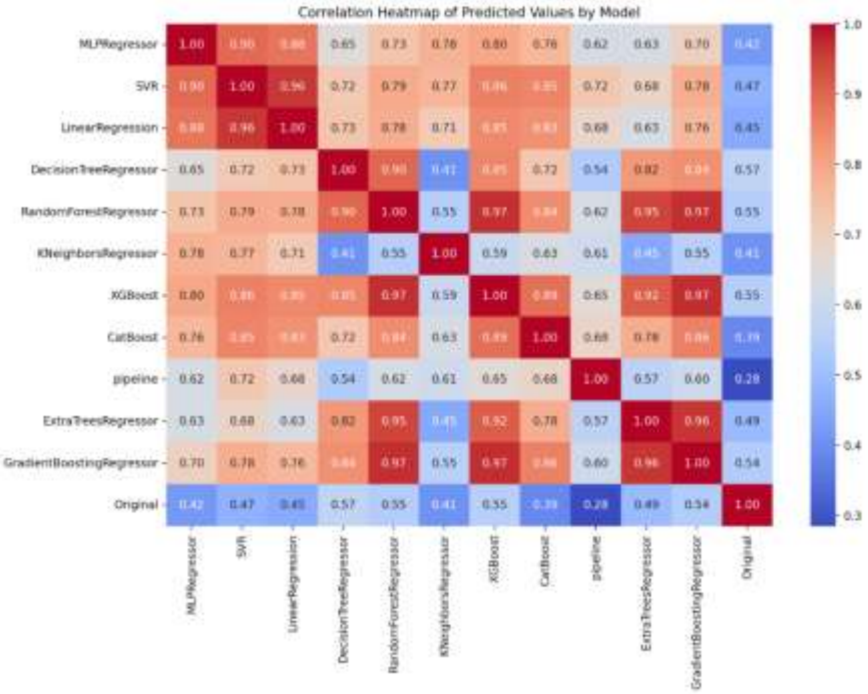

Figure 5: Correlation Heatmap of Predicted Values by Model

Figure 6 uses a radar plot to capture visually all of the model metrics. This figure enables us to compare per model performances across multiple metrics thus giving a bird eye view about their strengths and those areas that a model lags behind, hence models that are most suitable for a certain class of predictive tasks.

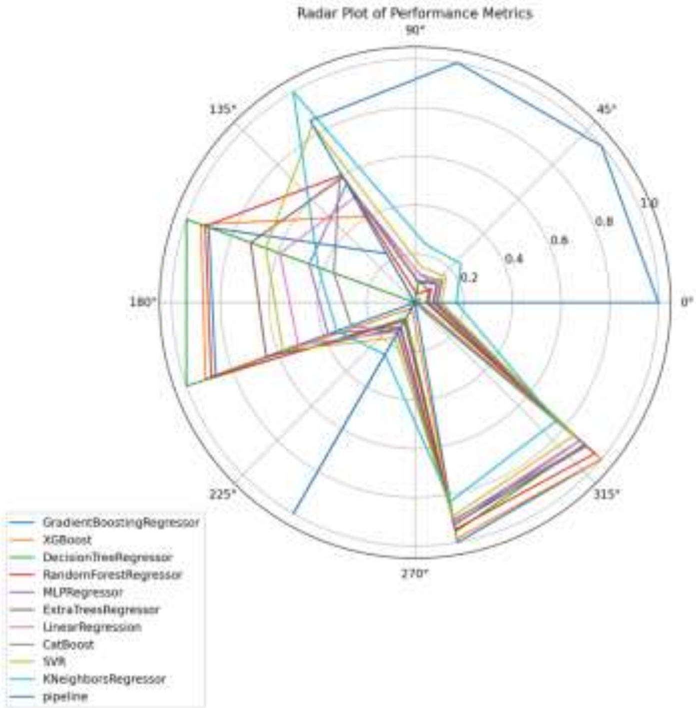

Figure 6: Radar Plot of Performance Metrics

4.2 Feature Selection Results

Recognizing the pivotal role of feature selection in model interpretability and efficiency, we delve into the outcomes of five distinctive feature selection methods: bWWPA, bPSO, bWAO, bGWO, and bFA. Table 3 unfolds a comprehensive narrative, illustrating the results of these methods in terms of average error, average select size, average fitness, best fitness, worst fitness, and standard deviation of fitness [29].

