International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V72I7P113

Volume 72 Issue 7, 118-123, July 2024

Original Article

Study Hours vs. Exam Results: Academic Success of Students through Performance Prediction

Lida Asencios-Trujillo[1], Lucia Asencios-Trujillo[2], Carlos La-Rosa-Longobardi[3], Djamila Gallegos-Espinoza[4]

[1, 2,3,4]*Graduate School, Enrique Guzmán y Valle National University of Education, Lima, Peru.*

[1]*Corresponding Author : asenciostrujillolida@gmail.com*

Received: 21 February 2024 Revised: 05 May 2024 Accepted: 15 June 2024 Published: 26 July 2024

Abstract *- This study investigates the correlation between study hours and students' exam results to identify key predictors of academic success. Using a dataset of 500 records that include study hours, previous exam scores, and pass/fail outcomes, we applied regression models and machine learning to analyze the influence of these factors on exam performance. Our findings indicate that while study hours have a significant impact on the likelihood of passing an exam, previous exam scores provide a more accurate prediction of the outcome. This study highlights the importance of considering multiple factors when assessing academic performance and suggests targeted strategies to improve educational outcomes. The implications of these results could guide educators and policymakers in developing more effective interventions to support student success.*

Keywords *- Academic performance, Student success prediction, Predictive models in education, Influence of study hours, Academic history.*

1. Introduction

Education is a fundamental pillar in individual and social development, where academic performance plays a critical role. In recent years, the prediction and analysis of student performance have become an area of increasing interest for researchers and educators, aiming to understand better the factors that contribute to academic success or failure [1]. Previous studies have explored a variety of variables, from psychological and socioeconomic factors to teaching methods and study habits, demonstrating the complexity of academic performance [2-4]. Among these factors, the hours dedicated to study and the scores of previous exams have been identified as significant indicators of academic success [5]. However, the relationship between these variables and their precise predictive capacity has not yet been fully defined. This study seeks to fill this gap in the literature, using statistical analysis techniques and machine learning to assess their impact on exam outcomes. The objective of this research is twofold: first, to analyze the correlation between study hours, scores from previous exams, and success in subsequent exams, and second, to develop a predictive model that educators can use to identify students at risk of underperformance. Through this approach, we aspire to offer practical insights for the improvement of educational strategies and the personalization of academic support. This document is organized as follows: Section II reviews the relevant literature in the area of student performance prediction. Section III describes the methodology employed, including data collection, statistical analysis, and the development of the predictive model. Section IV presents the obtained results, followed by a discussion in Section V on the implications of these findings. Finally, Section VI concludes the study and proposes future directions for research.

2. Literature Review

The academic performance of students has been the subject of extensive analysis in educational literature. Factors such as cognitive skills, socioeconomic environment, learning strategies, and emotional well-being have been identified as significant determinants of academic success [6-8]. Among these, study hours and results from previous exams emerge as key variables directly related to academic performance [9], [10]. The amount of time dedicated to study has been positively correlated with academic performance in multiple studies.

Research by [11] demonstrates that students who spent more time studying outside of class tended to achieve better grades on their exams. However, the relationship is not always linear, as indicated by [12], who found that beyond a certain threshold, an increase in study hours did not translate into a significant improvement in performance. Results from previous exams have been used as predictors of future performance. In a study conducted by [13], it was observed that scores from previous exams were strong indicators of success in subsequent exams, suggesting they reflect both the student's accumulated knowledge and study skills. With the advancement of information technology, predictive models have gained popularity in the educational field to identify students at risk of underperformance and personalize interventions. The work of [14] applied machine learning models to predict student performance based on variables such as study hours, class participation, and results from previous exams, finding significant accuracy in their predictions.

Despite advances in understanding the factors influencing academic performance, there are gaps in the literature, particularly in assessing how interactions between different variables affect student outcomes. Additionally, there is a need to explore further the role of learning techniques and the quality of study beyond the quantity of hours dedicated.

3. Methodology

This study adopts a quantitative approach to examine the relationship between study hours, results from previous exams, and performance in future exams to develop a predictive model of students' academic success.

