An AI Based Interpretable Model to Evaluate the Impact of Socioeconomic and Academic Factors on Engineering Students' Performance

Meet Tilavat[1], Jyotir Patel[1], Rajeev Kumar Gupta[1] and Komal Singh[1]

[1] Department of Computer Science and Engineering, School of Technology, Pandit Deendayal Energy University, Gandhinagar 382007, Gujarat, India

Abstract. This research investigates the factors influencing academic performance in engineering programs using a dataset of 12,411 students from Colombia. The data includes academic results from standardized tests at the secondary and university levels, as well as social and economic background information. We employed various machine learning algorithms to predict student performance based on these factors and analyzed feature importance to identify key contributors to academic success. Our results indicate that Convolutional Neural Network, Gradient Boosting and Logistic Regression models achieved the highest accuracy (over 90%) in predicting student performance quartiles, followed by Random Forest and Decision Tree models. Analysis of feature importance revealed that university-level academic competencies, particularly Critical Reading (CR PRO) and Citizen Competencies (CC PRO), played a significant role in determining student success. Additionally, secondary-level academic performance and socioeconomic factors demonstrated varying degrees of influence across different models. This study provides valuable insights for educators and policymakers to develop targeted interventions and support systems to enhance student success in engineering education.

Keywords: Machine Learning, Academic Performance Prediction, Engineering Education, Educational Data Mining, Socioeconomic Factors in Education, Feature Importance Analysis, Predictive Analytics in Higher Education, Machine Learning Algorithms in Education

1 INTRODUCTION

The field of education is experiencing a transformative shift driven by the abundance of data and advancements in data analysis techniques. Educational data mining and learning analytics have emerged as powerful tools to analyze student performance, identify at-risk individuals, and develop personalized learning interventions. This research paper delves into the factors influencing academic performance in engineering students using a dataset of 12,411 students from Colombia.

The data encompasses academic results from standardized tests at both secondary and university levels, along with social and economic background information. Through the analysis of this comprehensive dataset, we aim to gain insights into the relationship between various factors and academic success in engineering programs.

This study is guided by the following research questions:

- **How do academic competencies at the secondary level influence performance in engineering programs?**
- **What is the impact of social and economic factors on academic achievement in engineering?**
- **Can we identify specific factors that significantly contribute to student success in different engineering disciplines?**

By analyzing a large-scale dataset with diverse variables, this research contributes to understanding academic performance within engineering education. The findings can inform educators, policymakers, and researchers in developing targeted interventions and support systems to enhance student success in engineering.

2 RELATED WORK

Predicting academic performance in engineering education has become a significant research focus, with numerous studies utilizing data mining and machine learning to pinpoint factors contributing to student success. This field has seen a surge in using learning analytics (LA) and educational data mining (EDM) to understand and predict student performance.

Early research focused on traditional machine learning models. Delahoz-Dominguez et al. (2020) [1] and Soto-Acevedo et al. (2023) [2] used standardized test scores and socioeconomic data to analyze and predict academic performance in Colombia, emphasizing the importance of a strong academic foundation established in secondary education. Similar approaches have been explored by other researchers, employing models like decision trees [4], random forests [5], support vector machines [6], and Naïve Bayes [7] to predict student outcomes based on demographic, academic, and social factors.

The use of Learning Management Systems (LMS) has provided new avenues for performance prediction. Khan et al. (2023) [3] investigated the predictive power of LMS activity logs, identifying factors like resource views, activity gaps, and previous academic performance as strong predictors of student success. This highlights the value of student engagement and historical data in assessing learning progress.

More recently, researchers have explored the application of deep learning models for improved predictive accuracy. Alhazmi and Sheneamer (2023) [8] used dimensionality reduction techniques and machine learning models to predict student performance at early stages, using admission scores and first-year course grades. Sun et al. (2023) [9] proposed a model based on multifeature fusion and attention mechanisms to analyze historical academic data from multiple dimensions, demonstrating the potential of this approach for accurate performance prediction. Convolutional Neural Networks (CNNs), often used for image recognition, have al-so been adapted for student performance prediction. Chau et al. (2021) [18] used a 2D CNN to analyze temporal educational data by transforming it into 2D images, showcasing the potential of CNNs to uncover patterns in non-image data. Poudyal et al. (2022) [17] further explored this approach by developing a hybrid 2D CNN model trained on 1D numerical data converted into 2D grayscale images, achieving promising results in predicting academic performance. Like our exploration of advanced models, Nabil et al. (2021) [20] investigated the effectiveness of Deep Neural Networks (DNNs) for predicting student performance. Their study focused on predicting student success in a Data Structures course based on grades from previous courses. They found that DNNs outperformed other machine learning models, achieving an accuracy of 89%. This highlights the potential of DNNs in handling complex academic performance data and identifying students at risk of failure.

