**Introduction Student**

Performance analysis is a crucial area of educational research, aiming to identify factors that influence academic success. Traditionally, performance evaluation relies on grades and teacher assessments, but with advancements in data science, machine learning provides an automated approach to analyze large educational datasets.

The dataset used in this study is sourced from a Portuguese secondary school and contains 649 student records with 33 attributes related to demographics, family background, study habits, and academic performance. The key objective of this study is to explore the dataset using Exploratory Data Analysis (EDA) and predictive modeling to identify patterns influencing student success.

**Articles**

* <https://link.springer.com/article/10.1007/s10639-020-10230-3>
* <https://ieeexplore.ieee.org/abstract/document/7407798/>
* <https://link.springer.com/chapter/10.1007/978-981-13-0514-6_71>
* <https://ieeexplore.ieee.org/abstract/document/7407798/>

**Abstract Education**

Plays a critical role in shaping future careers, and student performance analysis helps in understanding the key determinants of academic success. Traditional approaches to evaluating student performance rely on teacher assessments and grades. However, data-driven techniques provide more objective and scalable insights. This study analyzes a dataset containing 649 student records with 33 attributes related to social, economic, and educational factors. Using EDA, correlation analysis, and predictive modeling, we aim to identify the most significant features influencing students' final grades. Machine learning algorithms such as Linear Regression, Decision Trees, and Random Forest are applied to predict academic outcomes. The findings highlight the importance of parental education, study time, and social engagement in determining student success.

**Methodology Data Collection**

The dataset consists of records of Portuguese students in secondary education, containing:

* *Demographics*: Age, gender, family background.
* *Parental Background:* Education level, job type.
* *Study Habits*: Study time, failures, extracurricular activities.
* *Academic Performance*: First-period, second-period, and final grades (G1, G2, G3).

**Data Cleaning**

* Checked for missing values and handled them appropriately.
* Converted categorical variables into numerical labels for analysis.
* Identified and removed duplicates.

**Exploratory Data Analysis (EDA)**

* Performed statistical analysis to examine the relationship between variables.
* Used visualization techniques (bar charts, heatmaps) to identify patterns in student performance.
* Conducted correlation analysis to determine the impact of various factors on grades.

**Machine Learning**

* Models Linear Regression: Used to predict students' final grades.
* *Decision Tree:* Applied to capture complex interactions between variables.
* *Random Forest*: Evaluated for its robustness in predicting student performance.

**Evaluation Metrics**

* Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) were used to assess model performance.
* Chi-Square test applied to analyze relationships between categorical features.

**Literature Review**

* Machine Learning has been widely used in educational analytics to predict student performance.
* Studies show that Random Forest performs well in handling complex relationships.
* Parental education and student study habits are strong indicators of academic success.

**Findings**

* *Parental Education:* Strong correlation with student performance; higher parental education leads to better grades.
* *Study Time:* Moderate correlation; students who study regularly tend to perform better.
* *Alcohol Consumption:* Negative impact on student grades, especially excessive weekend drinking.
* *Absences:* High absences correlate with lower academic performance.
* Random Forest emerged as the best model, achieving the highest accuracy in predicting final grades.

**Conclusion**

This study successfully applied machine learning models to predict student performance based on various socio-economic and academic factors. Among the models tested, Random Forest performed the best, effectively capturing complex relationships between variables. The insights gained from this analysis can aid educators in identifying at-risk students and implementing targeted interventions to improve academic outcomes.