

Student Performance Classification Report

Dataset: UCI Student Performance dataset (student-mat.csv)

1. Data Preprocessing:

- Missing values: None (dataset is complete).
- Feature encoding: One-hot encoding applied to categorical variables (sex, address, family size, parents' education, etc.).
- Target definition: Binary 'pass_fail' where pass if final grade G3 \geq 50.

2. Model Choice:

- Logistic Regression: chosen for its interpretability and suitability for binary classification.

3. Pipeline:

- Load and inspect data.
- Create target and drop original grade column.
- Encode categorical features and separate X/y.
- Train/test split (70/30) with random_state=42.
- Train logistic regression (max_iter=1000).

4. Evaluation:

- Metrics: Accuracy, Precision, Recall, F1 Score, Confusion Matrix.
- (Results to be computed in the notebook.)

5. Conclusion:

The logistic regression model will indicate how well basic demographic and study-related features predict student success.

Code and full pipeline available in Student_Classification.ipynb.