Bias Detection Report

Bias Detection Report

Analysis Summary

This report provides an analysis of potential bias in the dataset 'Student Performance.csv' focusing on the correlation between internet access at home (categorical feature) and final grades (numerical feature).

Dataset and Features

- **Dataset:** Student Performance.csv

- **Features Analyzed:**

  - `internet` (categorical)

  - `G3` (numerical, final grades)

Bias Type

- **Correlation** between the categorical feature `internet` and the numerical feature `G3`.

Detection and Analysis Methods

- **Toolset Method:** Used the `categorical_numerical_correlation_standardized_difference` tool to calculate the standardized difference.

- **Reference Literature (B-01-3):** Calculated the Standardized Difference (SD) for each category in the `internet` feature.

Results

- **Toolset Result:**

  - Standardized Difference of -0.335 for "no internet" suggests minimal bias.

- **Reference Method Result:**

  - SD for "no internet": -0.219

  - SD for "yes internet": 0.044

Both results align with a minimal bias level.

Visualizations

Below are the visualizations illustrating the distribution of final grades among students with and without internet access.

Conclusion and Recommendations

The analysis indicates a minimal correlation bias between internet access and final grades. The dataset can be utilized with confidence, considering slight variations in specific scenarios.