

Bias Detection Report

Bias Detection Report for 'native-country' Feature

Dataset Analyzed:

- **Dataset Path**: `source_files/Adult.csv`
- **Feature Examined**: "native-country"

Bias Detection Methods and Results:

1. **Bias Type**: Distribution
2. **Relevant Feature**: "native-country"
3. **Toolset Analysis**:
 - **Shannon Entropy and Balance Metric**:
 - Balance: 0.175
 - Shannon Entropy: 0.944
 - **Interpretation**: The balance metric is low, indicating a high level of imbalance in the distribution of native countries.

- **Max/Min Ratio**:

- Max/Min Ratio: 29170.0

- **Interpretation**: A very high max/min ratio suggests extreme concentration in certain categories, indicating a significant distribution bias.

- **Gini Index**:

- Corrected Gini Index: 0.199

- Adjusted Gini Index: 0.203

- **Interpretation**: The Gini Index indicates inequality in the distribution, supporting the presence of bias.

4. **Reference Literature Method (Chi-Square Test)**:

- The Chi-Square test was applied to compare observed frequencies with expected uniform distribution. However, the execution results were not available, indicating a need for further examination.

5. **Visualizations**:

- **Pie Chart**: See attached image.

- **Heatmap**: See attached image.

Bias Level Assessment:

The overall bias level in the "native-country" feature is categorized as "Extreme Bias." This suggests that the dataset may not be well-balanced in terms of geographic representation, and significant variations are present.

Recommendations:

- Consider rebalancing the dataset to ensure better geographic representation.
- Be cautious when using this feature in analysis, as the bias may significantly influence outcomes.

