JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 09 Issue 02
2025

Research Article
https://doi.org/10.56979/902/2025

A Predictive Model and Performance Evaluation in Mathematics for Primary Education

Marvi Abro[1], Imtiaz Husain[1*], Syed M. Hassan Zaidi[1], Faryal Sheikh[2], and Ghulam Murtaza[1]

[1]Department of Artificial Intelligence and Mathematical Sciences, SMIU, Karachi-74000, Pakistan.
[2]Department of English, SMIU, Karachi-74000, Pakistan.
[*]Corresponding Author: Imtiaz Hussain. Email: imtiaz@smiu.edu.pk

Received: June 30, 2025 **Accepted:** August 12, 2025

Abstract: This study investigates the application of predictive modelling to assess and forecast students' academic performance in primary mathematics education. Four regression techniques, Linear Regression, Decision Tree Regression, Random Forest Regression, and K-Nearest Neighbours Regression, were implemented and comparatively evaluated. Model performance was measured using Mean Squared Error (MSE) as the primary metric. Results indicate that Linear Regression achieved the lowest MSE (1.33), establishing a strong predictive baseline. Although Decision Tree Regression effectively captured non-linear patterns, it yielded a substantially higher MSE (62.38), highlighting the risk of overfitting. Random Forest Regression improved generalization by aggregating multiple decision trees, achieving an MSE of 25.21. Meanwhile, K-Nearest Neighbours Regression provided localized predictive accuracy with a competitive MSE of 19.29. Collectively, these findings contribute to the growing body of research on data-driven approaches in education, providing practical insights for educators and policymakers to leverage predictive analytics and enhance learning outcomes in primary mathematics.

Keywords: Predictive Modelling; Regression Analysis; Educational Data Mining; Mathematics Education; Machine Learning in Education; Model Performance Evaluation

1. Introduction

In contemporary education, the accurate assessment and prediction of student performance play a pivotal role in enhancing learning outcomes and ensuring academic success. With the exponential growth of data stored in educational databases, institutions are confronted with complex challenges, including identifying factors contributing to poor performance and mitigating risks of student attrition. In this context, the integration of machine learning (ML) techniques has emerged as a powerful approach for predicting learning outcomes and diagnosing students' academic deficiencies. A growing body of literature demonstrates the potential of ML-based models in educational data mining, drawing insights from interdisciplinary research in computer science, psychology, and pedagogy. For instance, studies have shown that ensemble methods such as Random Forest classifiers often outperform traditional algorithms like Support Vector Machines in predicting student performance with higher accuracy [1]. Similarly, other works have emphasized the importance of early prediction models to inform decision-making, enabling timely interventions that support student learning and retention [2].

Despite these advances, much of the existing research has predominantly focused on classification approaches or higher education datasets, with relatively limited exploration of regression-based predictive models tailored to primary mathematics education. Classification models, while useful for determining categorical outcomes such as 'pass' or 'fail,' often overlook the degree of variation in student achievement. In contrast, regression techniques forecast continuous performance scores, allowing educators to measure the extent of learning gaps and design interventions of appropriate intensity. Recent studies [3], [4], [5] highlight how regression-based approaches provide richer insights into performance trajectories, particularly in early education contexts. Addressing this critical gap, the present study evaluates and compares multiple regression models; Linear Regression, Decision Tree Regression, Random Forest Regression, and K-Nearest Neighbours Regression applied to primary mathematics performance. By rigorously assessing their predictive accuracy, this research contributes to the growing discourse on data-driven educational interventions and offers practical implications for enhancing student outcomes at the foundational level [6 - 8].

In 2023, several research efforts were directed toward the classification and prediction of student performance data using diverse machine learning (ML) algorithms. One such study demonstrated how data mining techniques, including Naïve Bayes, ID3, C4.5, and Support Vector Machines (SVM), could be effectively applied to the UCI student performance dataset, with algorithmic performance evaluated through accuracy and error rate parameters [9], [10]. In the same year, a combined strategy was introduced that integrated statistically enabled ML algorithms such as Fuzzy C-Means, Logistic Regression, and Random Forest, providing a comparative analysis of student performance prediction [11]. This study presented detailed findings across multiple algorithmic groupings, with performance indicators such as accuracy, detection rate, and false alarm rate offering practical insights. Building on these developments, further research applied four ML methods, Fuzzy C-Means, Multi-Layer Perceptron, Logistic Regression, and Random Forest, to classify students' academic achievements at the college level [12], [13].

