Bias Detection Report

Language Distribution Bias Analysis Report

This report analyzes the distribution of languages in the dataset to assess potential bias. The analysis was conducted using a combination of toolset methods and literature-based methods to provide a comprehensive view of the bias levels.

Bias Type: Distribution

Relevant Feature: patient_lang

Bias Level: Significant Bias (Level 4)

- Toolset methods indicate extreme to significant bias levels, with Shannon Entropy and Gini Index showing a high concentration in specific language categories.

- Literature methods suggest a moderate level of bias, with Jensen-Shannon and Kullback-Leibler Divergence indicating some imbalance.

Detailed Findings

1. Shannon Entropy & Balance: Extreme Bias (Level 5)

2. Max/Min Ratio: Significant Bias (Level 4)

3. Gini Index: Extreme Bias (Level 5)

4. Jensen-Shannon Divergence: Moderate Bias

5. Kullback-Leibler Divergence: Moderate Bias

Conclusion

The analysis highlights a prominent language bias within the dataset. The dominance of certain languages suggests the need for careful consideration in applications relying on balanced language representation. The visualizations further illustrate the disproportionate distribution.