Report: SHAP Analysis on Student Performance Dataset

ARCHITHA OJHA

2303A52170

BATCH 41

1. Introduction

The objective of this project is to apply SHAP (SHapley Additive exPlanations) to interpret a machine learning model for predicting student performance. We aim to identify the most important factors influencing student grades and evaluate the predictive power of the chosen model.

2. Dataset Description

Source: Kaggle – Students Performance in Exams dataset

Size: ~1000 rows

Features: Gender, race/ethnicity, parental education level, lunch type, test preparation course, math/reading/writing scores.

Target Variable: Grade (categorical: A, B, C, etc. – derived from scores).

3. Preprocessing

Removed irrelevant columns (e.g., IDs).

Encoded categorical variables using Label Encoding / One-Hot Encoding.

Handled missing values (none found).

Standardized numerical features using StandardScaler.

4. Model & Performance:

Accuracy: ~85%

Classification Report: High precision/recall for major classes, lower for minority classes.

5. SHAP Analysis :

Explainer Used: TreeExplainer

Plots Generated:

Summary Plot → Shows the global feature importance.

Force Plot → Highlights individual student predictions.

Waterfall Plot → Step-by-step contribution of features for one prediction.

Comparison with Model's Built-in Feature Importance:

Both SHAP and Random Forest importance agree that academic scores are the strongest predictors.

SHAP adds interpretability by showing direction of influence (e.g., higher math score increases probability of good grade).

6. Conclusion:

Academic scores (math, reading, writing) are the most decisive factors in predicting student performance.

SHAP enhances interpretability compared to traditional feature importance.