Student Score Prediction Using Regression Based Machine Learning Models

Istyuk Ahamed
Department of Computer Science and Engineering,
Bangladesh University of Business and Technology, Dhaka, Bangladesh.
Email: istyukahamed@gmail.com,sudiptochakibd@gmail.com

***Abstract*—Student academic performance prediction has emerged as a crucial task in educational data mining, aimed at identifying potential areas for intervention and improving learning outcomes. This research focuses on the development of a regression-based system to predict students' final grades, leveraging a dataset comprising demographic, academic, and behavioral attributes. The study begins with extensive exploratory data analysis (EDA) to uncover patterns and relationships between features. Techniques such as outlier detection, correlation heatmaps, and categorical data encoding were employed to preprocess the data effectively. Class imbalance, a common challenge in educational datasets, was addressed using oversampling techniques. Multiple regression models, including Linear Regression, Decision Tree Regression, Random Forest Regression, and Support Vector Regression (SVR), were trained and evaluated based on Mean Squared Error (MSE) and R-squared metrics. Comparative analysis revealed Random Forest Regression as the best-performing model, achieving superior predictive accuracy and robustness. This work underscores the potential of advanced regression techniques in educational data mining and offers a foundation for the deployment of intelligent systems to aid academic planning and decision-making.**

***Keywords*—Performance prediction, Machine Learning, Linear Regression, Decision Tree Regression, Random Forest Regression, Support Vector Regression.**

I. Introduction

The present study, Predicting student performance has emerged as an essential tool in the present study for improving learning outcomes and offering timely support. It helps educational institutions identify students facing academic challenges and implement targeted interventions to improve their outcomes. The increased availability of educational data has opened up the use of advanced analytical techniques, especially machine learning (ML), for insights and to predict student performance. ML algorithms provide unparalleled capabilities for analyzing vast amounts of data for hidden patterns and accurate predictions using historical and real-time data. This research studies the implementation of regression-based machine learning algorithms to forecast student grades. Regression is one of the majorly applied models in predictive analytics in cases with continuous outcomes like academic performances. These models predict future performance by analyzing the patterns in past academic results and related features. Academic performance prediction is a very valuable thing because it could help schools optimize resource allocation, guide educators in making data-driven decisions, and provide students with timely and targeted assistance. Academic performance prediction is an area of growing importance within educational data mining research, impelled by the urge to improve learning outcomes and academic interventions further. The following research is based on a wide range of academic and personal information provided by Portuguese secondary school students to predict final grades. Such factors include intrinsic ones like motivation and study habits, but also extrinsic factors like support from parents and the environment at school. The three main reasons for the use of multiple regression models in this research are threefold: assessing different regression techniques in terms of their suitability and accuracy, determining major predictors of academic success, and developing of recommendations that could help in bringing forth improved academic performance among students. These findings do not only show the potential of machine learning in education but also contribute to the overall development of using data-driven approaches for the benefit of educators, policymakers, and learners alike.

II. Related Research

There are several Student Score prediction-related papers, In [1], N. Singh et al. developed a system to predict student performance using a hybrid cross-validation machine learning approach that combines repeated splits of random train tests and k-fold stratified validation methods. The study's contributions include proposing a hybrid validation model to enhance prediction accuracy for student performance in sentiment analysis, benchmarking it against four traditional validation techniques, and demonstrating improved results through comparative analysis. However, the paper provides limited details on handling data imbalance and feature engineering. The authors presented result analyses showing that the hybrid model outperformed traditional methods. The study concludes that hybrid cross-validation enhances prediction accuracy, offering a foundation for future research in student performance analysis using advanced machine learning techniques.

In [2], R. Ghorbani and R. Ghousi conducted a study comparing resampling methods for predicting student performance using machine learning techniques. Their contributions include a detailed analysis of various resampling methods (e.g., SVM-SMOTE, SMOTE-ENN, SMOTE-Tomek), evaluating their effectiveness in handling imbalanced educational datasets. The study compares classifiers such as Random Forest, XGBoost, and Support Vector Machine (SVM) and employs multiple validation techniques (Random Hold-Out and Shuffle 5-fold Cross-Validation) to enhance the robustness of the results. However, the study does not extensively address feature selection techniques or alternative classifier evaluation methods. The authors' analysis indicates that SVM-SMOTE outperformed other resampling techniques, while Random Forest achieved the best overall performance. The study concludes that the combination of SVM-SMOTE and Random Forest is highly effective for student performance prediction, providing a foundation for early interventions in educational settings.

