

Dataset Analysis Report

Dataset contains demographic, academic and behavioral features of students.

Target variable selected: GPA (Regression).

1. Dataset Overview

- Machine Learning Task: Regression
- Target Variable: GPA (0.00 – 4.00)
- Contains academic, support and activity indicators

2. Feature Types

Numerical: Age, StudyTimeWeekly, Absences, GPA

Binary: Tutoring, Extracurricular, Sports, Music, Volunteering, Gender

Categorical: Ethnicity

Ordinal: ParentalEducation, ParentalSupport, GradeClass

Identifier: StudentID

3. Missing Values

Dataset contains 0 missing values across all features.

4. Observations

- Higher StudyTimeWeekly generally correlates with higher GPA
- Higher Absences tends to reduce GPA
- Tutoring and ParentalSupport positively influence academic performance
- Extracurricular and Sports participation represent behavioral factors

5. ML Suitability Assessment

Advantages:

- Clean dataset (no missing values)
- Relevant predictive features
- Real-world educational context
- Continuous numeric target suitable for regression

Challenges:

- Requires encoding of categorical and ordinal variables
- StudentID must be excluded for modeling

Conclusion:

The dataset is suitable for machine learning regression tasks for predicting GPA after encoding and preprocessing.