Table 3: Feature Selection Results

DOI: https://doi.org/10.54216/JAIM.070102
Received: April 27, 2023 Revised: August 11, 2023 Accepted: January 01, 2024

Feature Selection	bWWPA	bPSO	bWAO	bGWO	bFA
Average error	0.31814	0.35194	0.35174	0.33824	0.35034
Average Select size	0.45374	0.45374	0.61714	0.37654	0.48824
Average Fitness	0.38034	0.37874	0.38654	0.38644	0.43064
Best Fitness	0.30064	0.35904	0.35064	0.36424	0.34934
Worst Fitness	0.36754	0.42674	0.42674	0.44044	0.44694
Standard deviation Fitness	0.19114	0.19054	0.19274	0.19174	0.22734

These outcomes provide practitioners with valuable insights into the comparative effectiveness of different feature selection techniques, aiding in informed decision-making for predictive modeling.

Table 4 presents key descriptive statistics for five feature selection methods (bWWPA, bPSO, bWAO, bGWO, bFA), including measures such as minimum, maximum, mean, standard deviation, and more. This summary provides a quick glance at the central tendencies and distribution of values, aiding researchers and practitioners in understanding the performance characteristics of each feature selection method in the context of predictive modeling.

Table 4: Summary of Feature Selection Methods

	bWWPA	bPSO	bWAO	bGWO	bFA
Number of values	10	10	10	10	10
Minimum	0.3171	0.3432	0.3332	0.3282	0.3303
25% Percentile	0.3181	0.3519	0.3492	0.3382	0.3493
Median	0.3181	0.3519	0.3517	0.3382	0.3503
75% Percentile	0.3181	0.3519	0.3517	0.3382	0.3503
Maximum	0.3198	0.3632	0.3622	0.3408	0.359
Range	0.002674	0.02	0.029	0.01258	0.02869
Mean	0.3182	0.3522	0.3499	0.3375	0.3488
Std. Deviation	0.000646	0.004744	0.007606	0.003353	0.007225
Std. Error of Mean	0.000204	0.0015	0.002405	0.00106	0.002285

Figure 7 presents the average quality measures of feature selection models that exactly indicate their individual accuracy. This figure provides for practitioners, a useful reference in knowing the efficacy of each feature selection method in the reduction of prediction errors, aiding them to choose the right techniques that aligns with the objectives of their predictive model assessment activities.

DOI: https://doi.org/10.54216/JAIM.070102
Received: April 27, 2023 Revised: August 11, 2023 Accepted: January 01, 2024