The growing success of artificial intelligence and machine learning has profoundly influenced the landscape of learning and education, steering societies toward more innovative models of knowledge development [14]. Advances across disciplines such as cognitive science, psychology, and educational technology highlight the need for assessment systems that move beyond conventional testing practices and embrace data-driven approaches [15], [16]. In this regard, statistical models [17], when coupled with ML-based predictive frameworks, provide a robust means of generating comprehensive insights into student performance, thus enabling more adaptive and evidence-based educational strategies [18]. These studies collectively establish the foundation for exploring regression-based approaches in primary mathematics education, thereby addressing an important research gap and setting the stage for the present study.

Various studies have been conducted assessing students' academic performance generally and in mathematics specifically by using different techniques, including machine learning and predictive models. The role of student identities on academic achievement gave a central place to intersectionality as it influences the entire person [19], [17]. In primary level education, different subjects are taught, including mathematics. Students consider mathematics a difficult subject, and their performance is poor in mathematics [18]. Building competence in mathematics is considered essential for professional and personal development [19]. Mathematics provides a better future vision to solve and understand the problems [20].

Mathematics is the basic competence and has great importance in our lives; it has its fundamental role in different disciplines of life, such as technical education, banking, cartography, research and so on [21] and to improve individuals' lives with value and excellence [22]. Assessing students' performance is the key concept to keeping the students on the right track [23]. Comparative research on the performance of students in mathematics by using different ways, such as statistical analysis strategies, five classes of machine learning algorithms (ML), deep belief network (DBN), which is a deep learning method. Random Forest (RF) data set showed outstanding performance in predicting the performance of students. The performance of students in mathematics with different socioeconomic factors was evaluated, and a predictive model was used for evaluating their achievements by different machine learning algorithms, including ensemble methods, decision trees, and linear regression method [24], [25]. While classification methods are capable of predicting categorical outcomes, such as whether a student is likely to succeed or fail, they do not provide the granularity required to capture subtle differences in performance. Regression-based models, on the other hand, estimate exact achievement levels, which allows educators to identify not only students at risk but also the magnitude of their learning deficits. This fine-grained predictive power is especially critical in primary education, where early detection of performance variation can inform personalized interventions and long-term academic planning.

2. Materials and Methods

2.1. Dataset Description

The dataset comprised students' scores in Sindhi, English, and Mathematics from the previous academic year, supplemented with demographic variables including gender, socio-economic status (SES), and attendance. SES was quantified using parental education and household income, categorized into three levels (low, medium, high). Attendance was measured as the percentage of days present during the academic year and normalized to a 0–1 scale. These variables were integrated into the feature set to capture both academic and contextual predictors of student performance.

Sindhi: assessed through reading and dictation tasks,

English: evaluated by listening to an audio passage followed by comprehension questions,

Mathematics: measured to determine learners' quantitative reasoning and problem-solving abilities.

In addition to academic scores, the dataset incorporated demographic and contextual features, including students' gender, socio-economic status, and attendance records. Socio-economic status was quantified using a composite index of parental education and household income, categorized into three levels (low, medium, high). Attendance was measured as the percentage of school days attended during the academic year, which was normalized to a 0–1 scale for comparability across regression models. These demographic and behavioral variables were integrated into the feature set alongside subject scores to capture broader influences on student performance, which were considered as supplementary predictors of academic performance.

2.2. Feature Engineering

To construct a comprehensive dataset, students' historical performance in Sindhi, English, and Mathematics was systematically extracted from institutional records. These subject scores were merged with demographic and attendance-related variables to form the final feature set. Feature preprocessing steps, including normalization and scaling, were applied to ensure comparability of predictors across regression models.

2.3. Regression Models Applied

Preprocessing steps included normalization of continuous variables and one-hot encoding of categorical variables. Hyperparameter tuning was performed using grid search with 5-fold cross-validation to ensure optimal performance. For the KNN Regressor, $k = 5$ provided the best balance of bias and variance. The Decision Tree Regressor was tuned for maximum depth between 3 and 10, with depth = 6 yielding optimal results. For the Random Forest Regressor, the number of estimators was varied between 50 and 200, with 100 trees providing the lowest cross-validation error.

In addition to the four core models, two supplementary algorithms were included for benchmarking: a Multi-Layer Perceptron (MLP) Regressor and a Support Vector Regressor (SVR). Their inclusion enables comparison with more complex, non-linear models often applied in educational prediction, thereby contextualizing the performance of the primary regression approaches.

Four machine learning regression models were implemented and compared for predictive performance:

1. **Linear Regression –** Used as a baseline model to establish predictive accuracy under the assumption of linear relationships between predictors and outcomes.
2. **Decision Tree Regression –** Applied to capture non-linear patterns in the dataset, though susceptible to overfitting without careful parameter tuning.
3. **Random Forest Regression –** Leveraged as an ensemble method combining multiple decision trees to improve generalization and reduce variance.
4. **K-Nearest Neighbors (KNN) Regression –** Employed to provide localized predictions by estimating student performance based on similarities to neighboring data points.