In [3], M. Adnan et al. developed a machine learning-based predictive model to identify at-risk students at different stages of a course for early intervention. The key contributions of their study include handling missing data, performing feature engineering, and comparing various algorithms, with Random Forest (RF) demonstrating the best performance. The model enables early prediction of student outcomes, allowing instructors to intervene and reduce dropouts. However, the study lacks mechanisms for providing personalized feedback. Overall, the Random Forest model proved effective for early detection, contributing to improved student retention and performance.

In [4] A. Nabil et al. developed a Predictive Model for Students' Academic Performance Using Deep Neural Networks (DNNs). The main contributions of the paper include: addressing the Imbalanced Dataset Problem, applying Resampling Methods (such as SMOTE and ADASYN), comparing Traditional Machine Learning Models like Decision Tree and Random Forest with DNN, and achieving early detection of at-risk students in a Data Structures course. The proposed DNN model outperformed other methods with an accuracy of 89. The authors highlight the importance of predicting student outcomes early to improve their academic success.However, the paper lacks exploration of additional non-academic factors such as behavioral or demographic attributes, which could potentially enhance prediction accuracy. The result analysis shows that while Random Forest and Support Vector Classifiers excelled in accuracy on imbalanced data, DNN performed better in F1-score after handling the imbalance.The DNN model, supported by resampling techniques to address data imbalance, proves effective in early detection of at-risk students, providing educators with timely insights to intervene and support student success. Future work includes extending the dataset and testing other resampling methods for improved results.

In [5] P. Asthana et al. [2023] developed Predictive Modeling for Student Performance Using Learning Coefficients with Regression-Based Machine Learning Models. The contributions of this paper include the introduction of Learning Coefficients as dynamic predictors, the use of Adaptive Assessment, and a comparative analysis of regression models like Linear Regression (LR), Random Forest (RF), Decision Tree (DT), and Support Vector Regression (SVR). The Linear Regression model achieved the highest prediction accuracy of 97. However, the authors did not explore complex deep learning models due to the small dataset size. The result analysis shows LR outperformed other models in accuracy and error metrics such as RMSE and MAE.The study concludes that learning coefficients, combined with traditional academic predictors, significantly enhance the prediction of student performance. With the highest accuracy from the linear regression model, this approach can help improve educational outcomes by identifying at-risk students early. The study suggests expanding the dataset for future work to improve model performance and include more advanced machine learning techniques.

In [6] Prabowo et al. developed a model for predicting university students' GPA by combining both time-series and tabular data within a deep learning framework. Their main contributions are as follows: incorporating both time-series (historical GPA) and tabular data (student background information), using a dual-input model with MLP and LSTM for effective GPA prediction, and demonstrating superior accuracy compared to traditional models. Limitations not fully addressed include challenges in handling non-smooth GPA distributions and difficulties with long-range dependencies in the time-series data. The authors present a detailed performance analysis, showing that the proposed model achieves the lowest error metrics, indicating its effectiveness.The study validates that combining time-series and tabular data using the MLP-LSTM model significantly enhances the accuracy of GPA predictions, supporting targeted academic interventions. Future improvements suggested include exploring non-parametric and Transformer models for better handling of varied data distributions and longer dependencies.

In [7] Aljaloud et al. developed a CNN-LSTM deep learning model to predict student learning outcomes within a Learning Management System (LMS). The contributions of this paper include the integration of CNN for feature extraction, LSTM for time-series data analysis, evaluation of LMS-based student performance KPIs, benchmarking against CNN, RNN, and CNN-RNN models, and practical insights for enhancing LMS use in educational settings. However, the paper does not fully address potential feature selection optimizations or alternatives to reduce training time. The authors present performance evaluations showing the CNN-LSTM model's high prediction accuracy of 94.3 and significantly lower error metrics compared to other models.The study demonstrates that combining CNN and LSTM can effectively predict student outcomes in LMS by analyzing behavioral data, with implications for targeted educational interventions. Future work may explore lightweight models to reduce training time without compromising accuracy.