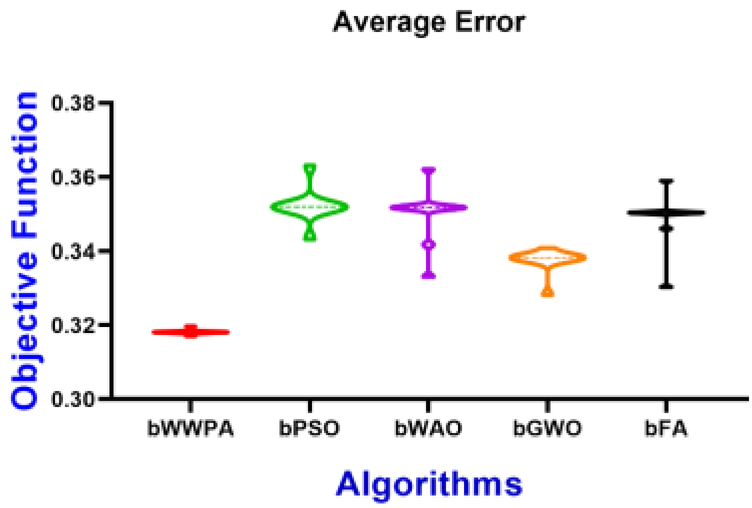

Figure 7: Feature Selection Models Average Error

4.3 ANOVA and Wilcoxon Signed Rank Test Results

We utilize both ANOVA and Wilcoxon signed-rank tests to strictly examine the statistical significance of the observed differences in model performance. Table 5 supplies an ANOVA table incorporating the sum of squares, degrees of freedom, mean squares, F -statistic, and the respective p-value [30].

Table 5: The ANOVA Table

ANOVA table	SS	DF	MS	F (DFn, DFd)	P value
Treatment (between columns)	0.004837	4	0.001209	F (4, 45) = 7.696	P<0.0001
Residual (within columns)	0.007071	45	0.000157		
Total	0.01191	49			

Figure 8 presents the ANOVA results which is a statistical technique that is used to compare the group means. The significances and the differences among the groups can be observed from this figure, that will help to understand the factors of the variations of the data.

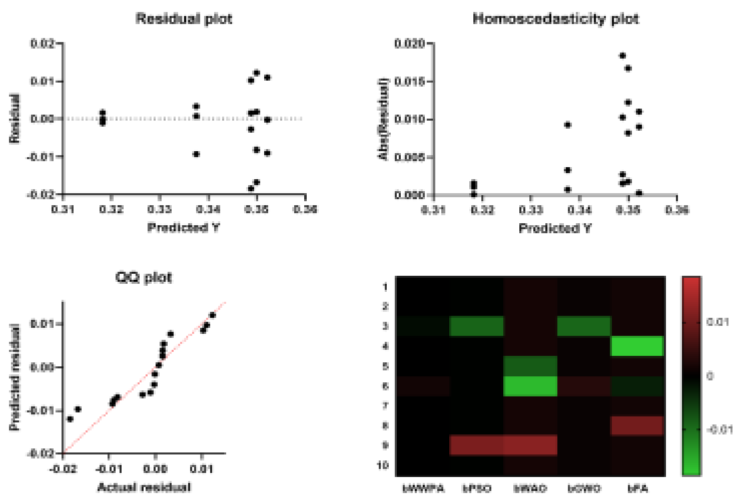

Figure 8: The ANOVA Test Plots

DOI: https://doi.org/10.54216/JAIM.070102
Received: April 27, 2023 Revised: August 11, 2023 Accepted: January 01, 2024

The obtained p-value is instrumental, being a key assessment measure of the significance of the model performance disparities among the compared models. Moreover, Table 6 presents the results of the Wilcoxon signed rank test, thus reinforcing the differences observed.

Table 6: Wilcoxon Signed Rank Test Results

	bWWPA	bPSO	bWAO	bGWO	bFA
Number of values	10	10	10	10	10
Sum of signed ranks (W)	55	55	55	55	55
Sum of positive ranks	55	55	55	55	55
Sum of negative ranks	0	0	0	0	0
P value (two tailed)	0.002	0.002	0.002	0.002	0.002
Exact or estimate?	Exact	Exact	Exact	Exact	Exact
P value summary	**	**	**	**	**
Significant (alpha=0.05)?	Yes	Yes	Yes	Yes	Yes

The statistical analyses also support the reliability and validity of the discerned distinctions in model performance. The substantiating our findings.

4.3 Machine Learning After Feature Selection

Now back to the effect of the model performance, Table 7 contains the parameters. The efficacy of transformation models is measured in this section as it pertains to increased predictive precision.