2.4. Model Evaluation

To ensure robust performance assessment, the dataset was divided into training and testing subsets using an 80:20 split ratio. The models were trained on the training set and subsequently evaluated on the unseen test set. In addition, k-fold cross-validation was employed to minimize bias and variance in the results, providing a more reliable estimate of model generalizability.

Model performance was quantified using Mean Squared Error (MSE), chosen for its sensitivity to large prediction errors and suitability for regression analysis. Comparative analysis of MSE values across models enabled the identification of the most effective regression approach for predicting primary mathematics performance.

2.4.1. Exploratory Data Analysis (EDA)

The EDAs were done for count, mean, standard deviation, minimum and maximum values, and first and third quartiles of the key subjects Sindhi, English, and Mathematics. Results revealed the mean, and the spread is the statistics given for the study participants' grade point average. Table 1 summarizes basic statistics of student performance across subjects, showing substantial variability, particularly in mathematics. Table 2 presents gender distribution, revealing a male-dominant sample (64.3%), which may affect subgroup analyses. Table 3 disaggregates scores across grade levels, demonstrating that performance improves in higher classes, while Table 4 consolidates subject averages for overall comparison.

Figures 1–3 (boxplots) confirm that mathematics scores are more widely dispersed, justifying the study's focus on this subject. Figure 4 illustrates gender imbalance, while Figure 5 shows dataset composition by data type. Yearly patterns in Figure 6 highlight fluctuations in enrollment and performance across 2017–2023, and Figure 7 (histograms) reveals that most mathematics scores skew below 60, underscoring challenges in prediction. Together, these descriptive analyses provide essential context for interpreting model performance.

Table 1. Basic Statistics

	Sindhi	English	Mathematics
count	661	661	661
mean	45.2	44.7	47.7
Std	19.9	19.2	21.9
min	0.0	0.0	15
25%	33	33	33
50%	33	33	33
75%	55	53	62
max	100	100	100

Table 2. Gender Distribution

Gender	Count
M	425
F	236

Table 3. Average Scores by Class

Class	Sindhi	English	Mathematics
Class 1	43.4	45.3	48.7
Class 2	42.2	42.6	42.7
Class 3	42.7	42.5	44.6
Class 4	54.3	48.8	56.1

Table 4. Overall Average Scores

Subject	Overall Average
Sindhi	45.2
English	44.6
Mathematics	47.7

The box plot is used to show the distributions of scores in each subject (Sindhi, English, and Mathematics), fig-1,2,3. Gender distributions are shown in fig-4. The pie chart (fig. 5) shows the proportion of different data types within a dataset, which is divided into two segments, representing the categories "int64" and "object". The blue segment labelled "int64", accounts for 57.1% of the data, while the red segment, labeled "object", accounts for 42.9% of the data. This indicates that a majority of the variables in the dataset are of integer type, while a smaller portion are of object type.

The yearly Distribution is shown in Figure 6 (the frequency of occurrences for each year from 2017 to 2023). The years are on the x-axis, and on the y-axis, we have taken the counts. The height of each bar corresponds to the number of occurrences for that particular year and reveals that the year 2019 had the highest frequency, followed by 2020 and 2018. The years 2022 and 2023 had the lowest frequencies, with 2021 falling somewhere in between with respect to the performance of the respondent students. Figure 7 shows the histogram distribution of three subjects among five classes. The x-axis represents the score (0-100), while the y-axis represents the class labels. Overall, it revealed that most students scored between 60 and 70 in Sindhi and English, with a smaller number of students scoring higher or lower. In Mathematics, the distribution is slightly skewed to the left, with a larger number of students scoring lower than 60.