The current research expands upon these efforts by leveraging a hybrid approach. We combine the feature extraction capabilities of CNNs with the sequential processing strength of Long Short-Term Memory (LSTM) networks, aiming to capture both complex relationships within data and temporal dependencies in student learning behavior. Furthermore, we delve into feature importance analysis to provide insights into the factors most significantly contributing to student success in engineering programs.

3 DATA DESCRIPTION

The data for this study was obtained from the Colombian Institute for the Evaluation of Education (ICFES) and compiled by Delahoz-Dominguez et al. (2020) [10]. It contains information on 12,411 engineering students and includes the following variables:

- **Academic Performance:**
    - **Secondary Level:** Scores on standardized tests (Saber 11) for Mathematics (MAT S11), Critical Reading (CR S11), Citizen Competencies (CC S11), Biology (BIO S11), and English (ENG S11).
    - **University Level:** Scores on standardized tests (Saber Pro) for Quantitative Reasoning (QR PRO), Critical Reading (CR PRO), Citizen Competencies (CC PRO), Written Communication (WC PRO), English (ENG PRO), and Formulation of Engineering Projects (FEP PRO). Global Score (G SC) is also provided.
- **Demographic Information:** Gender, socioeconomic level, and educational background of parents.
- **Social and Economic Factors:** Access to resources like internet, television, computers, and household appliances.
- **Educational Background:** Type of school attended (public or private), academic program enrolled in.

The data provides a rich source of information for exploring the factors impacting academic performance in engineering education.

4 METHODOLOGY

This study employs a quantitative approach to analyze the factors influencing academic performance in engineering students. We utilize various data preprocessing techniques, machine learning models, and evaluation metrics to achieve our research objectives.

4.1 Data Preprocessing:

1) **Categorical Encoding:** Categorical variables like gender, parents' education, and occupation are converted into numerical representations using Label Encoding.
2) **Target Variable Transformation:** The global score (G SC) is binned into quartiles (Q1, Q2, Q3, Q4) to create a categorical target variable for classification models. This allows us to predict students' performance levels rather than exact scores.
3) **Train-Test Split:** The data is split into training and testing sets with an 80/20 ratio. This ensures model evaluation on unseen data and prevents overfitting.

4.2 Model Selection:

We explore a range of classification models to predict student performance based on the binned global score:

- **Random Forest Classifier:** This ensemble learning method combines multiple decision trees to improve prediction accuracy and reduce overfitting.
- **K-Nearest Neighbors (KNN):** This non-parametric algorithm classifies data points based on the k nearest neighbors in the training data, considering the similarity between data points.
- **Gradient Boosting Classifier:** This ensemble learning technique combines weak learners sequentially to create a strong predictive model by focusing on correcting errors from previous models.
- **Logistic Regression with Cross-Validation (GLMNet):** This statistical model is suitable for binary and multiclass classification problems and is particularly useful when interpreting feature importance.
- **Support Vector Machine (SVM):** This algorithm finds the hyperplane maximizing the margin between classes, making it effective for high-dimensional data and complex relationships.
- **Decision Tree Classifier:** This tree-like model makes decisions based on a series of rules and is easy to interpret, providing insights into the factors influencing predictions.
- **Convolutional Neural Network (CNN):** This deep learning architecture is particularly well-suited for processing data with grid-like structures, such as

images. We will adapt CNNs to handle our tabular data to explore their potential for performance prediction.

4.3 Evaluation Metrics:

We evaluate the performance of each model using the following metrics:

- **Accuracy:** The proportion of correctly classified instances, providing an overall measure of the model's effectiveness.
- **F1 Score:** The harmonic mean of precision and recall, offering a balanced evaluation for imbalanced datasets and considering both false positives and false negatives.
- **Confusion Matrix:** This table visualizes the model's performance by showing the counts of true positives, true negatives, false positives, and false negatives for each class.
- **Classification Report:** This report provides a detailed breakdown of precision, recall, F1-score, and support for each class (quartile of global score), allowing for a comprehensive understanding of the model's behavior across different performance levels.
- **ROC Curve and AUC:** This visualization technique illustrates the trade-off between true positive rate and false positive rate, and the area under the ROC curve (AUC) summarizes the model's overall performance in distinguishing between classes.

4.4 Interpretability and Feature Importance:

- **Permutation Importance:** This technique assesses the importance of each feature by measuring the decrease in model performance when the feature's values are randomly shuffled.
- **SHAP (SHapley Additive exPlanations):** This gametheoretic approach explains the output of any machine learning model by providing insights into the contribution of each feature to the predictions.
- **LIME (Local Interpretable Model-agnostic Explanations):** This method focuses on explaining individual predictions by creating a locally faithful interpretable model around the prediction. It helps understand how specific features contribute to a particular prediction, offering insights into the model's behavior on a case-by-case basis.