3.1. Population and Sample

The target population for this study consists of secondary-level students from various schools in Lima, Peru. Due to confidentiality reasons, the educational institutions have decided that the students' data should remain anonymous. This information was collected during the 2023 academic year. A random sample of 500 students was selected, ensuring representativeness in terms of gender, age, and previous academic performance. Table 1 displays the data characteristics for the study.

Table 1. Team roles

Characteristics	Description
Total Number of Students	500
Age Range	[12 – 17]
Previous Academic Performance	**High Performance:** Students with scores in the 75th percentile and above, with grades over 85, representing the top 25% of the sample. These students have demonstrated an exceptional understanding of previous content and possess strong academic foundations. **Medium Performance:** Students whose scores are between the 25th and 75th percentiles, with grades from 70 to 85. This group constitutes the majority of the sample and reflects a competent mastery of the content, with room for growth and improvement. **Low Performance:** Students in the 25th percentile and below, with scores below 70, identifying 25% of the sample with greater academic challenges. These students could significantly benefit from targeted interventions and additional support.

3.2. Data Collection

Data was collected through academic records provided by the institution, including the study hours self-reported by students in the week leading up to their final exams and their scores on these exams. Additionally, students' scores from previous exams were collected to analyze their relationship with performance on future exams.

3.3. Variables

Independent Variable 1: Study hours, measured as the total number of hours dedicated to study in the week before the exam. Independent Variable 2: Previous exam score, measured on a scale from 0 to 100. Dependent Variable: Exam performance, categorized as 'passed' (1) if the student scored 60 or above and 'failed' (0) otherwise.

3.4. Data Analysis

To analyze the relationship between variables, descriptive analysis was used to characterize the sample. Subsequently, logistic regression techniques were applied to assess how study hours and scores from previous exams influence the probability of passing an exam. Finally, a machine learning model was developed using a random forest algorithm to predict students' performance on future exams based on these variables.

3.5. Analysis Tools

Data analysis was performed using the statistical software R and Python, leveraging specialized libraries such as pandas for data manipulation and scikit-learn for predictive modeling.

3.6. Techniques Applied to the Study

In this research, techniques for processing, transforming, and analyzing the data were utilized with the ultimate goal of obtaining predictive results regarding the investigation. Therefore, the following steps were followed.

3.6.1. Data Preparation and Preprocessing

Initially, data cleaning and preparation were conducted using pandas, a Python library, for data manipulation and analysis. This included removing duplicate or incomplete records, normalizing scores from previous exams, and transforming study hours into a continuous numerical variable. Additionally, encoding was applied to the target variable "Pass/Fail" to convert it into a binary format suitable for analysis.

3.6.2. Exploratory Data Analysis (EDA)

An Exploratory Data Analysis (EDA) was conducted using Matplotlib and Seaborn to visualize distributions, identify potential correlations between variables, and better understand the data structure. This included generating histograms, box plots, and correlation heat maps. This step was crucial for designing effective modeling strategies.

3.6.3. Feature Selection and Transformation

To enhance the predictive capability of the models, feature selection was conducted based on the importance of variables determined by a preliminary random forest model (RandomForestClassifier from scikit-learn). Additionally, feature transformation techniques, such as the standardization of numerical variables, were applied to optimize the model performance.

3.6.4. Development of Predictive Models

Several predictive models were experimented with to estimate the probability of success in exams based on study hours and scores from previous exams. The models included.

Logistic Regression

Used as a base model due to its simplicity and effectiveness in binary classification.

Random Forests

Selected for its ability to handle nonlinear relationships and its importance in feature selection.

Vector Support Machines (SVM)

Implemented with a linear kernel and RBF (Radial Basis Function) to explore different feature spaces.

Neural Networks

Keras was used on TensorFlow to develop a simple neural network model, exploring its ability to capture complexities in data. For each model, cross-validation was performed to ensure robustness and generalizability. Evaluation metrics such as accuracy, sensitivity, specificity, and Area Under the ROC Curve (AUC) were used to compare model performance.

3.6.5. Model Optimization and ValidationSeveral

Using scikit-learn's GridSearchCV, an exhaustive hyperparameter search was performed for each model to find the optimal configuration. Final validation of the models was carried out on a separate test dataset to assess their performance on unseen data.