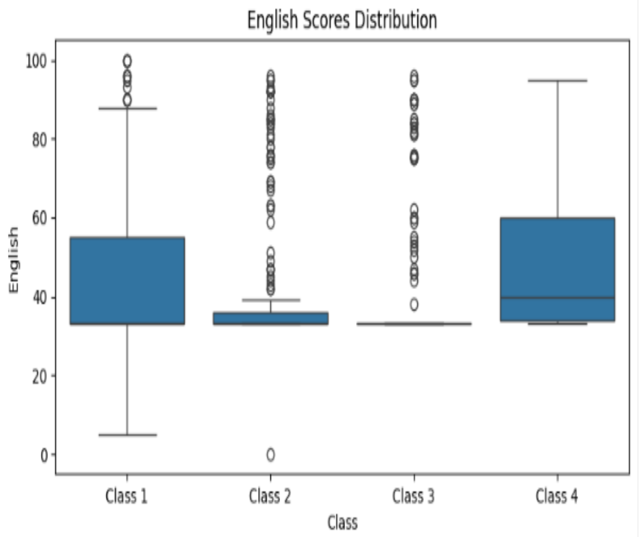

Figure 1. English Subject Score

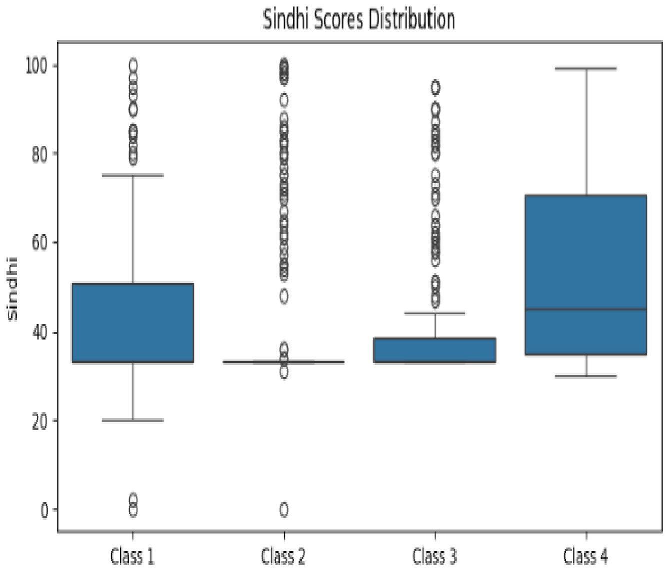

Figure 2. Sindhi Subject Score

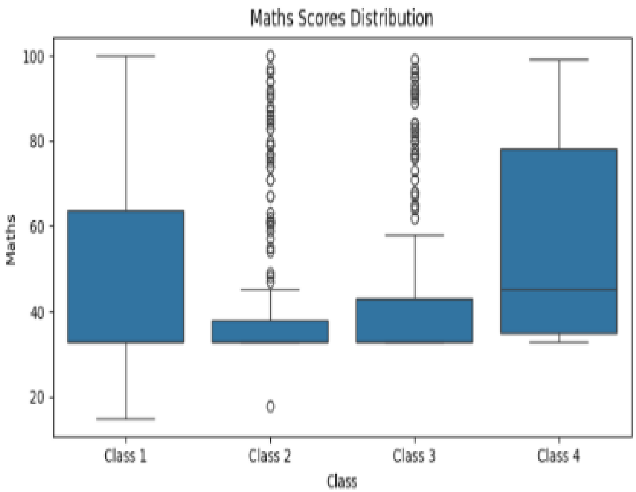

Figure 3. Maths Subject Score

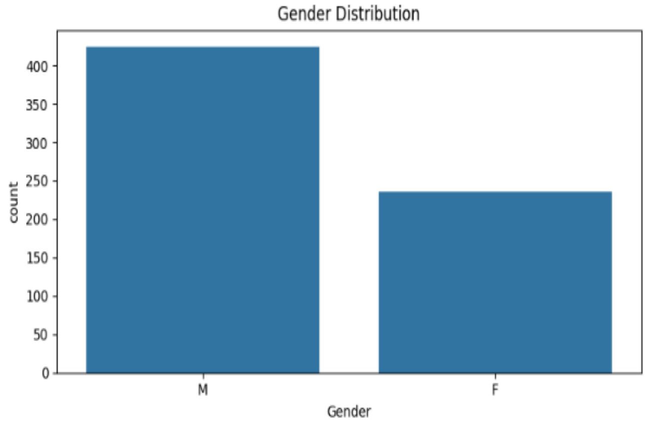

Figure 4. Gender Distribution

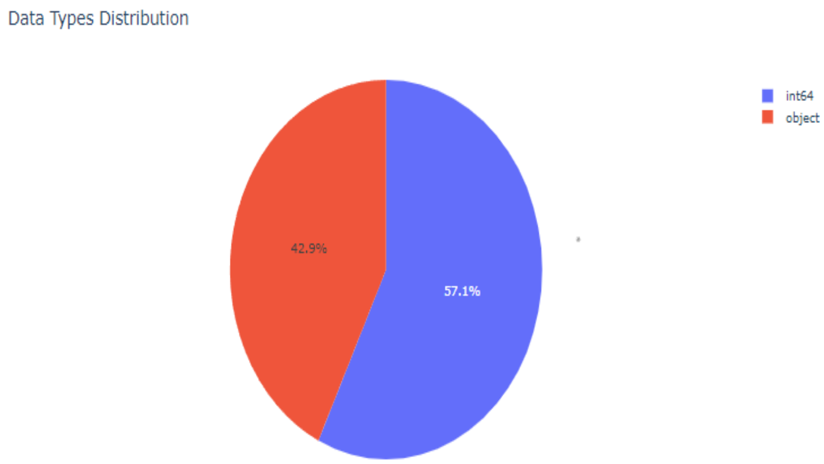

Figure 5. Yearly Distribution

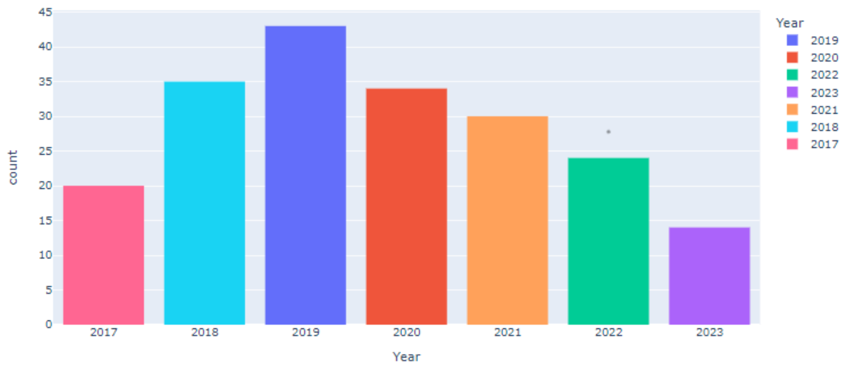

Figure 6. Histogram of three subjects

Histogram, located on the right side of the graph, shows the frequency of students achieving scores within each range for all classes combined. The dot plot, located on the left side of the graph, displays the individual Sindhi scores for each student in each class. Overall, the graph suggests that Class 1 has the highest number of students with high Sindhi scores, while Class 5 has the highest number of students with low Sindhi scores. The distribution of scores within each class can be observed in the dot plot, with more concentrated dots indicating a higher frequency of scores within that range for the

2.4.2. Subject-wise Analysis:

The subject-wise analyses depict the distribution of scores in each subject. The additive score of Sindhi, English and Mathematics was stored in three different rows for each student to get an overall idea of the student's knowledge and abilities in various areas.

2.4.3. Data Transformation:

Data transformation was done for the machine learning model training data set was converted into a dimensional array of lists, and the 'Gender' column was transformed to the desired form (0 for boy and 1 for girl). Splitting the listed values from the original dataset into a separate new dataset consisting of a single thread with a span of 3 rows in the working place.