5 RESULTS AND DISCUSSION

This chapter presents the findings of the data analysis, focusing on the performance of different machine learning models in predicting student performance quartiles based on academic and socioeconomic factors.

5.1 Model Performance Comparison:

Table 1 Precision, Recall, F1 Score, Support and Accuracy of the models tested.

Model	Class	Precision	Recall	F1 Score	Support	Accuracy
RandomForest	Q1	0.93	0.88	0.91	657	
	Q2	0.79	0.81	0.80	589	
	Q3	0.80	0.85	0.82	625	**0.86**
	Q4	0.94	0.90	0.92	612	
	Overall			0.86	2483	
KNN	Q1	0.84	0.88	0.86	657	
	Q2	0.69	0.67	0.68	589	
	Q3	0.67	0.71	0.69	625	**0.76**
	Q4	0.87	0.80	0.84	612	
	Overall			0.77	2483	
GMLBoost	Q1	0.97	0.93	0.95	657	
	Q2	0.88	0.90	0.89	589	
	Q3	0.89	0.93	0.91	625	**0.92**
	Q4	0.98	0.95	0.96	612	
	Overall			0.93	2483	
GLMNet	Q1	0.96	0.95	0.95	657	
	Q2	0.89	0.88	0.89	589	
	Q3	0.88	0.89	0.88	625	**0.91**
	Q4	0.95	0.94	0.95	612	
	Overall			0.92	2483	
SVM	Q1	0.96	0.80	0.87	657	
	Q2	0.69	0.78	0.73	589	
	Q3	0.64	0.72	0.68	625	**0.77**
	Q4	0.84	0.79	0.81	612	
	Overall			0.77	2483	
DecisionTree	Q1	0.88	0.89	0.89	657	
	Q2	0.74	0.71	0.73	589	
	Q3	0.76	0.76	0.76	625	**0.82**
	Q4	0.89	0.90	0.90	612	
	Overall			0.82	2483	
CNN	Q1	0.98	0.93	0.95	657	
	Q2	0.90	0.81	0.85	589	
	Q3	0.81	0.94	0.87	625	**0.90**
	Q4	0.94	0.94	0.94	612	
	Overall			0.90	2483	

The table provides a comprehensive comparison of the performance of various machine learning models in predicting academic performance across different quartiles (Q1-Q4). The RandomForest model achieved a high overall accuracy of 86.99% with strong precision and recall, particularly excelling in Q1 and Q4. The KNN model, while achieving an overall accuracy of 76.84%, showed moderate performance, particularly struggling with Q2 and Q3. GMLBoost emerged as the top performer with an overall accuracy of 92.99% and consistently high precision, recall, and F1 scores across all quartiles, indicating robust predictive capability. The GLMNet model also performed admirably with an overall accuracy of 91.78%, demonstrating balanced performance across all quartiles. The SVM model, with an overall accuracy of 77.16%, showed a notable drop in performance for Q3, reflecting its limitations in certain quartiles. The DecisionTree model achieved an overall accuracy of 82.12%, with commendable performance in Q1 and Q4 but lower scores in Q2 and Q3. Lastly, the CNN model demonstrated strong predictive power with an overall accuracy of 90.78%, particularly excelling in Q1 and Q4. This comparison underscores the effectiveness of ensemble methods like GMLBoost and RandomForest in handling complex predictive tasks involving academic performance data.

5.2 Confusion Matrices:

Referencing the submitted images, the confusion matrices show the numbers of accurate and inaccurate predictions for each quartile, giving a visual depiction of the models' performance. These matrices' analysis provides insightful information about the advantages and disadvantages of the models.

For instance, the Random Forest confusion matrix shows a higher number of misclassifications between Q1 (lowest performance) and Q2, suggesting challenges in distinguishing students within these performance levels. This could be attributed to similar underlying factors influencing students in these quartiles or the models' limitations in capturing subtle differences.

In contrast, the GMLBoost confusion matrix demonstrates a clear diagonal pattern with minimal misclassifications, indicating its ability to accurately differentiate between the performance quartiles.