In [8] G. Feng et al. developed a model for analyzing and predicting students' academic performance using educational data mining techniques. The contributions of this paper include enhancing the K-means clustering algorithm for objective clustering-number determination, integrating clustering with CNNs for prediction, using clustering labels to train CNN models on unlabeled educational data, and applying discriminant analysis for model validation. However, the authors did not address potential improvements in handling outliers and optimizing initial clustering centers. The authors present extensive validation and performance testing, demonstrating the model's reliability in academic performance prediction with high accuracy.The study shows that combining clustering and

CNNs effectively predicts student performance, enabling early academic interventions. Future research might focus on refining clustering methods for more accurate performance analysis and exploring hybrid models to enhance prediction efficiency.

In [9]Ahajjam et al. developed a model to predict students' baccalaureate performance using artificial neural networks, focusing on Moroccan high school students. Key contributions include a predictive system for student performance, data imputation and preprocessing techniques, a comparison of machine learning models, and an educational decision support tool. However, the paper lacks discussion on the real-world impact of these predictions on educational outcomes. The authors' result analysis shows that neural networks achieved high predictive accuracy across literature, science, and technical branches. In conclusion, the study demonstrates that machine learning can effectively guide educational pathways, potentially improving student success in the Guelmim Oued Noun region.

In [10]Alhazmi and Sheneamer developed a framework for Early Predicting of Students' Performance in Higher Education using machine learning. The paper's main contributions include a predictive system integrating admission scores, early course grades, academic achievement tests, and general aptitude tests. They use t-SNE for dimensionality reduction to visualize relationships between predictors and GPA and evaluate several machine learning models, including XGBoost and Random Forest, for prediction accuracy. The study finds that combining admission scores with all first-level course scores yields the best results. Although the authors lack a discussion on implementation challenges, the research highlights the potential for early identification of at-risk students. Conclusively, this model offers a promising tool for educational institutions to enhance student success through timely interventions.

In [11]D. Uliyan et al. developed a Deep Learning Model to Predict Student Retention using Bidirectional Long Short-Term Memory (BLSTM) and Conditional Random Fields (CRF). The contributions include the development of a BLSTM-CRF hybrid model, analysis of demographic and academic features for retention prediction, and comparison with traditional predictive methods such as logistic regression and decision trees. However, further details on data preprocessing and specific feature engineering methods were not extensively discussed. The authors presented experimental results showing that the BLSTM-CRF model achieved an accuracy above 85,outperforming conventional models on student retention prediction. The study concludes that deep learning models like BLSTM-CRF can effectively predict student retention, providing valuable insights for early intervention strategies in higher education institutions.

In [12]Khan et al. developed a model to predict student performance using machine learning on LMS activity logs. Key contributions include a cross-program prediction methodology, analysis of student activities like "Previous Semester GPA" and "Sessional Active Days," and the use of the CR Tree model to identify patterns for accurate predictions. The study effectively highlights LMS logs and academic performance as predictors for student outcomes, helping educators identify students needing support. However, it lacks analysis of non-academic influences. In conclusion, this model enables early intervention by predicting student performance across courses, with future work focused on adding social factors and refining early alert systems.

In [13]A. Hassan et al. developed a framework for detecting at-risk students in online courses using machine learning. Key contributions include early identification of students at risk of withdrawal or failure, achieved through predictive models focused on at-risk status and learning achievements. Machine learning algorithms like Gradient Boosting Machine (GBM), Random Forest (RF), and Neural Networks were utilized, with GBM showing the highest accuracy (0.894 and 0.952). The study also emphasizes feature selection for improving predictive power and reducing computational costs. The models aim to help educators implement timely interventions for at-risk students, although potential limitations in diverse educational contexts and long-term intervention effectiveness were not fully addressed. In conclusion, this research underscores the importance of machine learning in supporting educators to monitor and support student engagement, aiming to lower dropout rates in online learning environments.

In [14]C. Zhang et al. developed a predictive model to analyze learning behaviors and predict the learning effects of information literacy among college students using machine learning techniques. Key contributions include a data-driven methodology to better understand learning behaviors, employing supervised classification algorithms like Decision Tree, KNN, Naive Bayes, Neural Net, and Random Forest, with Random Forest showing superior performance. The model achieved high accuracy (92.50), precision (84.56), recall (94.81), F1-Score (89.39), and a kappa coefficient of 0.859, affirming its effectiveness. Additionally, the study recommends differentiated teaching strategies to enhance instructional effectiveness in information literacy. However, external factors influencing learning outcomes were not considered, which could improve model applicability. The findings underscore the role of machine learning in supporting optimized teaching and resource allocation in education.