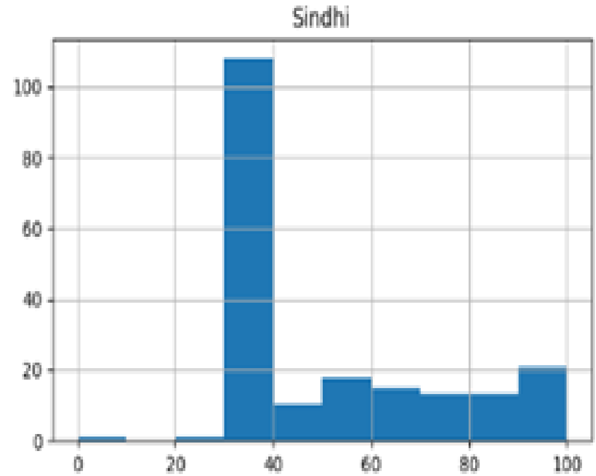

Figure 7. Sindhi Graph

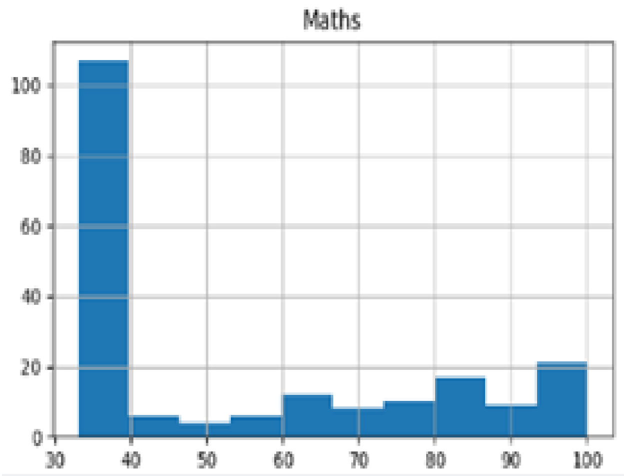

Figure 8. Maths Graph

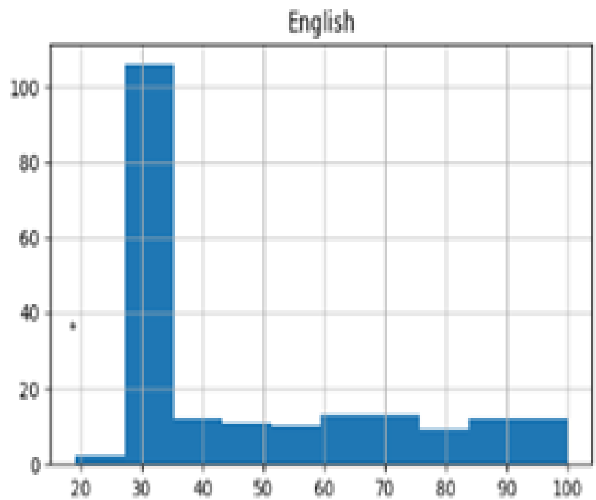

Figure 9. English Graph

2.4.4. Model Training and Evaluation:

Multi-output regressor model was adopted to train the MLP Regressor for the target merit and gender. The model's error was measured on the test set, too. The support vector regressor (SVR) and a decision tree (DT) regressor were implemented in parallel to compare among themselves and to find out which one possesses higher performance.

2.4.5. Predictions and Analysis:

The trained model has been used for prediction, and its performance has been evaluated by using random values from the data.

3. Results

The MLP Regressor model was trained with a configuration of hidden layers set to (100, 50) and a maximum of 1000 iterations. After training, the model would be able to make predictions. The Support Vector Regressor (SVR) model, implemented as a Multioutput Regressor with a linear kernel, demonstrated a higher MSE on the test set, measuring 687.54. Predictions on new data yielded scores of Sindhi 46.41, English 36.35, Mathematics 48.02, and Gender 0.63. The 'Gender' prediction is closer to a binary value but still deviates from the expected 0 or 1. The Decision Tree Regressor model, trained with default parameters, displayed the lowest MSE on the test set, measuring 62.38. However, predictions on new data resulted in Sindhi, English, and Mathematics scores all set to 33.31

Academic performance serves as a crucial indicator for evaluating students' learning progress and overall educational outcomes. Regression models, traditionally employed to examine relationships among variables, continue to provide valuable baselines for predictive analysis. However, with the advancement of machine learning (ML), more sophisticated models now enable the identification of hidden patterns in educational datasets, thereby enhancing predictive accuracy and offering actionable insights for stakeholders.

In this study, multiple regression-based ML algorithms were implemented and comparatively analyzed. The Decision Tree Regressor yielded a relatively high Mean Squared Error (MSE) of 62.38, which can be attributed to the inherent non-linear behavior of academic outcomes and the model's tendency toward overfitting when applied to small or heterogeneous datasets. In contrast, the Random Forest Regressor demonstrated a more balanced performance with an MSE of 25.21, reflecting its ability to aggregate multiple decision trees and reduce variance, thereby achieving a trade-off between accuracy and flexibility. This suggests that ensemble approaches are particularly well-suited for addressing the complexities of educational data.

The K-Nearest Neighbors (KNN) Regressor offered localized predictions, effectively capturing trends specific to subgroups of students and providing potential for personalized educational insights. Such characteristics highlight the model's relevance in contexts where learner diversity and subgroup behaviors strongly influence outcomes. On the other hand, models such as the Multi-Layer Perceptron (MLP) Regressor and Support Vector Regression (SVR) encountered challenges in handling categorical or binary attributes (e.g., gender), which limited their predictive effectiveness in this study.

The Decision Tree Regressor, though capable of capturing non-linear patterns, recorded a relatively high Mean Squared Error (MSE = 62.38 for females, 62.71 for males), indicating susceptibility to overfitting. In contrast, the Random Forest Regressor achieved lower MSE values (25.21 and 25.14), demonstrating the advantage of ensemble methods in improving generalization. The KNN Regressor provided competitive results (MSE = 19.29 and 19.36), highlighting its suitability for subgroup-specific predictions. Linear Regression remained the most effective baseline (MSE = 1.33 and 1.47).