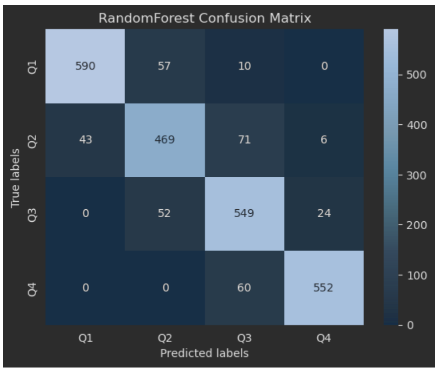

Fig. 1 *Random Forest*

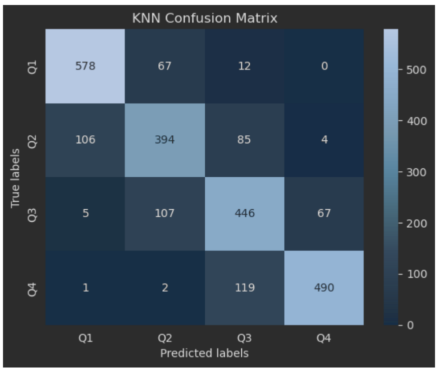

Fig. 2 KNN

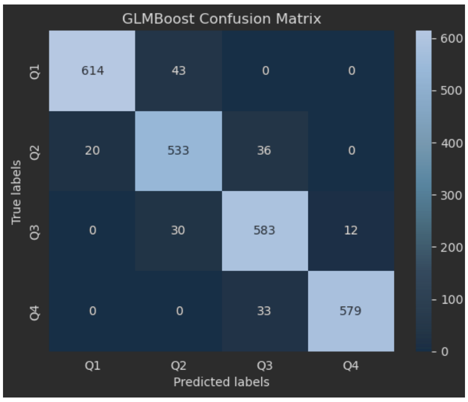

***Fig. 3** GLMBOOST*

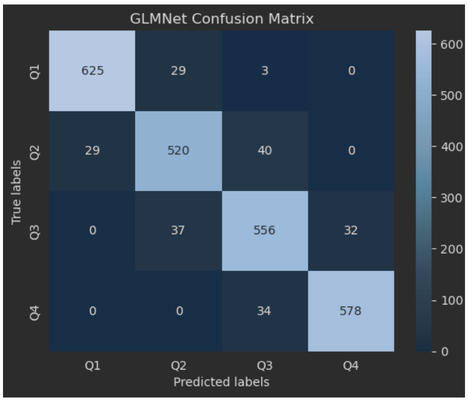

Fig. 4 GLMNet

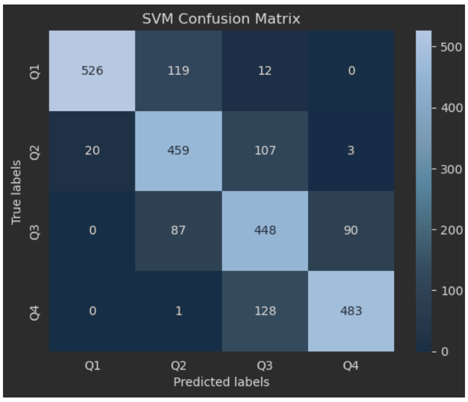

Fig. 5 SVM

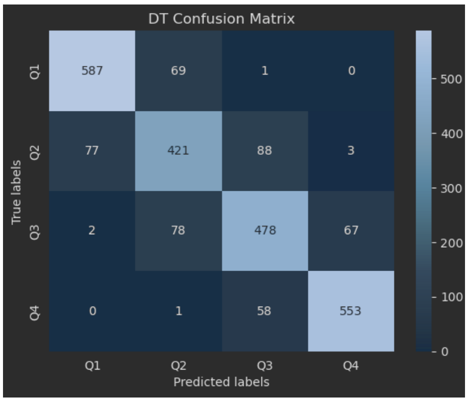

Fig. 6 Decision Tree

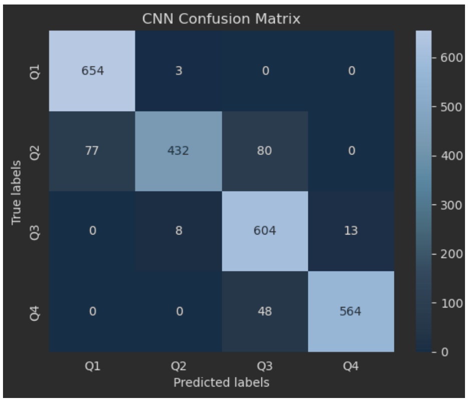

Fig. 7 CNN

5.3 ROC Curves and AUC:

The trade-off between each model's true positive rate (sensitivity) and false positive rate (specificity) is shown by the comparison of the ROC curves (see the uploaded image). AUC offers a solitary measure that encapsulates the mod-els' comprehensive efficacy. Every model shows an AUC value that is much higher than 0.5, meaning that it can distinguish between various performance quartiles more accurately than a random guess. CNN, GLMNet, and GMLBoost have the greatest AUC values, indicating their superiority.