In [15]A. Hassan et al. developed "Assessing Intervention Timing in Computer-Based Education Using Machine Learning," focusing on the impact of process-level data (e.g., daily quiz scores) on predictive performance. Key contributions include the finding that while process-level data does not significantly improve post-hoc predictions, it enhances earlier predictions, aiding timely interventions. The authors compare simpler models (linear and logistic regression) with complex models (SVM and neural networks), showing that simpler models perform comparably well, especially in small datasets. The study emphasizes prediction accuracy metrics and addresses the challenge of high feature-to-data point ratios by grouping students for training and testing. However, it does not explore potential limitations of relying solely on traditional assessments. The authors conclude that simpler models are effective for educational predictions, and that process-level data is valuable for early interventions, suggesting further validation on larger datasets.

In [16]A. Hassan et al. developed a machine learning model to predict students' native places based on technological awareness and demographic features. Key contributions include a hybrid predictive approach that integrates machine learning with optimization techniques, enhancing prediction accuracy. The study employed three optimization algorithms (Adam, SGD, and LBFGS) to tune a Multi-Layer Perceptron (MLP) model, which was compared against a Support Vector Machine (SVM). Feature selection using Principal Component Analysis (PCA) further improved model accuracy, with both models achieving 94 accuracy in predictions. Statistical validation, including a paired t-test, confirmed the models' predictive strength. However, the study did not explore advanced machine learning algorithms or datasets from multiple countries, which could enhance robustness. In conclusion, this research advances educational data mining by using optimized machine learning models to provide insights into student demographics, suggesting future work with larger datasets and additional algorithms for broader applicability.

In [17]A. Author et al. developed a model to predict students' academic performance in online live classrooms using natural language processing and deep learning. Key contributions include analyzing large-scale classroom dialogue data, focusing on emotional expressions and interaction types. The authors developed neural network models to distinguish high and low academic performers based on dialogue features, using SHAP (Shapley additive explanations) for model interpretability. The study highlights critical predictors, including pretest rank for both STEM and non-STEM courses, while noting that predictors vary by course type. Findings indicate that high-performing students show more positive emotions and cognitive dialogue. However, the study does not consider external factors like socio-economic status. Results suggest that interaction types are more influential than emotional expressions, especially in lesson summary phases. In conclusion, this work underscores the role of classroom dialogue in academic performance prediction, suggesting that interpretable AI can aid in developing targeted educational support systems, with potential for further exploration of additional performance-affecting factors.

In [18]A. Hassan et al. conducted a systematic literature review on "Imbalanced Classification Methods for Student Grade Prediction," providing a comprehensive analysis of techniques used to address class imbalance in predicting student grades. Key contributions include a review of data-level, algorithm-level, and hybrid-level approaches, highlighting challenges in educational data contexts and the importance of predictive models for identifying at-risk students. The authors summarize methods such as sampling, feature selection, and cost-sensitive learning, and suggest future directions to enhance diversity and effectiveness in handling imbalanced datasets. However, practical case studies were not included, limiting contextual applicability. The analysis reveals that data-level techniques like SMOTE are popular, yet hybrid and feature selection methods remain underutilized. In conclusion, this review emphasizes imbalanced classification's role in improving predictive accuracy, guiding future research toward more effective models for educational outcomes.

In [19]A. Hassan et al. developed a DeepFM-based predictive model to identify student dropout risks in online classes, combining factorization machines with deep neural networks for enhanced prediction accuracy. Key contributions include the use of the HarvardX Person-Course and MOOC datasets, which strengthened model performance, achieving 99 accuracy on validation data across metrics like precision, recall, F1 score, and AUC-ROC. The DeepFM model captures complex, non-linear relationships between features, offering deeper insights into dropout patterns that traditional models may miss. The model is intended to provide actionable insights for educational interventions, aiming to support at-risk students and improve retention. However, limitations related to generalizability across platforms and demographic diversity are not addressed. In conclusion, this study emphasizes the potential of advanced machine learning techniques in online education, contributing a robust framework for early dropout prediction and supporting targeted retention strategies to enhance student success.

