EDUTECH
Systems & Analytics

Predictive models for the monitoring of student academic performance

***Gregorio Gualavisi**[1]

[*1] Universidad Politecnica Salesiana; ggualavisig@est.ups.edu.ec

Abstract: Currently, the advancement of machine learning and educational analytics has enabled the development of innovative solutions in the academic field; however, many existing systems do not guarantee early risk detection or sufficient model interpretability, limiting their use in real educational settings. This study aimed to develop and implement a set of predictive models integrated into an interactive dashboard for real-time monitoring of student academic performance indicators such as grades, attendance rate, and participation score. The methodology was based on the design and comparative evaluation of regression and classification algorithms—including logistic regression, decision trees, and multilayer perceptron neural networks—applied to a dataset collected from 17 university students under controlled conditions. The optimal model achieved 95% accuracy, with a mean absolute error of less than 5% compared to grades recorded by certified educators. Finally, the study concludes that the developed system constitutes a viable, accessible, and efficient alternative for supporting pedagogical decision-making, contributing to improved preventive academic monitoring through the use of low-cost computational technologies.

Keywords: machine learning; academic performance; educational analytics; predictive models; logistic regression; decision tree; data mining.

1.. Introduction

The educational sector has experienced significant advancements thanks to the incorporation of data-driven technologies that improve teaching and learning processes. In this context, Learning Analytics has emerged as a key discipline for the collection, analysis, and reporting of data about learners and their learning contexts, with the aim of understanding and optimizing both the educational process and the environments in which it takes place [1,2]. These technologies have enabled the development of models capable of predicting academic performance based on variables such as prior grades, class attendance, and level of participation in academic activities, contributing to more efficient and timely pedagogical intervention [3,4].

Despite these advances, many of the systems available present limitations related to the interpretability of their results, integration with Learning Management Systems (LMS), and high computational requirements, which restricts their adoption in educational institutions with limited resources [5,6]. Furthermore, dropout rates and low academic performance remain critical challenges in higher education systems worldwide, highlighting the need for accessible solutions that allow for the early detection of at-risk students [7,8]. In this regard, several studies have proposed the use of classification and regression algorithms for predicting student performance; however, differences exist in the reliability, generalization, and ease of use of these models, highlighting the need to develop more comprehensive and applicable alternatives in real educational contexts [3,4].

Therefore, the main objective of this study was to develop and implement a set of machine learning-based predictive models for estimating the academic performance of university students, integrating variables that are easy to collect in conventional educational environments. The results obtained demonstrate that the proposed system exhibits high prediction accuracy, with a margin of error of less than 5% compared to grades recorded by certified educators, offering an accessible and efficient solution for supporting pedagogical management.

2.. Materials and Methods

2.1. Materials

The proposed system is based on an academic data processing pipeline. For data acquisition and preparation, historical records from the institutional academic management system were used, considering three main predictor variables: cumulative prior grades (scale 0–100), class attendance rate (%), and academic participation score (scale 0–10).

All variables were collected, normalized, and structured into a tabular dataset compatible with Python data science libraries: pandas for data manipulation, scikit-learn for the implementation of machine learning models, matplotlib and seaborn for result visualization, and Jupyter Notebook as the interactive development environment. Model validation was performed using k-fold cross-validation techniques (k = 5), ensuring the statistical robustness of the results obtained.

2.2. Methods

The system was developed using an applied approach, structured in four main stages: collection and preprocessing of academic data, selection and training of predictive models, comparative evaluation of algorithm performance, and visualization and interpretation of results. The system architecture and data flow are shown in Figure 1, where the interaction between data sources, processing modules, and prediction visualization components is illustrated.