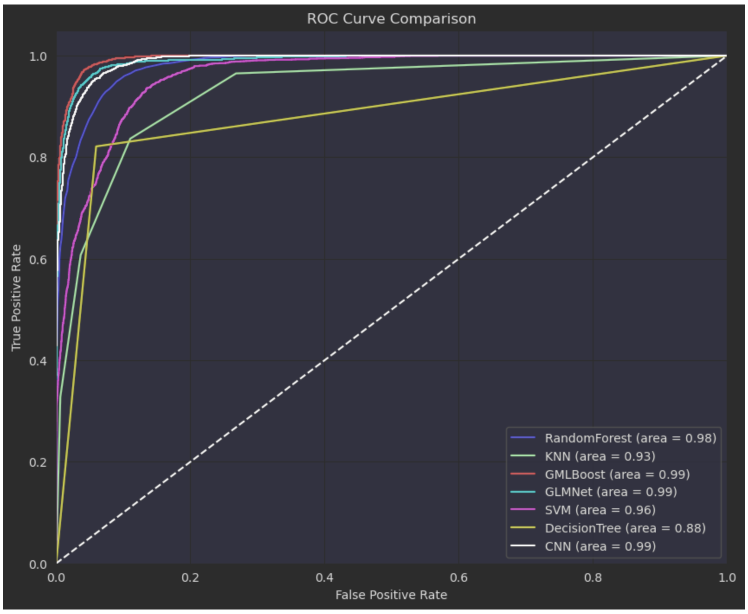

Fig. 8 ROC Curve

5.4 Feature Importance:

The feature importance analysis, using both permutation importance and SHAP values, revealed valuable insights into the factors influencing student performance (refer to the uploaded feature importance plots):

- **University-Level Competencies:** Scores on the Saber Pro tests, particularly Critical Reading (CR PRO) and Citizen Competencies (CC PRO), consistently ranked among the top predictors across multiple models. This underscores the importance of these skills for success in engineering programs.
- **Secondary-Level Performance:** Scores from the Saber 11 tests, especially in mathematics and science, also exhibited moderate importance in several models, suggesting that a strong foundation in these areas contributes to success at the university level.

- **Socioeconomic Factors:** Variables like socioeconomic level and parents' education demonstrated varying degrees of influence across different models. This indicates the complex interplay between socioeconomic background and academic achievement.

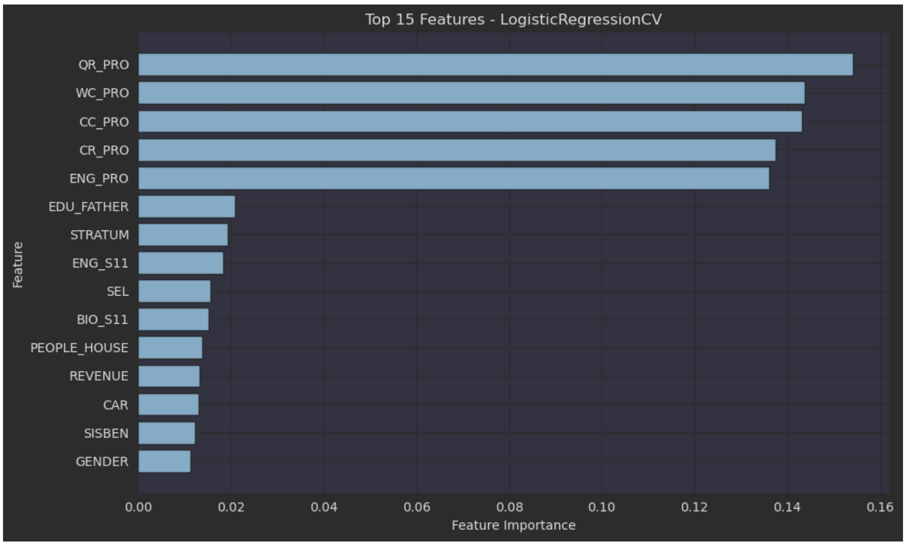

Fig 9 Top 15 Features for GLMNet

The SHAP dependence plot for the Citizen Competencies score at the university level (CC PRO) reveals a distinctive pattern, indicating its influence on the Random Forest model's predictions for student performance. The plot shows a negative correlation: as the CC PRO score increases, the SHAP value decreases, which suggests that higher scores on CC PRO tend to contribute to a higher predicted performance quartile (such as Q3 or Q4), while lower scores on this feature are associated with lower performance quartiles (such as Q1 or Q2).