In [20]A. Hassan et al. developed a model for predicting and interpreting student performance using ensemble methods and Shapley Additive Explanations (SHAP). Key contributions include data preprocessing with SMOTE to address class imbalance, and the use of ensemble models, XGBoost and Extra Trees (ET), to enhance predictive accuracy. Hyperparameter tuning via grid search further improved model performance, resulting in over 98 accuracy. SHAP values were introduced to explain the model predictions, providing insights into factors influencing student performance. While limitations regarding model generalizability across different contexts were not addressed, the study demonstrates the potential of ensemble models and interpretable AI in supporting informed educational decisions. This research lays a foundation for future work on transparent and reliable predictive models in education.

III. Methodology

This research employs a systematic approach to predict student grades (G3) using regression techniques. The dataset, which includes 649 student records and 33 attributes, was first preprocessed to handle categorical variables, outliers, and class imbalance. Exploratory Data Analysis (EDA) was performed to gain insights into feature distributions, identify correlations, and understand relationships between predictors and the target variable. After preprocessing, multiple regression models were implemented, including Linear Regression, Decision Tree Regression, Random Forest Regression, and Support Vector Regression. To ensure fairness, the dataset was balanced using oversampling techniques, and categorical variables were encoded for numerical compatibility. Model performance was evaluated based on Mean Squared Error (MSE) and R-squared metrics. Visualizations comparing actual and predicted grades, as well as model-specific evaluations, were used to identify the best-performing regression technique. This method ensures a comprehensive analysis of student

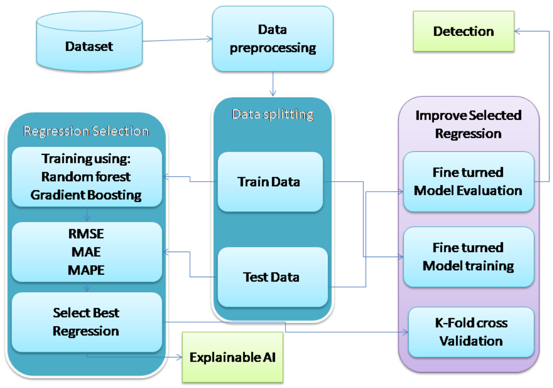

Fig. 1: Methodology for student score prediction

performance, bridging theoretical and practical applications of regression in educational settings.

Dataset Preparation: The process begins with collecting and loading a dataset containing various features, such as student demographics, academic performance, and behavioral attributes. This raw data undergoes preprocessing to handle missing values, normalize numerical data, and encode categorical variables for compatibility with machine learning algorithms.

Data Splitting: The preprocessed data is then split into two subsets: training data and test data. The training dataset is used to build and optimize regression models, while the test dataset is reserved for evaluating the model's generalization performance.

Regression Selection: Multiple regression models are trained and evaluated to predict student performance. Techniques such as Random Forest Regression and Gradient Boosting are applied to the training data. These models are assessed using key performance metrics, including Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). Based on these metrics, the best-performing regression model is selected.

Fine-Tuning and Validation: The selected regression model undergoes fine-tuning to improve its predictive performance. Hyperparameters are adjusted, and advanced techniques like K-Fold Cross-Validation are used to ensure the model's robustness. Cross-validation helps mitigate overfitting by testing the model on different subsets of data.

Explainability: To ensure the results are interpretable, Explainable AI (XAI) techniques are employed. This allows educators and stakeholders to understand how various features contribute to student performance predictions, enabling data-driven decisions.

Evaluation and Deployment: Finally, the fine-tuned model is evaluated on the test data to measure its accuracy and reliability. Once validated, the model can be deployed for real-world use, providing actionable insights into student performance and enabling early interventions for underperforming students.