Table 7: Transformed Model Performance Metrics

Models	mse	Rmse	Mae	mbe	R	R2	RRMSE	NSE
GradientBoostingRegressor	0.0064	0.0293	0.0235	0.0085	0.9175	0.8577	22.8638	0.7911
XGBoost	0.0069	0.0304	0.0236	0.0119	0.8850	0.8684	24.3036	0.7311
DecisionTreeRegressor	0.0078	0.0323	0.0269	0.0041	0.9033	0.8888	26.9172	0.6165
RandomForestRegressor	0.0081	0.0329	0.0248	0.0157	0.8795	0.8623	27.6481	0.5831
MLPRegressor	0.0086	0.0341	0.0277	0.0149	0.7500	0.7376	29.1827	0.6447

Such results indicate the potency of our technique of model transformation for improving prediction accuracy, thus providing necessary scaffold for further refinement of the predictive framework in the educational context.

A heatmap shown in Figure 9 displays the performance metrics of the machine learning models after the feature selection. The heatmap intensity color legend is a synopsis on model scores. These graphs allow the user to see the characteristics, patterns, trends, and areas that need to be enhanced, and thus improve the understanding of the role which the features play in the overall performance of the machine-learning model.

DOI: https://doi.org/10.54216/JAIM.070102
Received: April 27, 2023 Revised: August 11, 2023 Accepted: January 01, 2024

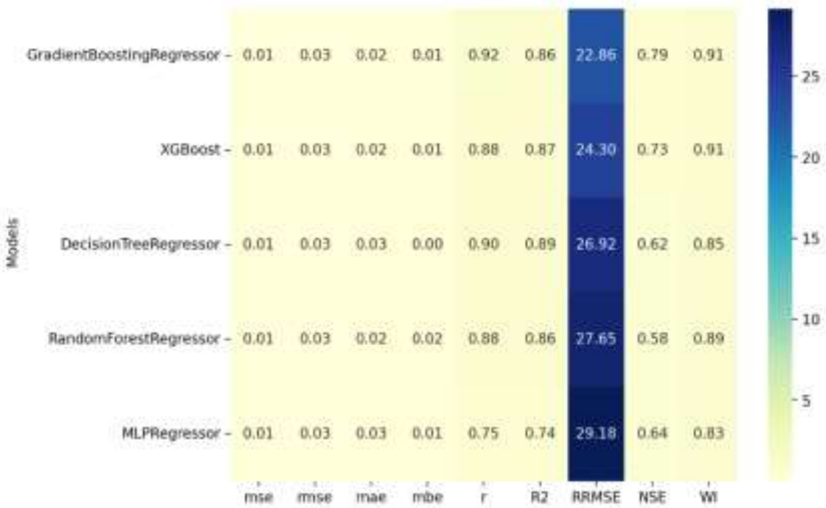

Figure 9: Performance Heatmap Plot of Machine Learning Models with Feature Extraction applied

In Figure 10 The radar plot is used subsequently to showcase the performance characteristics of the ML models after the thorough feature selection procedure. This graphical display allows us to see how multiple metrics occur in various models distributed from all over the world. Radar plots give a compact yet illusionary perspective on each model by using the bullets that show in detail and at the same time in a summarized manner the overwhelming points of different models.