3.6.6. Interpretation of Results

The interpretation of the models focused on the importance of the characteristics and their impact on the probability of passing an exam. SHAP (SHapley Additive exPlanations) was used to explain the predictions of the more complex models, providing insights into how study hours and previous exam scores influence the model's predictions.

4. Result

This study assessed the ability of study hours and previous exam scores to predict students' academic performance using two modeling approaches: Logistic Regression and Random Forests. These models were applied to a dataset of 500 student records, seeking to identify significant patterns and predictors of academic success.

4.1. Descriptive Analysis

Initially, a descriptive analysis was performed to understand the general characteristics of the sample. Students spent an average of 5.49 hours studying before exams, with pre-exams averaging 68.92. Visual exploratory analysis indicated a varied distribution in both variables, suggesting significant individual differences in study habits and previous academic performance. Figure 1 shows how study hours are distributed between students who passed and those who did not. You can look at differences in the means and dispersion of study hours, which can indicate whether longer study time is associated with a higher chance of passing. Figure 2 compares previous test scores between students who passed and those who did not. This analysis can reveal the influence of previous exam scores on current academic performance, highlighting the possible relationship between accumulated knowledge and success in future exams. The correlation matrix, as shown in Figure 3, shows the relationships between study hours, previous test scores, and academic performance. Correlation coefficients can provide a quantitative understanding of how these variables are related to each other.