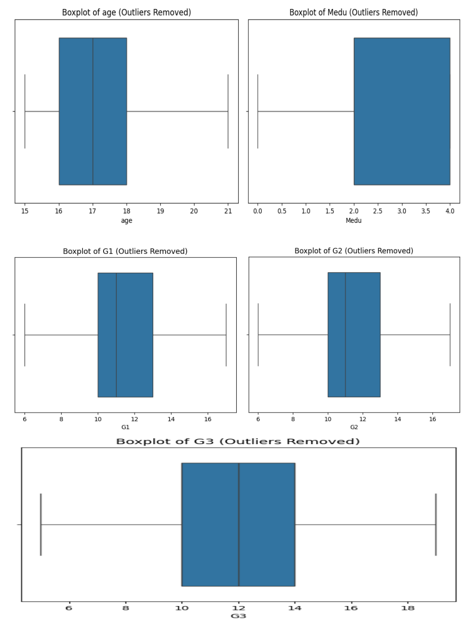

Fig. 2: Outlier removal

IV. Result Analysis

1. Overview of Data Preprocessing and Feature Handling Data preprocessing was conducted to ensure a clean, analyzable dataset. The primary steps included sorting, handling outliers, and balancing the dataset:

Outlier Detection and Removal: Using the IQR method, significant outliers in numerical features were detected and removed. For example, extreme age values were filtered to ensure reliable regression analysis. Class Balancing: The target variable, G3 (final grades), exhibited class imbalance. Oversampling using RandomOverSampler increased the minority class samples, ensuring an equitable model training process. This step prevented bias toward dominant classes during prediction. These preprocessing techniques enhanced data quality, leading to improved model performance.

2. Model Performance Analysis Linear Regression: The simplest model, Linear Regression, showed moderate performance with an R-squared value of 0.8800. The model assumes linearity in relationships between features, which might have limited its accuracy due to the dataset's nonlinear patterns.

Decision Tree Regression: This model captured nonlinear relationships, yielding better performance (R-squared = 0.9792). However, its tendency to overfit limited its generalizability.

Random Forest Regression: Random Forest achieved the best results (MSE = 0.3529, R-squared = 0.9882). Its ensemble nature combined predictions from multiple decision trees, reducing overfitting and improving accuracy. This model excelled in capturing complex patterns in the data.

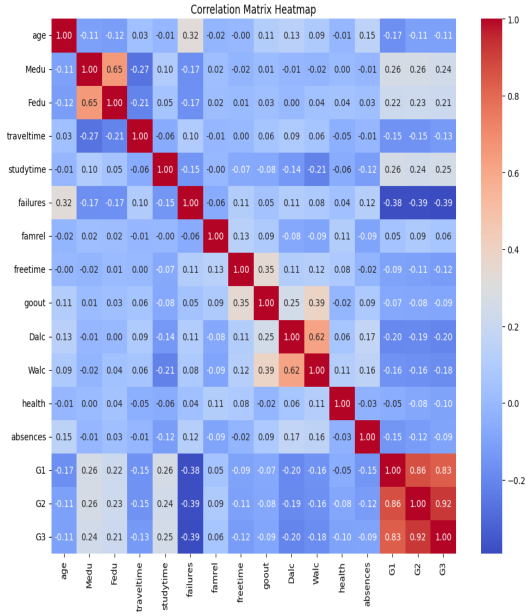

Fig. 3: Correlation matrix.

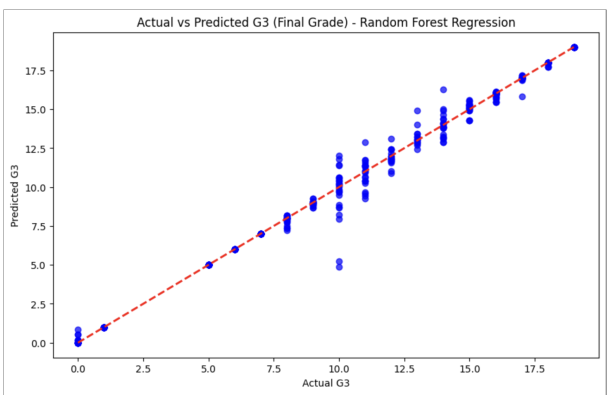

Fig. 4: scatter plot of Actual vs Predicted values

Support Vector Regression (SVR): SVR performed moderately well, with an R-squared of 0.8263. While effective in handling nonlinearity, the model struggled with scalability and tuning challenges given the dataset size.

3.Comparison of Model Performances The results demonstrate that Random Forest Regression outperformed the other models in both MSE and R-squared metrics. It effectively balances bias and variance, capturing intricate relationships within the dataset. In contrast, simpler models like Linear Regression fell short in handling nonlinear dependencies, while single-tree methods like Decision Tree Regression were prone to overfitting.

