Bias Detection Report

Bias Detection Report

Dataset Analyzed

The dataset used for this analysis is the MIMIC-IV dataset, which includes features such as admission type, hospital expiration flag, patient race, and more.

Features Examined

The primary features examined in this analysis are 'patient_race' and 'hospital_expire_flag'.

Types of Bias Detected

The analysis focused on detecting correlation bias between the categorical features of patient race and hospital mortality rates.

Tools and Methods Used

- **Toolset Methods:** Cramér's V, Elift, and Statistical Parity methods were planned but not executed due to initial data type issues.

- **Reference Literature Method:** Wasserstein-2 Distance was successfully applied to evaluate distributional differences between categories.

Values Obtained and Extent of Bias

The Wasserstein-2 distances ranged from 0 to 0.199, indicating varying levels of bias between different patient race categories and hospital mortality rates. Distances close to 0 suggest minimal bias, while higher distances indicate more pronounced bias.

Stacked Bar Chart of patient_race vs hospital_expire_flag
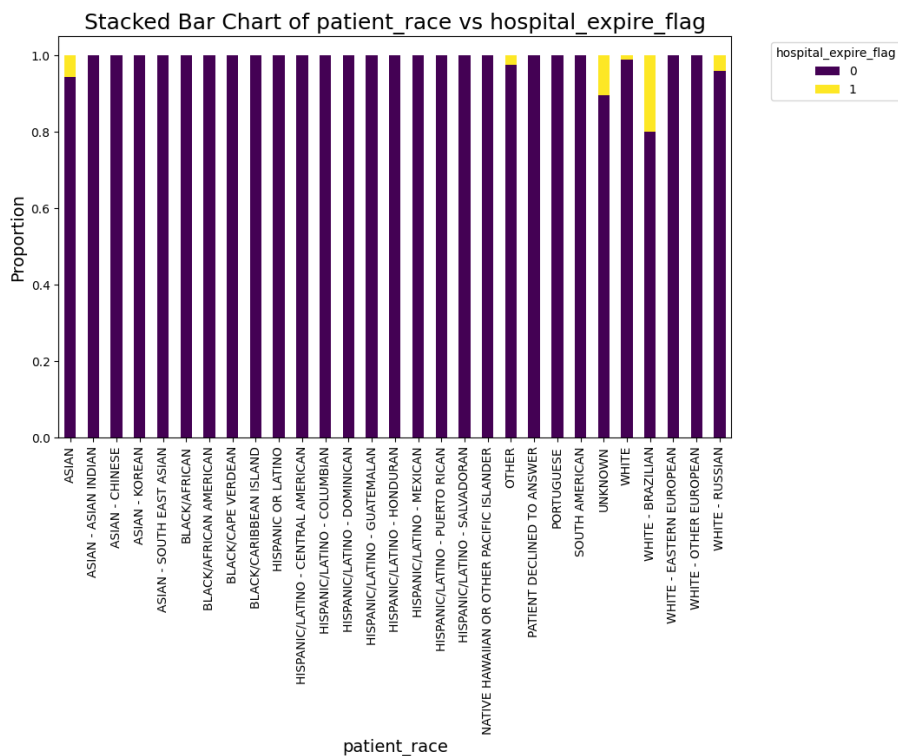
Visualizations

The stacked bar chart illustrates the relationship between patient race and hospital mortality rates, providing a visual representation of the distribution across categories.

Interpretation of Bias Severity

The analysis shows that some race categories have more pronounced biases in terms of their distribution concerning hospital mortality rates. The highest Wasserstein-2 distance observed was 0.199, suggesting a notable bias that may warrant further exploration in specific contexts.

Recommendations

Given the presence of varying levels of bias, particularly with higher Wasserstein-2 distances, it is recommended to consider these biases in any analysis or decision-making processes involving this dataset. Addressing these biases may involve further data exploration or adjustments to ensure fair and accurate outcomes.