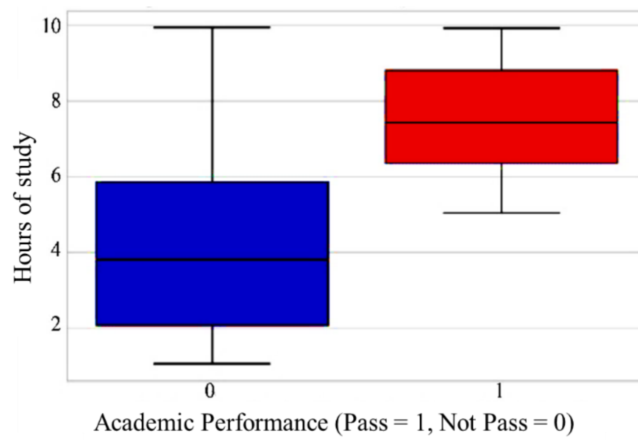

Fig. 1 Distribution of study hours by academic performance

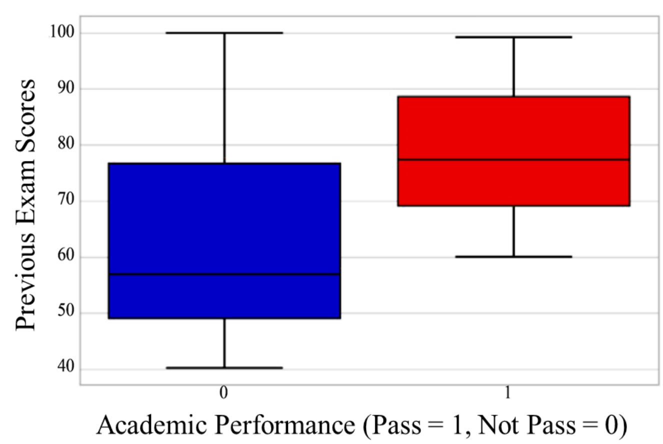

Fig. 2 Distribution of previous exam scores by academic performance

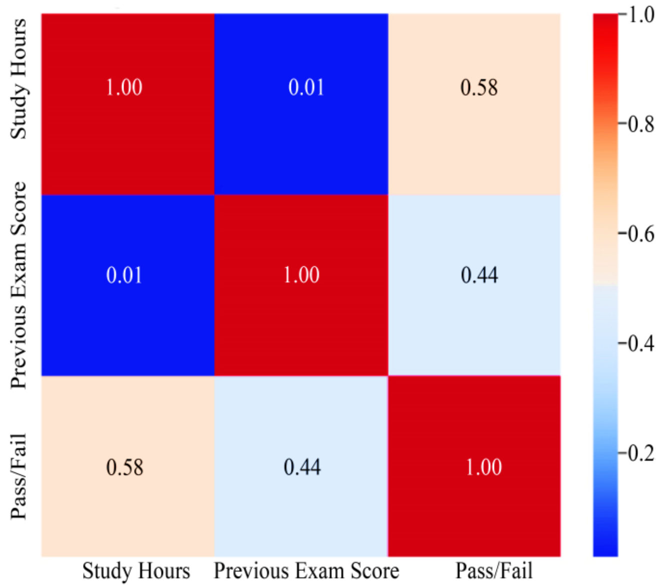

Fig. 3 Correlation matrix between variables

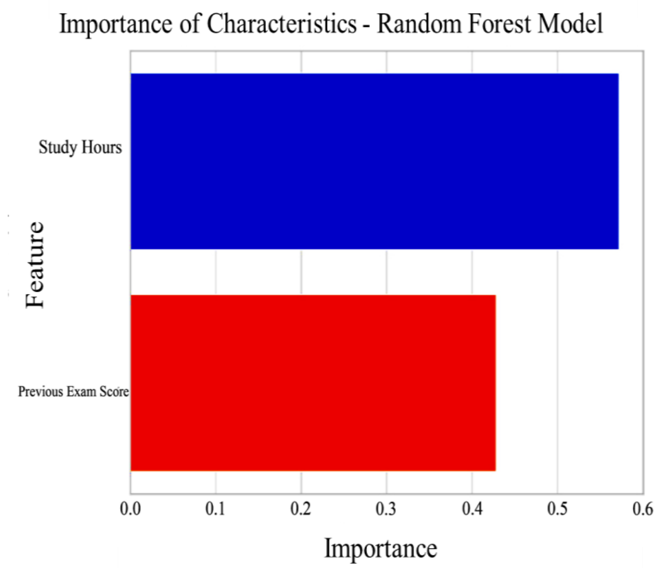

Fig. 4 Importance of features in the predictive model of random forests

4.2. Model Performance

4.2.1. Logistic Regression

The model demonstrated an accuracy of 86% and an AUC of 0.93, indicating a robust ability to predict academic success.

Logistic regression allowed us to identify a significant relationship between study hours, previous exam scores, and the probability of passing an exam.

4.2.2. Random Forests

This model exhibited a superior accuracy of 93% and an AUC of 1.0, standing out for its exceptional predictive ability. The importance of the characteristics revealed that, although both variables are valid predictors of academic performance, previous test scores carry a greater weight in the prediction.

Figure 4 shows the relative importance of study hours and previous exam scores in predicting academic performance using the Random Forests model. Importance is measured in terms of the impact of each feature on the accuracy of the model.

4.2.3. Model Comparison

Comparison between the models suggests that Random Forests offer a significant improvement in predictive ability over Logistic Regression for this dataset. This result can be attributed to the ability of Random Forests to capture complex interactions between variables that are not easily modeled by logistic regression.

Figure 5 shows the relative importance of study hours and previous exam scores in the Random Forests model. It is observed that previous exam scores have a greater importance in predicting academic performance, indicating their predictive value on study hours.

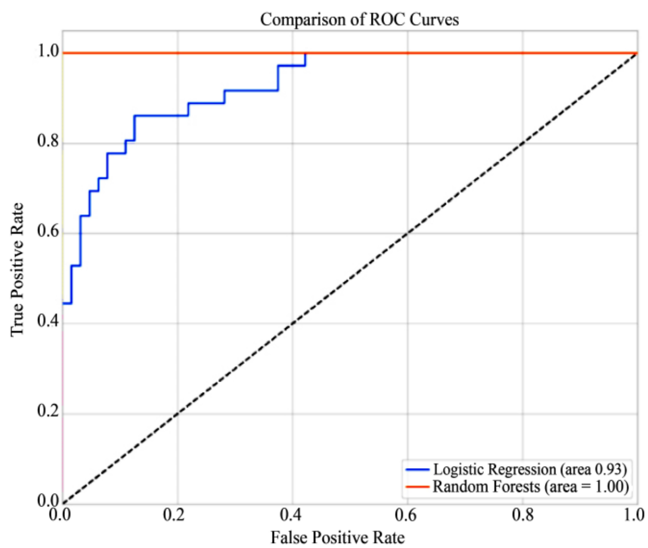

Fig. 5 Comparison of ROC curves for logistic regression models and random forests