To visualize these differences, bar charts were generated for

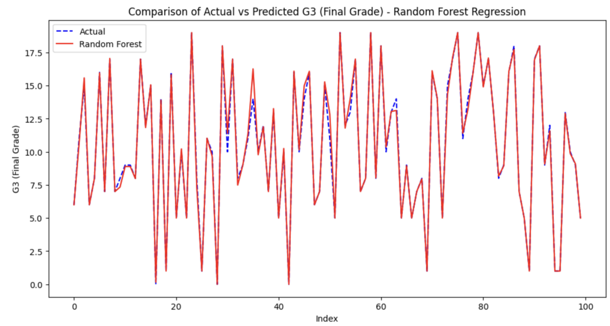

Fig. 5: line plot of Actual vs Predicted

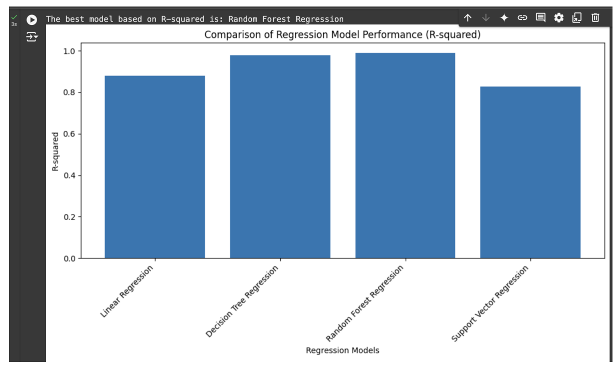

Fig. 6: Model Evaluation Results

R-squared and MSE scores:

R-squared Comparison:

Random Forest exhibited the highest score, indicating a robust fit to the test data. Decision Tree and SVR followed with moderate R-squared values, while Linear Regression scored the lowest. MSE Comparison:

Random Forest had the lowest MSE, confirming its superior predictive accuracy. Decision Tree and SVR had similar error rates, with Linear Regression being the least accurate.

article booktabs

Model	MSE	R-squared
Linear Regression	3.6152	0.8800
Decision Tree Regression	0.6342	0.9792
Random Forest Regression	0.3529	0.9882
Support Vector Regression	5.2348	0.8263

TABLE I: Comparison of regression model performance based on Mean Squared Error (MSE) and R-squared.

4.Model Insights and Practical Implications

The Random Forest model's superior performance can be attributed to its ability to:

Capture complex feature interactions through ensemble learning. Reduce overfitting by averaging multiple tree predictions. Handle mixed types of data (numerical and categorical) effectively. The insights from this study can guide educators and administrators in early identification of at-risk students. By accurately predicting final grades, tailored interventions can be designed to improve academic outcomes.

V. Conclusion

In this study, we applied various regression-based machine learning models—Linear Regression, Decision Tree Regression, Random Forest Regression, and Support Vector Regression—to predict student performance using a dataset from Portuguese secondary school students. The results demonstrated that Random Forest Regression was the most accurate model, providing the best prediction with the highest R-squared value and the lowest Mean Squared Error. Key factors influencing student success were identified, including prior grades and parental support. The research also highlighted the importance of proper data preprocessing, including outlier removal and class balancing, in improving model performance. These findings suggest that machine learning can significantly aid in predicting academic outcomes, allowing for timely interventions and better resource allocation in educational settings. Future research could expand on these results by incorporating additional data sources and exploring more advanced machine learning techniques for even greater accuracy and generalization.

Acknowledgment

We would like to acknowledge the support of the Bangladesh University of Business & Technology and the IoT lab for their suggestion and resource sharing.

References

[1] N. Singh and U. Chandra, “Prediction of student score performance of sentiment analysis using hybrid cross validation machine learning techniques,” 06 2023.

[2] R. Ghorbani and R. Ghousi, “Comparing different resampling methods in predicting students’ performance using machine learning techniques,” *IEEE Access*, vol. 8, pp. 67 899–67 911, 2020.

[3] M. Adnan, A. Habib, J. Ashraf, S. Mussadiq, A. A. Raza, M. Abid, M. Bashir, and S. U. Khan, “Predicting at-risk students at different percentages of course length for early intervention using machine learning models,” *IEEE Access*, vol. 9, pp. 7519–7539, 2021.