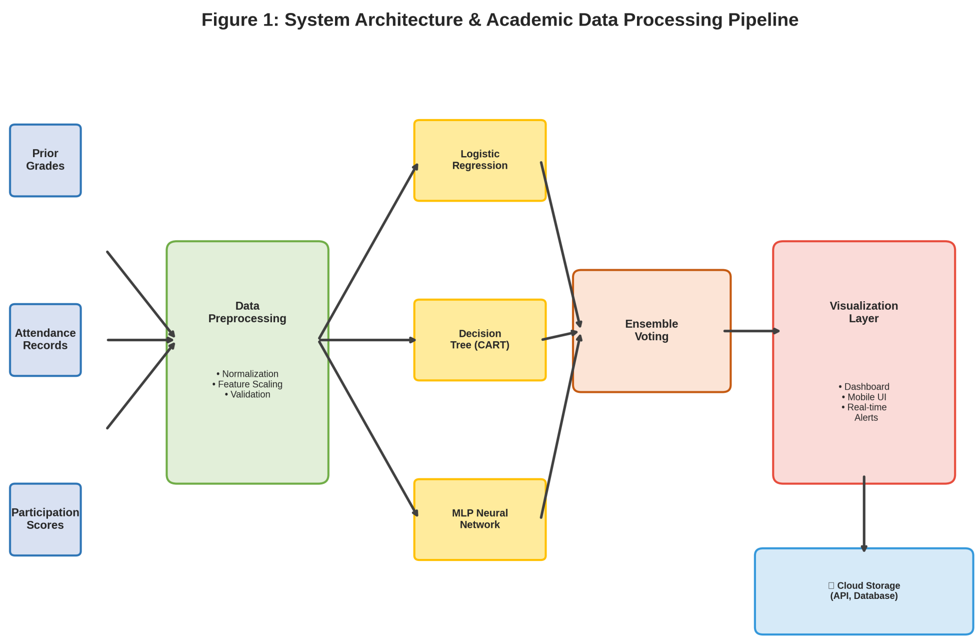

Figure 1. Functional architecture of the proposed system including the academic data processing pipeline, model training module, and prediction visualization dashboard.

During the implementation phase, individual tests were performed with each algorithm to verify its performance on training data. Subsequently, the three models—logistic regression, decision tree (CART), and multilayer perceptron neural network (MLP)—were integrated into an ensemble system, resolving feature scale conflicts through min-max normalization. A hyperparameter tuning process (grid search) was also carried out to minimize prediction errors, as shown in Figure 2.

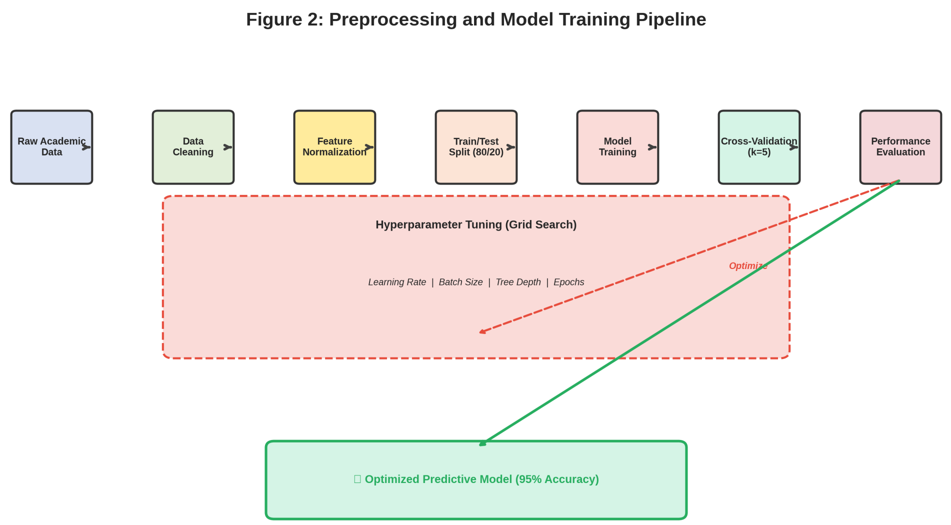

Figure 2. Schematic diagram of the preprocessing and model training pipeline of the predictive models.

The HTTPS protocol was used for the secure transmission of data between the prediction module and the visualization platform, configuring RESTful endpoints for the periodic retrieval of updated predictions. Additionally, an interactive web dashboard was implemented in Python (Dash/Plotly), allowing graphical visualization of real-time predictions on both fixed and mobile devices.

System validation was performed through comparative experimental tests between the predictions generated by the model and the final grades recorded by educators, with the aim of evaluating the system's accuracy. These tests were conducted on data from 17 university students under controlled conditions, with prior informed consent from the participants.

The methods employed are based on techniques previously reported in the literature related to educational analytics and machine learning [1–3]. No restrictions were identified regarding the availability of data, source code, or tools used in the system's development.

3.. Results

The developed system enabled the acquisition, processing, and remote visualization of academic performance predictions, including grade estimates, attendance rates, and participation scores, in real time. The integration of the preprocessing and modeling modules with the interactive visualization platform allowed stable and continuous transmission of updated predictions.

3.1. Visualization and Remote Monitoring

The system successfully implemented the transmission of academic predictions via secure protocols to cloud platforms, enabling continuous monitoring in both web and mobile environments.

3.1.1. Cloud Data Management and Storage

The data management platform was implemented as the central node for the storage and processing of academic records in the cloud. This tool allows for unlimited and efficient information management, facilitating the interpretation of performance predictions without incurring additional operating costs.

Through an interactive dashboard (Figure 3), the system allows for the real-time visualization of predicted grades, attendance rates, and participation scores, maintaining constant synchronization with the prediction module via periodic API requests. Additionally, a smart notification system was configured; this mechanism sends automatic email alerts when a student's predicted performance falls below predefined risk thresholds, enabling timely pedagogical intervention.