4.2.4. Interpretation of Importance of Features

Analysis of the importance of features in the Random Forests model revealed intriguing findings about predictors of academic achievement. As illustrated in Figure 2, previous exam scores emerged as the most significant predictor, surpassing hours of study. This result not only underscores the critical role of prior academic preparation but also highlights the predictive ability of past achievements over future academic successes.

Table 2. Importance of characteristics in the predictive model of random forests

Feature	Importance
Study Hours	0.5719
Previous Exam Scores	0.4281

Table 2 shows that, although both characteristics are important in predicting academic performance, study hours have a slightly higher weight in this specific model. This result suggests that, within the context of our analysis, the amount of time spent studying is a stronger predictor of academic performance than previous exam scores.

The predominance of previous test scores suggests that educational interventions that focus on improving continued academic achievement may be particularly effective. On the other hand, although study hours were identified as an important factor, their lower relative weight indicates that the quality and not just the quantity of study could be a key area for future research and pedagogical strategies.

5. Discussion

This study explored the relationship between study hours, previous exam scores, and academic performance using Logistic Regression and Random Forest models. The results indicate that both variables are significant predictors of exam success, although with different degrees of influence.The findings support previous research underscoring the importance of previous exam scores as a robust indicator of future performance [15]. However, unlike certain studies that highlight the predominance of study hours [9], this analysis reveals that although study hours are important, previous exam scores carry considerable weight in predicting academic performance. This result may reflect the accumulation of knowledge and skills over time beyond immediate preparation before an exam.The identification of previous test scores as the strongest predictor suggests that educational interventions should not only focus on short-term study strategies but also ongoing support throughout academic cycles. This could include mentoring programs, monitoring learning progress, and implementing adaptive pedagogical strategies that respond to individual student needs.This study has several limitations. First, the sample is limited to students at a specific institution, which may affect the generalizability of the results. Second, the study is based on self-reported data from study hours, which may be subject to memory or perceptual biases. Future research should consider using more objective methods to measure study time and expand the sample to diverse student populations. Future research could explore the interaction between study hours and previous test scores in more detail, as well as include other variables that may influence academic performance, such as psychosocial factors and the learning environment. In addition, it would be valuable to examine the effectiveness of different intervention strategies based on the findings of this study to develop more specific and effective pedagogical recommendations.

6. Conclusion

This study addressed the relationship between study hours, previous exam scores, and student's academic performance using statistical analysis methods and advanced predictive models. The findings reveal that although both hours of study and previous test scores are significant predictors of test performance, past scores have a greater influence on predicting academic success.

Past Test Scores: This study confirms the importance of past test scores as a robust predictor of academic performance, suggesting that academic history provides a solid foundation for projecting future success.Although study hours also correlate with academic performance, their impact is relatively minor compared to previous test scores. This result underscores the need to focus educational interventions not only on the quantity of study but also on improving the quality and effectiveness of learning. This study contributes to the understanding of the factors that influence academic performance, offering empirical evidence on the relevance of considering both academic history and current study practices. The predictive models developed can serve as tools to identify students at risk of underachievement, enabling early and personalized interventions. Educators and administrators can use these findings to design academic support programs that focus on strengthening students' knowledge base and optimizing study strategies. In addition, this study emphasizes the importance of monitoring academic progress over time, suggesting that a longitudinal approach in education may be key to improving learning outcomes. It is recommended to conduct future studies with larger and more diverse samples and to explore other potentially influential variables, such as specific study techniques, emotional well-being, and socioeconomic factors. Investigating the implementation and effectiveness of different educational interventions based on risk profiles will also be crucial to advancing the personalization of student learning and support.