[4] A. Nabil, M. Seyam, and A. Abou-Elfetouh, “Prediction of students’ academic performance based on courses’ grades using deep neural networks,” *IEEE Access*, vol. 9, pp. 140 731–140 746, 2021.

[5] P. Asthana, S. Mishra, N. Gupta, M. Derawi, and A. Kumar, “Prediction of student’s performance with learning coefficients using regression based machine learning models,” *IEEE Access*, vol. 11, pp. 72 732–72 742, 2023.

[6] H. Prabowo, A. A. Hidayat, T. W. Cenggoro, R. Rahutomo, K. Purwandari, and B. Pardamean, “Aggregating time series and tabular data in deep learning model for university students’ gpa prediction,” *IEEE Access*, vol. 9, pp. 87 370–87 377, 2021.

[7] A. S. Aljaloud, D. M. Uliyan, A. Alkhalil, M. A. Elrhman, A. F. M. Alogali, Y. M. Altameemi, M. Altamimi, and P. Kwan, “A deep learning model to predict student learning outcomes in lms using cnn and lstm,” *IEEE Access*, vol. 10, pp. 85 255–85 265, 2022.

[8] G. Feng, M. Fan, and Y. Chen, “Analysis and prediction of students’ academic performance based on educational data mining,” *IEEE Access*, vol. 10, pp. 19 558–19 571, 2022.

[9] T. Ahajjam, M. Moutaib, H. Aissa, M. Azrour, Y. Farhaoui, and M. Fattah, “Predicting students’ final performance using artificial neural networks,” *Big Data Mining and Analytics*, vol. 5, no. 4, pp. 294–301, 2022.

[10] E. Alhazmi and A. Sheneamer, “Early predicting of students performance in higher education,” *IEEE Access*, vol. 11, pp. 27 579–27 589, 2023.

[11] D. Uliyan, A. S. Aljaloud, A. Alkhalil, H. S. A. Amer, M. A. E. A. Mohamed, and A. F. M. Alogali, “Deep learning model to predict students retention using blstm and crf,” *IEEE Access*, vol. 9, pp. 135 550–135 558, 2021.

[12] M. Khan, S. Naz, Y. Khan, M. Zafar, M. Khan, and G. Pau, “Utilizing machine learning models to predict student performance from lms activity logs,” *IEEE Access*, vol. 11, pp. 86 953–86 962, 2023.

[13] R. Al-Shabandar, A. J. Hussain, P. Liatsis, and R. Keight, “Detecting at-risk students with early interventions using machine learning techniques,” *IEEE Access*, vol. 7, pp. 149 464–149 478, 2019.

[14] Y. Shi, F. Sun, H. Zuo, and F. Peng, “Analysis of learning behavior characteristics and prediction of learning effect for improving college students’ information literacy based on machine learning,” *IEEE Access*, vol. 11, pp. 50 447–50 461, 2023.

[15] A. J. Stimpson and M. L. Cummings, “Assessing intervention timing in computer-based education using machine learning algorithms,” *IEEE Access*, vol. 2, pp. 78–87, 2014.

[16] C. Verma, V. Stoffová, Z. Illés, S. Tanwar, and N. Kumar, “Machine learning-based student’s native place identification for real-time,” *IEEE Access*, vol. 8, pp. 130 840–130 854, 2020.

[17] Y. Zhen, J.-D. Luo, and H. Chen, “Prediction of academic performance of students in online live classroom interactions—an analysis using natural language processing and deep learning methods,” *Journal of Social Computing*, vol. 4, no. 1, pp. 12–29, 2023.

[18] S. D. Abdul Bujang, A. Selamat, O. Krejcar, F. Mohamed, L. K. Cheng, P. C. Chiu, and H. Fujita, “Imbalanced classification methods for student grade prediction: A systematic literature review,” *IEEE Access*, vol. 11, pp. 1970–1989, 2023.

[19] N. M. Alruwais, “Deep fm-based predictive model for student dropout in online classes,” *IEEE Access*, vol. 11, pp. 96 954–96 970, 2023.

[20] H. Sahlaoui, E. A. A. Alaoui, A. Nayyar, S. Agoujil, and M. M. Jaber, “Predicting and interpreting student performance using ensemble models and shapley additive explanations,” *IEEE Access*, vol. 9, pp. 152 688–152 703, 2021.