Bias Detection Report

Bias Detection Report

Dataset Analyzed

The dataset analyzed is from the COMPAS dataset, focusing on the relationship between ethnicity and recommended supervision levels.

Features Examined

- Ethnicity (`Ethnic_Code_Text`)

- Recommended Supervision Level (`RecSupervisionLevelText`)

Types of Bias Detected

The analysis focused on correlation bias between the two categorical features.
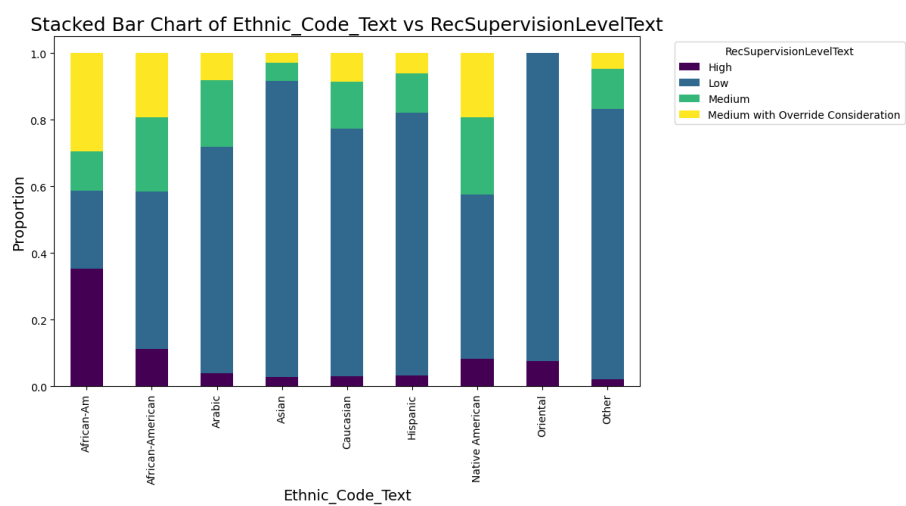
Tools Used

1. **Cramér's V**

2. **Elift**

3. **Statistical Parity Analysis (J-divergence)** from reference literature.

Values Obtained and Extent of Bias

- **Cramér's V**: 0.1804 (Minimal Bias)

- **Elift**: 5.2788 (Extreme Bias)

- **J-divergence**: Ranges from minimal to extreme values, indicating varying levels of bias across different ethnic pairs.

Visualizations

The stacked bar chart below illustrates the distribution of recommended supervision levels across different ethnic groups.

Interpretation of Bias Severity

The tools suggest a range of bias levels:

- **Minimal Bias**: Cramér's V indicates minimal association.

- **Extreme Bias**: Elift and some J-divergence results show significant bias.

Recommendations

- Carefully consider the implications of the biases detected, especially the extreme biases highlighted by Elift and J-divergence.

- Further investigation or adjustments may be needed to ensure fairness in decision-making processes involving these features.