Bias Detection Report

Bias Detection Report for Feature: Ethnic_Code_Text

Overview:

This report examines potential distribution and correlation biases related to the 'Ethnic_Code_Text' feature in the COMPAS dataset.

Distribution Bias Analysis:

1. Chi-Square Test:

   - Chi-Square Statistic: 129646.89

   - Critical Value: 11.07

   - Conclusion: Significant distribution bias, as the statistic is much higher than the critical value.

2. Kolmogorov-Smirnov Test:

   - KS Statistic: 0.333

   - p-value: 0.730

   - Conclusion: The test does not indicate significant distribution bias.

Correlation Bias Analysis:

1. Chi-Square Test (with 'Sex_Code_Text'):

- Chi-Square Statistic: 219.46

- p-value: 5.00e-43

- Conclusion: Significant correlation bias between 'Ethnic_Code_Text' and 'Sex_Code_Text'.

2. HSIC Analysis:

- HSIC Value: 1.31e-18

- Conclusion: No significant correlation bias detected by HSIC.

Summary:

- Distribution Bias Level: Extreme Bias (Level 5) based on Chi-Square Test.

- Correlation Bias Level: Significant Bias (Level 4) based on Chi-Square Test, mitigated by HSIC results.

Recommendations:

- Further analysis may be considered for additional features.

- Caution is advised when using this dataset due to the identified biases.