Table 5. Performance of different machine learning algorithms

Model	Gender	Sindhi	English	Maths	MSE
Linear Regression	Female	60.6	70.2	80.9	1.33
	Male	60.1	70.4	80.3	1.47
Decision Tree Regressor	Female	51.4	60.2	80.3	62.38

	Male	51.5	60.3	80.2	62.71
Random Forest Regressor	Female	60.7	64.1	77.7	25.21
	Male	59.2	64.3	78.2	25.14
K-Neighbors Regressor	Female	60.4	63.4	73.1	19.29
	Male	60.4	63.4	73.2	19.36

Supplementary models confirmed these findings: the MLP Regressor showed higher error despite its complexity, while the SVR struggled with categorical predictors. Table 5 consolidates these outcomes, illustrating that simpler regression models, particularly Linear Regression, outperformed more complex algorithms in this dataset.

Overall, while the Decision Tree Regressor exhibited the highest error rate, it still provided moderate forecasting capability when applied to new test data, underscoring its potential for improvement through parameter tuning or hybridization with ensemble methods. These findings collectively emphasize that no single model is universally optimal; instead, the choice of algorithm must be context-driven, balancing interpretability, accuracy, and the specific characteristics of the dataset under consideration. The comparative results, summarized in Table 5, provide a nuanced perspective on the strengths and limitations of different regression approaches for predicting primary mathematics performance.

4. Conclusions

This study examined the effectiveness of regression-based machine learning models in predicting academic performance, with a particular focus on primary mathematics education. Four algorithms, Linear Regression, Decision Tree Regression, Random Forest Regression, and K-Nearest Neighbours Regression, were implemented and evaluated using Mean Squared Error (MSE) as the primary performance metric. The results revealed that Linear Regression achieved the lowest MSE (1.33 for females and 1.47 for males), establishing a strong predictive baseline under the assumption of linear relationships. The Decision Tree Regressor, while capable of capturing non-linear patterns, exhibited a substantially higher MSE of 62.38 (62.71 for males), underscoring the model's tendency to overfit in small or heterogeneous datasets. This result contrasts with the Random Forest Regressor, which achieved a lower MSE of 25.21, demonstrating improved generalization through ensemble averaging. Meanwhile, the K-Nearest Neighbours Regressor offered localized predictive insights with competitive MSE scores (19.29 and 19.36), highlighting its potential for subgroup-specific or personalized academic forecasting. The findings provide valuable guidance for educators and policymakers, demonstrating that regression models can complement traditional assessment methods by offering predictive insights into student achievement. Importantly, the results emphasize that the selection of predictive models should be context-sensitive, balancing interpretability, accuracy, and the complexity of educational datasets.

In conclusion, this research highlights the potential of regression-based machine learning models as robust tools for forecasting academic outcomes in primary education. By producing fine-grained predictions, these approaches can inform evidence-based decision-making, support early interventions, and guide resource allocation toward equitable and effective educational systems.

Nonetheless, limitations must be acknowledged. The dataset, while feature-rich, was relatively modest in size (N = 661) and drawn from a single regional context, which may constrain generalizability. Moreover, socio-economic and attendance indicators were partly self-reported and may not fully capture underlying complexities. Future work should adopt larger, more diverse datasets, incorporate longitudinal designs, and integrate additional behavioral and contextual features to improve predictive validity.

Finally, ethical considerations must be central to the use of predictive analytics in education. The reliance on demographic data raises concerns about privacy and fairness, while gender imbalance in the dataset may introduce bias into predictions. Addressing these challenges will require both technical solutions, such as bias-mitigation techniques and balanced data sampling, and adherence to ethical guidelines in deploying machine learning for educational decision-making.

Data Availability Statement:

Primary data sets were gathered through a questionnaire from the government and private school-going children. The data set would be available upon request.