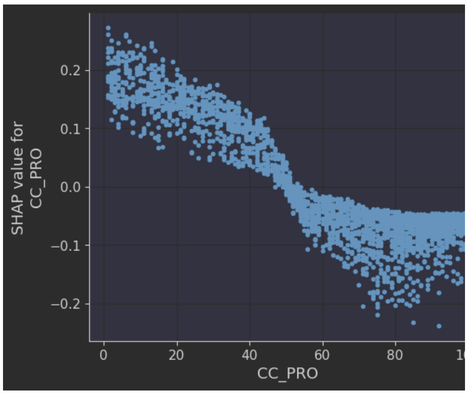

Fig 10 Dependence Plot for Q1

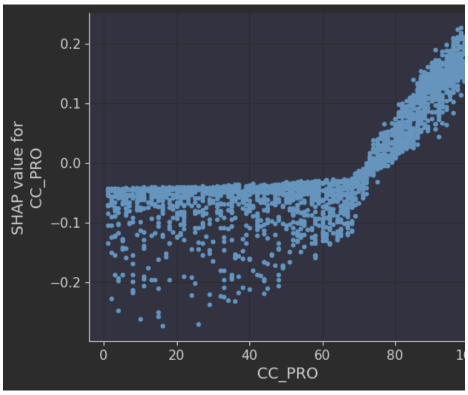

Fig 11 Dependence Plot for Q4

This trend is compelling, suggesting that students who demonstrate stronger capabilities in areas assessed by the Citizen Competencies are likely to perform better academically in their engineering programs. The distribution of SHAP values across the range of CC PRO scores also implies a nonlinear relationship, where the incremental impact on the model's output varies depending on the score range.

The SHAP dependence plot underscores the importance of Citizen Competencies as a strong predictor of academic success, while also highlighting the multidimensional nature of educational achievement. It suggests that while CC PRO is influential, its effect must be understood in the context of a multifaceted model that includes a diverse set of academic and socioeconomic variables.

For educational policymakers and curriculum designers, these insights emphasize the need to cultivate citizen competencies within engineering education frameworks. It prompts an inquiry into how these competencies are integrated into coursework and assessment, ensuring that students are equipped not only with technical knowledge but also with the social and ethical understanding necessary for the engineering profession.

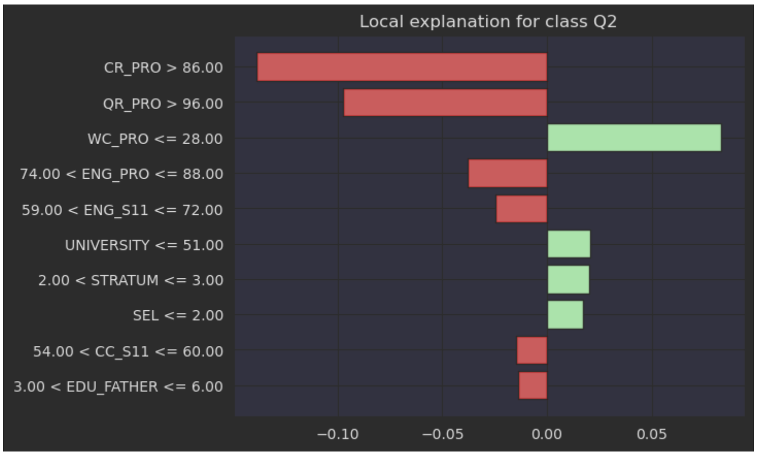

Fig 12 Lime Plot Local Explaination

The LIME plot provides an insightful explanation of the key features influencing the prediction of a student's classification into quartile Q2. Notably, higher scores in critical reading (CR_PRO) and quantitative reasoning (QR_PRO) from professional assessments significantly decrease the likelihood of being classified into Q2, indicating that better performance in these areas correlates with placement in higher quartiles. Conversely, moderate scores in professional English (ENG_PRO) positively influence the classification into Q2, while lower scores in written communication (WC_PRO) decrease it. Socioeconomic factors also play a crucial role, with students from mid-range socioeconomic strata and lower socioeconomic levels being more likely to be classified into Q2. Additionally, certain universities and the father's education level between three and six years negatively impact the classification, suggesting complex relationships between these variables and academic performance. This analysis highlights the multifaceted nature of academic success and the importance of considering various educational and socioeconomic factors in predictive modelling.

5.5 Comparative Analysis

Table 2 Comparing the outcomes of other similar studies.

Study	Dataset	Model	Accuracy
Current Research	ICFES, Colombia (Standardized Test Scores, Socioeconomic)	Gradient Boosting Machine (GBM)	92 %
Alhazmi and Sheneamer (2023) [8]	Jazan University, Saudi Arabia (Admission Scores, First-Year Grades)	Gaussian Naive Bayes (GNB)	74%
Soto-Acevedo et al. (2023) [2]	ICFES, Colombia (Standardized Test Scores, Socioeconomic)	Generalized Linear Network Model (GLM Net)	82%
Poudyal et al. (2022) [17]	OULAD (LMS Interaction, Demographics, Assessment)	Hybrid 2D CNN	88%
Aljaloud et al. (2022) [21]	University of Ha'il, Saudi Arabia (Blackboard Interaction)	CNN-LSTM	94%

The varying accuracy across different studies is attributable to factors like the specific dataset, the chosen models, the target variable being predicted, and the selection of relevant features. The findings of this comparative analysis strengthen the importance of considering both academic and socioeconomic factors in predicting engineering student performance. By leveraging these factors and employing powerful machine learning models, we can develop accurate predictive tools that support targeted interventions and improve educational outcomes in engineering programs.

