

Bias Detection Report

Bias Detection Report

Introduction

This report explores potential biases in students' plans for higher education, focusing on distribution and correlation biases in the 'Student Performance' dataset.

Data Preprocessing

Relevant features were extracted for analysis, including 'higher', 'school', 'age', and 'studytime'.

Distribution Bias Analysis

- **Shannon Entropy for 'higher'**: 0.2004
- **Maximum Entropy**: 0.6931
- **Potential Bias Detected**: True

The 'higher' feature shows potential distribution bias, with a significantly lower Shannon Entropy than the maximum entropy.

Correlation Bias Analysis

- **Chi-Square Test (higher vs school)**:
- Chi-Square Statistic: 0.0149

- p-value: 0.9027
- Significant Association: False

No significant association was found between 'higher' and 'school'.

- **Wasserstein-2 Distance**:
- **Age**: 1.2153 (Notable distributional difference)
- **Studytime**: 0.6693 (Moderate distributional difference)

These values indicate notable differences in age and moderate differences in studytime distribution based on higher education plans.

Conclusions

The analyses suggest moderate bias in the dataset:

- **Bias Type**: Distribution and Correlation
- **Relevant Features**: 'higher', 'age', 'studytime'
- **Overall Bias Level**: Moderate Bias

While there's no significant correlation with 'school', notable distributional differences exist with 'age' and 'studytime'.

Recommendations

Consider these findings in educational policy planning and intervention strategies to address potential biases in higher education plans.