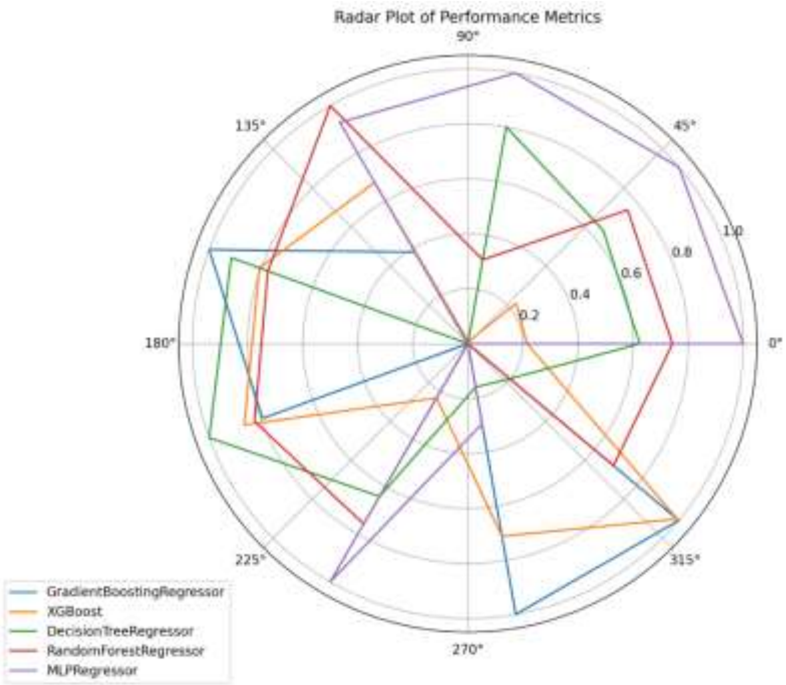

Figure 10: ML Models Performance after Feature Selection

In short, our worked results tell a vivid story of our experimental work. Starting from the subtleties of predictive capabilities of regression models, through the outcome of feature selection approaches, statistical significance of the observed differences, and culminating in the transformative aptness of the perspectives considered, our findings provide multifaceted insights on the area of student performance prediction. They serve not only the scholarly discourse but also provide practical solutions to move predictive frameworks further in educational contexts.

5. Conclusion

This research, through a novel integrated approach of feature engineering and model transformation, aims at the continuous improvement of student performance prediction models. The study brings a new perspective on relationships between regression models and feature selection methods. Amongst the

DOI: https://doi.org/10.54216/JAIM.070102
Received: April 27, 2023 Revised: August 11, 2023 Accepted: January 01, 2024

different models studied, GradientBoostingRegressor and XGBoost stand out as champions, demonstrating the best predictive accuracy. As such, methodologies like bWWPA and bFA show encouraging results, eliminating errors and maximizing fitness. The model performance differences were meaningfully evidenced by the strongly supported statistical results from full data analyses, including ANOVA and Wilcoxon signed-rank tests. Going beyond the academic realm, our results inform concrete action for schools, recommending targeted interventions in the light of influential factors. This study, however, acknowledges its inherent limitations, like the case with dataset constrictions, and it is in this regard that it paves the way for future inquiries. Such research may concentrate on alternative models, more diverse datasets, and longitudinal studies, all in a bid to enlarge predictive efficiency and, at the same time, shift focus to other issues within the education trajectory. With the closing of this research chapter, we envisage our work steering the evolution of educational data analytics, cementing a symbiotic relationship between machine learning and education, and firing up finely crafted interventions that enhance the student's journey.

Funding: "This research received no external funding"

Conflicts of Interest: "The authors declare no conflict of interest."

References

[1] Mishra, P., Biancolillo, A., Roger, J. M., Marini, F., & Rutledge, D. N. (2020). New data preprocessing trends based on ensemble of multiple preprocessing techniques. *TrAC Trends in Analytical Chemistry*, *132*, 116045. https://doi.org/10.1016/j.trac.2020.116045

[2] Naser, M. Z., & Alavi, A. H. (2023). Error Metrics and Performance Fitness Indicators for Artificial Intelligence and Machine Learning in Engineering and Sciences. *Architecture, Structures and Construction*, *3*(4), 499–517. https://doi.org/10.1007/s44150-021-00015-8

[3] Karasu, S., Altan, A., Bekiros, S., & Ahmad, W. (2020). A new forecasting model with wrapper-based feature selection approach using multi-objective optimization technique for chaotic crude oil time series. *Energy*, *212*, 118750. https://doi.org/10.1016/j.energy.2020.118750

[4] Cerqueira, V., Torgo, L., & Mozetič, I. (2020). Evaluating time series forecasting models: An empirical study on performance estimation methods. *Machine Learning*, *109*(11), 1997–2028. https://doi.org/10.1007/s10994-020-05910-7

[5] Vargas-Madriz, L. F., & Nocente, N. (2023). Exploring students' willingness to provide feedback: A mixed methods research on end-of-term student evaluations of teaching. *Social Sciences & Humanities Open*, *8*(1), 100525. https://doi.org/10.1016/j.ssaho.2023.100525