5.6 Findings

Our findings align with previous research [1], [2], [10]– [12] highlighting the significant influence of both secondary and university-level academic competencies on student performance in engineering programs. As emphasized by Kabakchieva (2013) [13], a strong foundation in secondary education, particularly in mathematics and science as noted by Ramesh, Parkavi & Ramar (2013) [11], plays a crucial role in predicting success at the university level. Furthermore, our analysis, along with the work of Bydzovsk ˇ a (2020) [14], ´ reveals that university-level academic competencies, especially critical reading (CR PRO) and citizen competencies (CC PRO), are strong predictors of student performance. These findings underscore the importance of fostering these skills throughout the educational journey of engineering students. Kanani et al. (2023) [19] demonstrated the effectiveness of LSTM models in time-series predictions, achieving high accuracy in rainfall forecasting, which is comparable to our use of advanced models for predicting student performance.

Similar to the research by Ramesh, Parkavi & Ramar (2013) [11], our study found that parental occupation, among other socioeconomic factors, has a notable impact on student performance. This aligns with the broader discussion on the complex interplay between socioeconomic background and academic achievement within engineering education [1], [10], [15]. Daud et al. (2017) [15] further emphasize the importance of considering family expenditures and student personal information as potential predictors of academic success. These findings suggest the need for a more nuanced understanding of how various socioeconomic and personal factors interact and influence student outcomes in engineering programs.

While our research primarily employed traditional machine learning models, studies such as Bydzovsk ˇ a (2020) [14] and ´ Thai-Nghe et al. (2010) [16] demonstrate the potential of utilizing social behavior data and collaborative filtering techniques for performance prediction. These approaches could offer valuable insights into student performance by leveraging similarities among students and courses, potentially leading to more personalized interventions and support systems. Future research could explore the integration of collaborative filtering and recommender systems within the context of engineering education to enhance predictive models and develop targeted strategies for student success.

6 CONCLUSION AND FUTURE WORK

This research explored the factors influencing academic performance in engineering education using a comprehensive dataset of Colombian students. By employing various machine learning models and feature importance analysis, we gained valuable insights into the relationship between academic competencies, socioeconomic background, and student success in engineering programs.

Our findings highlight the significant role of university level academic competencies, particularly critical reading and citizen competencies, in determining student success. Additionally, secondary-level performance in mathematics and science subjects demonstrated a moderate influence on academic achievement. Socioeconomic factors exhibited varying degrees of importance across different models, suggesting a complex interplay between these variables and academic outcomes.

6.1 Implications for Engineering Education:

These findings offer valuable implications for educators and policymakers:

- **Curriculum Development:** Emphasize the development of critical reading, problem-solving, and communication skills within engineering curricula to better prepare students for academic success and future careers.
- **Targeted Interventions:** Implement early interventions and support systems for students struggling in key areas identified by the feature importance analysis, such as critical reading and citizen competencies.
- **Bridging the Gap:** Address the potential impact of socioeconomic disparities on academic achievement through initiatives that provide equitable access to resources and support for students from disadvantaged backgrounds.

- **Data-Driven Decision Making:** Utilize data mining and machine learning techniques to gain a deeper understanding of student performance patterns and develop evidence-based strategies for improving learning outcomes.

6.2 Limitations and Future Research:

While this study provides valuable insights, it is essential to acknowledge its limitations:

- **Dataset Specificity:** The findings are based on data from Colombia and may not be directly generalizable to other contexts. Future research could explore similar analyses in different countries and educational systems.
- **Target Variable Transformation:** The binning of the global score into quartiles may result in information loss. Exploring regression models to predict continuous performance scores could be a valuable direction for future research.
- **Model Interpretability:** While we analyzed feature importance, further investigation into the models' decision-making processes and the relationships between different variables could provide a more nuanced understanding of the factors influencing academic performance.

6.3 Future research directions could include:

- **Incorporating additional factors:** Exploring the influence of other relevant variables, such as learning styles, motivation, engagement, and psychological factors, on academic performance.
- **Longitudinal studies:** Tracking student performance over time to understand how different factors interact and evolve throughout their academic journey.
- **Personalized learning:** Developing adaptive learning systems and interventions tailored to individual student needs and learning styles based on predictive models and data analysis.
- **Causal inference:** Utilizing advanced statistical techniques to establish causal relationships between identified factors and academic performance, moving beyond correlation-based analyses.