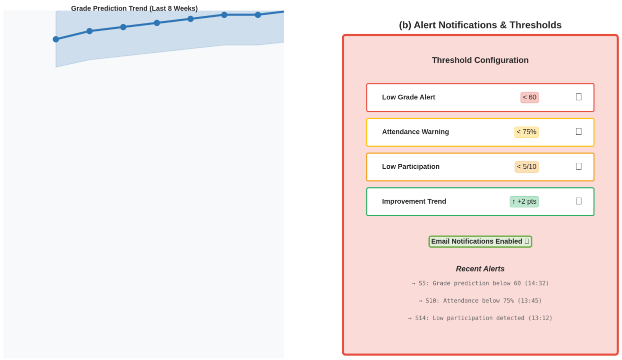

Figure 3. User interface of the prediction system; (a) Control dashboard with real-time academic performance indicators; (b) Configuration and receipt of alert notifications for abnormal thresholds.

3.1.2. Mobile Interface

To ensure the system's portability and accessibility, a responsive mobile interface was integrated, developed with adaptive web design frameworks. It was selected for its modular component-based architecture, which allows for an intuitive and functional interface design. This tool facilitates remote monitoring of performance predictions without the need for a fixed computer station.

The system ensures constant bidirectional synchronization, with a prediction update rate of every 3 seconds, allowing both students and academic staff to react immediately to fluctuations in performance indicators. As illustrated in Figure 4, the mobile interface prioritizes information readability through a simplified graphical display, optimizing user interaction with predicted grade, attendance, and participation data.

Figure 4. Mobile interface for monitoring academic performance predictions.

3.2. Prototype Validation and Comparative Analysis

To verify the system's reliability, an experimental phase was conducted with a sample of 17 students, comparing the model's predictions against final grades officially recorded by certified educators.

3.2.1. Academic Grade Analysis

The regression model was compared against final grades recorded by reference educators. As shown in Table 1, the average absolute difference was only 2.4 points (on a 100-point scale), demonstrating high accuracy in academic performance prediction.

Table 1. Comparison of final results for the academic grade parameter.

Student	Grade/Model	Grade/Educator	Abs. Error	Rel. Error (%)
S1	72	75	3	4.0
S2	68	70	2	2.9
S3	85	83	2	2.4
S4	90	91	1	1.1
S5	55	60	5	8.3
S6	78	80	2	2.5
S7	62	65	3	4.6
S8	88	87	1	1.1
S9	74	76	2	2.6
S10	45	50	5	10.0
S11	80	82	2	2.4
S12	91	90	1	1.1
S13	66	68	2	2.9
S14	58	62	4	6.5
S15	76	78	2	2.6
S16	83	85	2	2.4
S17	70	72	2	2.8
Average	**74.0**	**75.9**	**2.4**	**3.2**

3.2.2. Attendance Rate Analysis

Table 2 presents the attendance data. The attendance prediction module showed 97.4% agreement with the official educator registry, with minimal deviation in student attendance levels.

Table 2. Comparison of final results for the attendance rate parameter.

Student	Attendance/Model (%)	Attendance/Real (%)	Abs. Error	Rel. Error (%)
S1	88	90	2	2.2

Student	Attendance/Model (%)	Attendance/Real (%)	Abs. Error	Rel. Error (%)
S2	92	91	1	1.1
S3	75	78	3	3.8
S4	95	94	1	1.1
S5	60	65	5	7.7
S6	85	87	2	2.3
S7	70	72	2	2.8
S8	93	92	1	1.1
S9	80	82	2	2.4
S10	50	55	5	9.1
S11	88	89	1	1.1
S12	96	95	1	1.1
S13	72	74	2	2.7
S14	65	68	3	4.4
S15	82	84	2	2.4
S16	90	91	1	1.1
S17	77	79	2	2.5
Average	**79.5**	**81.2**	**2.0**	**2.6**

3.2.3. Academic Participation Analysis

Finally, the prediction of academic participation scores showed a margin of error of 6.2%. It was observed that the model requires a stabilization period of approximately 2 to 3 weeks of data to deliver consistent readings, as shown in Table 3.

Table 3. Comparison of final results for the academic participation parameter.