[6] Gordeeva, T., Sheldon, K., & Sychev, O. (2020). Linking academic performance to optimistic attributional style: Attributions following positive events matter most. *European Journal of Psychology of Education*, *35*(1), 21–48. https://doi.org/10.1007/s10212-019-00414-y

[7] Khan, A., & Ghosh, S. K. (2018). Data mining based analysis to explore the effect of teaching on student performance. *Education and Information Technologies*, *23*(4), 1677–1697. https://doi.org/10.1007/s10639-017-9685-z

[8] An, M., Zhang, X., Wang, Y., Zhao, J., & Kong, L. (2022). Reciprocal relations between achievement goals and academic performance in a collectivist higher education context: A longitudinal study. *European Journal of Psychology of Education*, *37*(3), 971–988. https://doi.org/10.1007/s10212-021-00572-y

[9] Li, C., Yao, J., Tang, Z., Tang, Y., & Zhang, Y. (2023). The Influence of the Student's Online Learning Behaviors on the Learning Performance. In B. Li, L. Yue, C. Tao, X. Han, D. Calvanese, & T. Amagasa (Eds.), *Web and Big Data* (pp. 28–36). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-25158-0_3

[10] Khan, A., & Ghosh, S. K. (2021). Student performance analysis and prediction in classroom learning: A review of educational data mining studies. *Education and Information Technologies*, *26*(1), 205–240. https://doi.org/10.1007/s10639-020-10230-3

[11] Shum, A., & Fryer, L. K. (2023). Grade goal effects on the interplay between motivation and performance in undergraduate gateway mathematics courses. *Contemporary Educational Psychology*, *75*, 102228. https://doi.org/10.1016/j.cedpsych.2023.102228

DOI: https://doi.org/10.54216/JAIM.070102
Received: April 27, 2023 Revised: August 11, 2023 Accepted: January 01, 2024

[12] Waters, L. E., Loton, D., & Jach, H. K. (2019). Does Strength-Based Parenting Predict Academic Achievement? The Mediating Effects of Perseverance and Engagement. *Journal of Happiness Studies*, *20*(4), 1121–1140. https://doi.org/10.1007/s10902-018-9983-1

[13] Schweder, S., & Raufelder, D. (2022). Students' interest and self-efficacy and the impact of changing learning environments. *Contemporary Educational Psychology*, *70*, 102082. https://doi.org/10.1016/j.cedpsych.2022.102082

[14] Otchere, D. A., Ganat, T. O. A., Ojero, J. O., Tackie-Otoo, B. N., & Taki, M. Y. (2022). Application of gradient boosting regression model for the evaluation of feature selection techniques in improving reservoir characterisation predictions. *Journal of Petroleum Science and Engineering*, *208*, 109244. https://doi.org/10.1016/j.petrol.2021.109244

[15] Ji, C., Zou, X., Hu, Y., Liu, S., Lyu, L., & Zheng, X. (2019). XG-SF: An XGBoost Classifier Based on Shapelet Features for Time Series Classification. *Procedia Computer Science*, *147*, 24–28. https://doi.org/10.1016/j.procs.2019.01.179

[16] Singh Kushwah, J., Kumar, A., Patel, S., Soni, R., Gawande, A., & Gupta, S. (2022). Comparative study of regressor and classifier with decision tree using modern tools. *Materials Today: Proceedings*, *56*, 3571–3576. https://doi.org/10.1016/j.matpr.2021.11.635

[17] Xue, L., Liu, Y., Xiong, Y., Liu, Y., Cui, X., & Lei, G. (2021). A data-driven shale gas production forecasting method based on the multi-objective random forest regression. *Journal of Petroleum Science and Engineering*, *196*, 107801. https://doi.org/10.1016/j.petrol.2020.107801

