Student Result Analysis

Muskaan Fatima Karachi, Pakistan

*Abstract*—This report scrutinizes the influence of a Student Academic performance Analysis that uses ethnic group, gender, parents' education, marital status, daily sports activity, test preparation, and weekly study hours to better understand the factors contributing to students’ academic performance encompassing math grades, reading scores, and writing scores. Applying machine learning algorithms, the Student Result Analysis uses students’ scores and other variable to gauge the relationship between academic performance and different factors affecting it. Through a detailed examination of a student dataset encompassing math grades, reading scores, and writing scores which includes a wide range factors such as ethnic group, gender, parents' education, marital status, daily sports activity, test preparation, and weekly study hours, aiming to discern patterns and associations that contribute to academic success. This research offers a workable answer by providing suitable and meaningful predictions. Student Academic performance Analysis can improve understanding of the relationships between these variables and explore further variables that contribute to academic success.

*Keywords*—Machine Learning Algorithms, User Preferences, Performance Analysis, Data Analysis, Academic Success, Educational Research

# Introduction

In today’s dynamic educational environment, comprehending the complex dynamics that plays a crucial role in student academic performance is of utmost significance. The pursuit for academic success is a journey influenced by a complicated interplay of socio-demographic factors that spreads beyond traditional settings. Student Performance analysis system uses a large amount of data that has essential fields such as ethnic group, gender, parents' education, marital status, daily sports activity, test preparation, and weekly study hours. Reliable evaluations allow users to understand the impacts of such variable on an individual. Each students’ ID serve as distinct labels, with their math scores, reading scores, and writing scores.

This analysis embarks on a comprehensive study of the influence of various variables on students’ performance, delving into the delicate relationships between academic outcomes and factors such as ethnic group, gender, parents' education, marital status, daily sports activity, test preparation, and weekly study hours. The approaches used to deal with the data columns are covered in this paper below, providing an understanding of how each element affects the students’ performance and the extent to which they are accurate. The results and findings are presented in the following parts to help recognize how student academic performance analysis accomplishes its objectives of bringing change in a student’s lifestyle to ensure their academic success.

# Problem Statement

A student's achievement in this educational environment, or schooling, is influenced by several societal factors. Understanding how these elements impact academic success is crucial for educators, politicians, scholars, and legislators. This study aims to identify:

1. To what degree are math grades, reading scores, and writing scores influenced by factors including gender, ethnicity, marital status, parents' educational attainment, level of daily sports participation, test-taking strategies, and weekly study hours?
2. Do differences or patterns in student performance between various sociodemographic groups stand out?
3. In what ways might knowledge have gained from the analysis aid in the creation of evidence-based treatments and educational support systems?

# Literature Review

According to the studies, machine learning techniques and data analysis in educational backgrounds has gained a standing in recent years. Understanding the features that influence student academic performance and utilizing advanced data analysis that can provide valuable insights for individuals and educational institutions. This literature review provides the insights of the key frames related to educational data analysis, student’s educational performance, and the use of machine learning in predicting these trends.

In addition, the advent of learning platforms and student management systems has enabled the collection of huge amounts of student’s academic data. Researchers have identified trends in academic data analysis to discover patterns, identify risks, and enhance predictive outcomes. Research has examined the multiple factors that influence student performance; these factors include social background, parental involvement, study practices, and additional activities. This study offers an in depth understanding of why and how a student’s performance is impacted by the socio-demographical features.

# Methodology

### Data Collection: We have taken the dataset from Kaggle, a website providing different variety of datasets, and in this research we are using the dataset Student Academic Results. This dataset spans math grade, writing grade, reading grade, ethnic group, gender, parents' education, marital status, daily sports activity, test preparation, and weekly study hours. Importantly, the dataset includes data more than ten thousand students.

### Data Preprocessing: Initially, our dataset comprised 30,642 rows and 11 columns. To enhance the data quality, we employed both Python and Power BI to address various issues. The following steps were taken:

#### Duplicate Table for Further Filtering: To avoid disturbing the original dataset, a duplicate table was created for subsequent filtering on rows columns.

#### Empty Row Removal: We analytically eliminated empty rows using Python.

#### Column Filtering: Useless columns such as "Unnamed4” was removed to simplify the dataset.

#### Grading: Math, Reading and Writing Scores were converted into grades to identify the number of passed and failed students.

### Subject-Specific Analysis: For Math, Reading, and Writing; maximum, minimum, and mean scores are calculated and visualized using tables and charts. Students failing or passing in each subject are identified. Box plots and pie charts are generated to envision score distributions.

### Machine Learning Algorithm and Data Analysis: Random Forest Regressor is used to predict Overall Grades. The model is evaluated using Mean Squared Error (MSE) and R-squared which is then visualized to understand factors influencing grades.

### Predictions and Insights: A sample of data is used for prediction. Overall grades are projected for the entire dataset using the trained model.

### Additional Analysis: Heatmaps are used to visualize the relationship between selected categories of variables and scores.

# Conclusion

In conclusion, an overview of the observations, and overall effects of various factors on student performance has been explained and visualized. To sum up, this study offers insightful information about the complicated web of socio-demographic variables affecting students' academic performance. The results emphasizes the need of modified educational approaches that take variable traits into account. This approach offers a thorough explanation of the analysis and machine learning technology used in the code. In order to extract meaningful insights from the dataset, data exploration, cleaning, visualization, and model evaluation has been done to examine variables that affect academic success and dive deeper into the multifaceted interactions between these features.

# Results

The Resulting analysis and findings are consistent with the research on how factors such as ethnic group, gender, parents' educational attainment, marital status, daily sports activity, test preparation, and weekly study hours affects students' outcomes. The notion that ‘physical activity might have a good impact on cognitive ability’ is supported by the relation found between reading scores and daily sports participation. Test preparation has been shown to have an influence on writing scores, which emphasizes the importance of focused involvements for particular areas.

Various tabulate, charts, graphs and matplotlib visualizations are used to display the findings, insights, and meaningful predictions.

##### References

1. Granata, F. Evapotranspiration evaluation models based on machine learning algorithms—A comparative study. Agric. Water Manag. 2019, 217, 303–315. [Google Scholar] [CrossRef].
2. Jansen, S. Machine Learning for Algorithmic Trading: Predictive Models to Extract Signals from Market and Alternative Data for Systematic Trading Strategies with Python, 2nd ed.; Packt Publishing: Birmingham, UK, 2020; ISBN 978-1839217715. [Google Scholar].
3. George AF Seber and Alan J. Lee, Linear regression analysis, John Wiley & Sons, vol. 329, 2012.