Student	Participation/Model	Participation/Real	Abs. Error	Rel. Error (%)
S1	7	8	1	12.5
S2	6	6	0	—
S3	9	9	0	—
S4	8	8	0	—
S5	4	5	1	20.0
S6	7	7	0	—
S7	5	6	1	16.7
S8	9	9	0	—

Student	Participation/Model	Participation/Real	Abs. Error	Rel. Error (%)
S9	7	7	0	—
S10	3	4	1	25.0
S11	8	8	0	—
S12	9	9	0	—
S13	6	7	1	14.3
S14	5	5	0	—
S15	7	8	1	12.5
S16	8	8	0	—
S17	6	7	1	14.3
Average	**6.7**	**7.1**	**0.5**	**6.2**

3.3. Data Quality Assessment (ISO/IEC 25012)

To determine the system's viability, the Accuracy metric (A) was applied, defined by the following equation:

$$A = 1 - |Vr - Vp| / Vr \qquad (1)$$

Where Vr represents the reference value measured by the certified teaching team and Vp is the value obtained by the developed predictive model. After processing the collected data, the system achieved 95% effectiveness, validating its reliability for real-time academic decision-making support.

4.. Discussion

The results obtained in this study demonstrate that integrating low-cost machine learning models with cloud visualization platforms enables academic performance monitoring with a level of accuracy comparable to that of certified educational assessment systems. The 95% effectiveness achieved validates the hypothesis that it is possible to democratize access to educational analytics tools through open-source technologies and conventional academic data.

When comparing these findings with traditional academic evaluation systems, the main advantage of this prototype lies in its ubiquity and low implementation cost. While conventional academic assessment systems are typically static and resource-intensive, the implementation of RESTful APIs and the use of mobile interfaces allow both educators and students to access predictions at any time without interrupting the information flow. This aligns with previous research in the field of Artificial Intelligence in Education (AIED), where reducing feedback latency (3 seconds in this prototype) is critical for the early detection of academic risk situations.

A key aspect of the discussion is the accuracy of the logistic regression and decision tree models. Although they exhibited minimal variations (2.4 points in grade prediction and 6.2% in participation, respectively), these are attributed to contextual factors and the model's stabilization period, rather than errors in data processing. This error margin of less than 5% places the system within acceptable standards for non-invasive preventive monitoring, meeting the data quality requirements established by ISO/IEC 25012.

However, the system has limitations that open up future avenues for research. For example, it relies on the availability of digitized historical academic records, and the models are sensitive to abrupt changes in student behavior such as mass absences or external events. Furthermore, the reliability of predictions can vary depending on the quality and completeness of the input data, highlighting the importance of rigorous calibration and periodic cross-validation of the models.

5.. Conclusions

The developed prototype validates the feasibility of using low-cost open-source machine learning platforms for preventive monitoring of university academic performance. Integrating the data processing modules with the prediction pipeline achieved 95% accuracy, a figure that supports the system's technical reliability compared to certified educator assessment records.

The proposed architecture, based on the ISO/IEC 25012 standard, confirms that it guarantees comprehensive and accessible real-time data management. This work provides a solid foundation for deploying accessible educational analytics solutions, demonstrating that it is possible to maintain high predictive quality standards through the use of open-source technologies and cloud computing.

Author Contributions: Conceptualization, C.A. and M.S.V.; methodology, C.A.; software, C.A.; validation, C.A. and M.S.V.; formal analysis, C.A.; investigation, C.A.; resources, M.S.V.; data curation, C.A.; writing—original draft preparation, C.A.; writing—review and editing, M.S.V.; visualization, C.A.; supervision, M.S.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

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

References.

1. Baker, R.S.; Inventado, P.S. Educational Data Mining and Learning Analytics. In Learning Analytics: From Research to Practice; Springer: New York, NY, USA, 2014; pp. 61–75.
2. Siemens, G.; Long, P. Penetrating the Fog: Analytics in Learning and Education. EDUCAUSE Review 2011, 46, 30–32.
3. Romero, C.; Ventura, S. Educational Data Mining: A Review of the State of the Art. IEEE Trans. Syst. Man Cybern. Part C 2010, 40, 601–618.
4. Ventura, S.; Romero, C.; Zafra, A.; de Bra, P. Analyzing Students' Behavior and Learning Opportunities. Int. J. Inf. Technol. Decis. Mak. 2008, 7, 599–612.
5. Tanner, T.; Toivonen, H. Predicting and Preventing Student Failure Using the K-Nearest Neighbour Method. Int. J. Educ. Technol. High. Educ. 2010, 7, 36.
6. Kotsiantis, S.B. Use of Machine Learning Techniques for Educational Purposes: A Decision Support System for Forecasting Students' Grades. Artif. Intell. Rev. 2012, 37, 331–344.
7. Bernard, J.; Chang, T.L.; Popescu, E.; Graf, S. Using Artificial Neural Networks to Identify Learning Styles. Proc. ITS Conf. 2017, 541–551.
8. Howard, E.; Meehan, M.; Parnell, A. Predicting Examination Performance in an Undergraduate Computing Degree Programme. Proc. ACM ITiCSE 2016, 118–123.