[18] Maqbool, J., Aggarwal, P., Kaur, R., Mittal, A., & Ganaie, I. A. (2023). Stock Prediction by Integrating Sentiment Scores of Financial News and MLP-Regressor: A Machine Learning Approach. *Procedia Computer Science*, *218*, 1067–1078. https://doi.org/10.1016/j.procs.2023.01.086

[19] Ossai, C. I., & Egwutuoha, I. P. (2020). Anomaly Detection and Extra Tree Regression for Assessment of the Remaining Useful Life of Lithium-Ion Battery. In L. Barolli, F. Amato, F. Moscato, T. Enokido, & M. Takizawa (Eds.), *Advanced Information Networking and Applications* (pp. 1474–1488). Springer International Publishing. https://doi.org/10.1007/978-3-030-44041-1_124

[20] Schmidt, A. F., & Finan, C. (2018). Linear regression and the normality assumption. *Journal of Clinical Epidemiology*, *98*, 146–151. https://doi.org/10.1016/j.jclinepi.2017.12.006

[21] Jabeur, S. B., Gharib, C., Mefteh-Wali, S., & Arfi, W. B. (2021). CatBoost model and artificial intelligence techniques for corporate failure prediction. *Technological Forecasting and Social Change*, *166*, 120658. https://doi.org/10.1016/j.techfore.2021.120658

[22] Anand, P., Rastogi, R., & Chandra, S. (2020). A class of new Support Vector Regression models. *Applied Soft Computing*, *94*, 106446. https://doi.org/10.1016/j.asoc.2020.106446

[23] Ortiz-Bejar, J., Graff, M., Tellez, E. S., Ortiz-Bejar, J., & Jacobo, J. C. (2018). K-Nearest Neighbor Regressors Optimized by using Random Search. *2018 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC)*, 1–5. https://doi.org/10.1109/ROPEC.2018.8661399

[24] Akinlar, M. A., Tchier, F., & Inc, M. (2020). Chaos control and solutions of fractional-order Malkus waterwheel model. *Chaos, Solitons & Fractals*, *135*, 109746. https://doi.org/10.1016/j.chaos.2020.109746

[25] Qasim, O. S., & Algamal, Z. Y. (2018). Feature selection using particle swarm optimization-based logistic regression model. *Chemometrics and Intelligent Laboratory Systems*, *182*, 41–46. https://doi.org/10.1016/j.chemolab.2018.08.016

[26] Rana, N., Latiff, M. S. A., Abdulhamid, S. M., & Chiroma, H. (2020). Whale optimization algorithm: A systematic review of contemporary applications, modifications and developments. *Neural Computing and Applications*, *32*(20), 16245–16277. https://doi.org/10.1007/s00521-020-04849-z

[27] Tikhamarine, Y., Souag-Gamane, D., Najah Ahmed, A., Kisi, O., & El-Shafie, A. (2020). Improving artificial intelligence models accuracy for monthly streamflow forecasting using grey Wolf optimization (GWO) algorithm. *Journal of Hydrology*, *582*, 124435. https://doi.org/10.1016/j.jhydrol.2019.124435

[28] Zhang, L., Mistry, K., Lim, C. P., & Neoh, S. C. (2018). Feature selection using firefly optimization for classification and regression models. *Decision Support Systems*, *106*, 64–85. https://doi.org/10.1016/j.dss.2017.12.001

[29] Bommert, A., Sun, X., Bischl, B., Rahnenführer, J., & Lang, M. (2020). Benchmark for filter methods for feature selection in high-dimensional classification data. *Computational Statistics & Data Analysis*, *143*, 106839. https://doi.org/10.1016/j.csda.2019.106839

[30] Weissgerber, T. L., Garcia-Valencia, O., Garovic, V. D., Milic, N. M., & Winham, S. J. (2018). Why we need to report more than "Data were Analyzed by t-tests or ANOVA." *eLife*, *7*, e36163. https://doi.org/10.7554/eLife.36163

DOI: https://doi.org/10.54216/JAIM.070102
Received: April 27, 2023 Revised: August 11, 2023 Accepted: January 01, 2024