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

References

1. Rai, Rahul, Manoj Kumar Tiwari, Dmitry Ivanov, and Alexandre Dolgui. (2021). Machine learning in manufacturing and industry 4.0 applications." *International Journal of Production Research* 59(16), pp.4773-4778.
2. Canagareddy, D., Subarayadu, K., &Hurbungs, V. (2019). A machine learning model to predict the performance of university students. In Smart and Sustainable Engineering for Next Generation Applications: Proceeding of the Second International Conference on Emerging Trends in Electrical, Electronic and Communications Engineering (ELECOM 2018), November 28–30, 2018, Mauritius 2 (pp. 313-322).
3. Batool, S., Rashid, J., Nisar, M. W., Kim, J., Kwon, H. Y. (2022). Educational data mining to predict students' academic performance: A survey study. Education and Information Technologies.28(4), DOI: 10.1007/s10639-022-11152-y
4. Wakelam, E., Jeferies, A., Davey, N., & Sun, Y. (2020). The potential for student performance prediction in small cohorts with minimal available attributes. British Journal of Educational Technology, 51(2), 347–370.
5. Dwivedi, D.N., and Ghanashyama, M. (2024). Predictive Analytics in Educational Outcomes: Analyzing High School Students' Performance in Mathematics. In Adaptive Learning Technologies for Higher Education, pp. 293-316. IGI Global.
6. Hassan, S. M., Ali, F., Wasi, S., Husain, I., and Ashraf, S. N. (2019). Roman-Urdu News Headline Classification with IR Models using Machine Learning Algorithms. Indian Journal of Science and Technology, 12(35):1-9.
7. Husain, I., Zaidi, S.M.H., Khan, U., Ahmed, A. (2024). Sentiment Analysis Classification of ChatGPT Tweets Using Machine Learning and Deep Learning Algorithms. Journal of Computing & Biomedical Informatics. 8(1): 1-8.
8. Iqbal, S., Ahmed, S., Hassan, B., Abid, M. K., and Husain, I. (2024). Enhancing Lung Cancer Detection with Hybrid Machine Learning: Integrating Ant Colony Optimization. Technical Journal, University of Engineering and Technology (UET) Taxila. 29(3): 74-86.
9. Pallathadka, H., Wenda, A., Ramirez-Asís, E., Asís-López, M., Flores-Albornoz, J., Phasinam, K., (2023). Classification and prediction of student performance data using various machine learning algorithms. Materials Today: Proceedings, 80(3), 3782-3785. https://doi.org/10.1016/j.matpr.2021.07.382
10. Awan, S. A. Ali, S. A., Husain, I., Quraishi, B. H., Ashraf. S.M.A. (2021). International journal of circuits, systems and signal processing. 15(189). 7172-7176.
11. Husain, I., Khan, N.U., and Shaukat, S. S. (2015). Mathematical Techniques to Study the Spatial Pattern of Meloidogynejavanica in relation to soil fungi and soil characteristics in a tomato-grown field. Research Journal of Recent Sciences, 2277, Pp. 2502.
12. Baig, M. A., Sarmad, A. S., Kamlesh, K. K., Muneer, A. S., Muhammad, Z. K., and Mahira, A. R. (2023). Prediction of Students Performance Level Using Integrated Approach of ML Algorithms. Int. J. Emerg. Technol. Learn. 18(1): 216-234.
13. Hassan, S.M., Ali, S. A., Hassan, B., Husain, I., Rafiq, M., and Awan, S.A. (2022). Hybrid Features Binary Classification of Imbalance Stroke Patients Using Different Machine Learning Algorithms. International Journal of Biology and Biomedical Engineering, 16(1): 154-160.
14. Shaukat, S. S., Husain, I., Siddiqui, I. A. (2005). An application of correspondence analysis, detrended correspondence analysis and canonical correspondence analysis with reference to the vegetation and environment of calcareous hills around Karachi. International Journal of Biology and Biotechnology. 2(3): 617-627.
15. Chen, S., and Yuanzhao, D. (2023). A machine learning approach to predicting academic performance in Pennsylvania's schools. Social Sciences 12(3): 118.
16. Hassan, S. M., Alghamdi, A., Hafeez, A., Hamdi, M., Husain, I., and Alrizq, M. (2021). An Effective Combination of Textures and Wavelet Features for Facial Expression Recognition. Engineering, Technology & Applied Science Research. 11(3): 7172-7176.
17. Sternberg, R. J. (2000). The concept of intelligence. Handbook of intelligence (2000): 3-15.
18. Segers, M., Filip, D., and Eduardo, C., (2006). Optimising new modes of assessment: In search of qualities and standards. Vol. 1. Springer Science & Business Media.
19. Boekaerts, M., and Eduardo, C., (2006). How far have we moved toward the integration of theory and practice in self-regulation? Educational psychology review 18, pp. 199-210.
20. Brauer, F., Castillo, C., and Zhilan, F. (2019). Mathematical models in epidemiology. Vol. 32. New York: Springer
21. Hussain, S., and Khan, M. Q. (2021). Student performulator: Predicting students' academic performance at secondary and intermediate levels using machine learning.10(5), Annals of Data Science. https://doi.org/10.1007/s40745-021-00341-0

22. Tomasevic, N., Gvozdenovic, N., & Vranes, S. (2020). An overview and comparison of supervised data mining techniques for student exam performance prediction. Computers & Education. 143, https://doi.org/10.1016/j.compedu.2019.103676
23. Sekeroglu, B., Dimmililer, K., & Tuncal, Kubra (2019). Student Performance Prediction and Classification Using Machine Learning Algorithms. The 8th International Conference, https://doi.org/10.1145/3318396.3318419
24. Sokkhey, P, and Takeo, O. (2019). Comparative study of prediction models on high school student performance in mathematics. In 2019 34th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC), pp. 1-4. IEEE.