References

1. Delahoz-Domínguez, E., Zuluaga-Ortiz, R., Herrera, T.J.F.: Dataset of academic performance evolution for engineering students. Data in Brief 30, 105537 (2020).
2. Soto-Acevedo, M., Abuchar-Curi, A.M., Zuluaga-Ortiz, R., Delahoz-Domínguez, E.: A machine learning model to predict standardized tests in engineering programs in Colombia. IEEE-RITA 18(3), 211–218 (2023).
3. Khan, M., Naz, S., Khan, Y., Zafar, M., Khan, M., & Pau, G., "Utilizing machine learning models to predict student performance from LMS activity logs," IEEE Access, vol. 11, pp. 86953–86962, (2023)

4. Rizvi, S., Rienties, B., & Khoja, S. A., “The role of demographics in online learning; A decision tree based approach,” Computers and Education/Computers & Education, vol. 137, pp. 32–47, (2019)
5. Rivas, A., Gonzalez-Briones, A., Hernandez, G., Prieto, J., Chamoso, P., “Artificial neural network analysis of the academic performance of students in virtual learning environments,” Neurocomputing, vol. 423, pp. 713–720, (2021)
6. Zohair, Abu, and Lubna Mahmoud. "Prediction of Student’s performance by modelling small dataset size." International Journal of Educational Technology in Higher Education 16.1 (2019): 1-18.
7. Azizah, E.N.; Pujianto, U.; Nugraha, E. Comparative performance between C4. 5 and Naive Bayes classifiers in predicting student academic performance in a Virtual Learning Environment. In Proceedings of the 4th International Conference on Education and Technology (ICET), Malang, Indonesia, pp. 18–22. (2018)
8. Alhazmi, E., Sheneamer, A.: Early predicting of students performance in higher education. IEEE Access 11, 27579–27589 (2023).
9. Sun, D., et al.: A university student performance prediction model and experiment based on Multi-Feature Fusion and Attention Mechanism. IEEE Access 11, 112307–112319 (2023).
10. Delahoz-Domínguez, E., Zuluaga-Ortiz, R., Herrera, T.J.F.: Dataset of academic performance evolution for engineering students. Data in Brief 30, 105537 (2020).
11. Ramesh, V., Parkavi, P., Ramar, K.: Predicting Student Performance: A statistical and data mining approach. International Journal of Computer Applications 63(8), 35–39 (2013).
12. Bydžovská, H.: A Comparative Analysis of Techniques for Predicting Student Performance. International Educational Data Mining Society (2016).
13. Kabakchieva, D.: Predicting student performance by using data mining methods for classification. Cybernetics and Information Technologies 13(1), 61–72 (2013).
14. Rastrollo-Guerrero, J.L., Gómez-Pulido, J.A., Durán-Domínguez, A.: Analyzing and Predicting Students’ Performance by Means of Machine Learning: A Review. Applied Sciences 10(3), 1042 (2020).
15. Daud, A., Aljohani, N.R., Abbasi, R.A., Lytras, M.D., Abbas, F., Alowibdi, J.S.: Predicting Student Performance using Advanced Learning Analytics. In: 26th International Conference on World Wide Web Companion, pp. 1-2 (2017).
16. Thai-Nghe, N., Drumond, L., Krohn-Grimberghe, A., Schmidt-Thieme, L.: Recommender system for predicting student performance. Procedia Computer Science 1(2), 2811–2819 (2010).
17. Poudyal, S., Mohammadi-Aragh, M.J., Ball, J.E., “Prediction of student academic performance using a hybrid 2D CNN model,” Electronics, vol. 11, no. 7, p. 1005, (2022)
18. Chau, V.T.N., Phung, N.H., “Enhanced CNN models for binary and multiclass student classification on temporal educational data at the program level,” Vietnam Journal of Computer Science, vol. 08, no. 02, pp. 311–335, (2020)
19. S. Kanani, S. Patel, R. K. Gupta, A. Jain, and J. C.-W. Lin, “An AI-Enabled ensemble method for rainfall forecasting using Long-Short term memory,” Mathematical Biosciences and Engineering, vol. 20, no. 5, pp. 8975–9002, (2023)
20. Nabil, A., Seyam, M., Abou-Elfetouh, A., “Prediction of students’ academic performance based on courses’ grades using deep neural networks,” *IEEE Access*, vol. 9, pp. 140731–140746, (2021)
21. Aljaloud, A.S., Uliyan, D.M., Alkhalil, A., Abd Elrhman, M., Alogali, A.F.M., Altameemi, Y.M., Altamimi, M., Kwan, P., “A deep learning model to predict student learning outcomes in LMS using CNN and LSTM,” IEEE Access, vol. 10, pp. 85255–85265, (2022)