References

[1] James S. Cole, David A. Bergin, and Tiffany A. Whittaker, “Predicting Student Achievement for Low Stakes Tests with Effort and Task Value,” *Contemporary Educational Psychology*, vol. 33, no. 4, pp. 609-624, 2008. [CrossRef] [Google Scholar] [Publisher Link]

[2] Fernando Robert Ferrel Ortega, Juana Vélez Mendoza, and Lucia Fernanda Ferrel Ballestas, “Psychological Factors in School Adolescents with Low Academic Performance: Depression and Self-Esteem,” *Encuentros*, vol. 12, no. 2, pp. 35-47, 2014. [Google Scholar] [Publisher Link]

[3] Lorraine C. Taylor, Jennifer D. Clayton, and Stephanie J. Rowley, “Academic Socialization: Understanding Parental Influences on Children’s School-Related Development in the Early Years,” *Review Of General Psychology*, vol. 8, no. 3, pp. 163-178, 2004. [CrossRef] [Google Scholar] [Publisher Link]

[4] Nahid Shirani Bidabadi et al., "Effective Teaching Methods in Higher Education: Requirements and Barriers," *Journal of Advances in Medical Education and Professionalism*, vol. 4, no. 4, pp. 170-178, 2016. [Google Scholar] [Publisher Link]

[5] Jocelyne Matar Boumosleh, and Doris Jaalouk, "Depression, Anxiety, and Smartphone Addiction in University Students- A Cross Sectional Study," *Plos One*, vol. 12, no. 8, pp. 1-14, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[6] Diego Danilo Rodriguez Rosero, Ruber Erlinton Ordonez Ortega, and Mario Eduardo Hidalgo-Villota, "Determinants of Academic Performance in Secondary Education in the Department of Nariño, Colombia," *Economics Readings*, no. 94, pp. 87-126, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[7] Villena Alcantara, and Mirian Noemy, "*Factors Determining the Academic Performance of University Students in Peru, between 2009 and 2013*," National University of Trujillo, pp. 1-54, 2017. [Google Scholar] [Publisher Link]

[8] Andrés Escarbajal Frutos, Marta Sánchez, and Isabel Guirao, "Factors Determining the Academic Performance of Students who are Children of Immigrants and Natives of Foreign Origin in Contexts of Social Exclusion," *Journal of Childhood and Adolescence*, no. 9, pp. 1-16, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[9] Cindy Rodriguez Tapia et al., "Analysis of the Impact on the Academic Performance of Scholarship Students Taking Into Account the Hours of Study and Grades of a Group from the Autonomous Regional University of the Andes," *Contemporary Dilemmas: Education, Politics and Values*, no. 1, pp. 1-17, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[10] María Dolores Cáceres Zapatero, and Enrique Morales Corral, "Chapter 6.4. Case Study on 24-Hour Media Disconnection in Spanish University Students," *Mirror of Monographs of Social Communication*, no. 2, pp. 287-301, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[11] Ramon Cladellas et al., "Effects of Sleeping Hours and Sleeping Habits on the Academic Performance of Six- and Seven-Year-Old Children: A Preliminary Study," *Culture and Education*, vol. 23, no. 1, pp. 119-128, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[12] Eva Fita Lladó, Sebastian Rodriguez Espinar, and Mercè Torrado Fonseca, "Academic Performance in the Transition from Secondary School to University," *Journal of Education*, no. 334, pp. 391-414, 2004. [Google Scholar] [Publisher Link]

[13] Á. Menéndez-Aller et al., "Academic Resilience: the Influence of Effort, Expectations and Academic Self-Concept," *Latin American Journal of Psychology*, vol. 53, pp. 114-121, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[14] Javier Mella Norambuena, María Graciela Badilla-Quintana, and Yaranay López Angulo, "Predictive Models Based on the Use of Learning Analytics in Higher Education," *Texto Livre: Linguagem E Tecnologia*, vol. 15, pp. 1-22, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[15] Manuel Ángel Río Ruiz, and MaRía luisa JiMénez RodRigo, "6000 Scholarships Under Review, Results, Practices, Expectations and School Opportunities for Families and Students Included in the Program," *International Journal of Sociology*, vol. 72, no. 3, pp. 609-632, 2014. [CrossRef] [Google Scholar] [Publisher Link]