Siyamthanda Dlakavu – Bias Audit Report Documentation

Getting Started

1. Access the tool: https://bias-audit-report.netlify.app/
2. Select any of the uploaded CSV dataset using the dropdown interface.
3. Configure your audit parameters.
4. Click "Run Audit" to generate your bias report.
5. Explore the visualizations and metrics in the report dashboard.

User Interface Overview

The application features three main sections:

1. Upload Section:

- File showing the pre-uploaded CSV datasets area.
- Dataset preview table.
- Column type indicators.

2. Configuration Panel:

- Protected attribute selection.
- Target variable selection.
- Privileged group specification.
- Audit execution button.

3. Report Dashboard:

- Metric cards (Disparate Impact, Statistical Parity, etc.).
- Interactive bar charts.
- Data tables with detailed metrics.
- Color-coded fairness indicators.

Core Features

Audit Configuration

1. *Protected Attribute*: Select the sensitive attribute to analyse (e.g., gender, race).
2. *Target Variable*: Choose the outcome variable to measure bias against.
3. *Privileged Group*: Specify which group is considered advantaged.
4. *Run Audit*: Initiates the bias analysis process.

Siyamthanda Dlakavu – Bias Audit Report Documentation

Report Visualization

The report dashboard displays:

Key Metrics

- *Disparate Impact Ratio*: Measures favourable outcome rates between groups.
- *Statistical Parity Difference*: Difference in selection rates.
- *Equal Opportunity Difference*: Difference in true positive rates.
- *Average Odds Difference*: Average of FPR and TPR differences.
- *Theil Index*: Economic inequality measure adapted for fairness.

Visualizations

- Comparative bar charts for privileged vs. unprivileged groups.
- Color-coded fairness indicators (green = fair, yellow = warning, red = biased).
- Interactive charts with hover tooltips showing exact values.

Data Tables

- Detailed metric values for each subgroup.
- Raw numbers supporting visualizations.
- Download option for report data.

Understanding Metrics

Disparate Impact Ratio

- Formula: (Rate_unprivileged / Rate_privileged)
- Fair range: 0.8 - 1.25
- <0.8 indicates bias against unprivileged group.
- 1.25 indicates bias against privileged group.

Statistical Parity Difference

- Formula: Rate_unprivileged - Rate_privileged.
- Ideal value: 0
- Negative values indicate bias against unprivileged group.

Equal Opportunity Difference

- Measures equality in true positive rates.
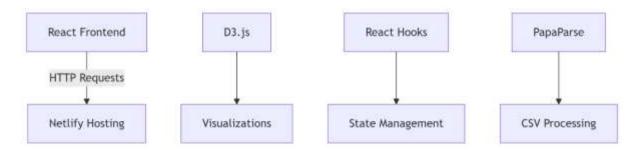- Formula: TPR_unprivileged - TPR_privileged

- Fair range: -0.1 to +0.1

Theil Index

- Measures inequality across all groups.

- Lower values indicate more equality.

- 0 represents perfect equality.

Technical Implementation

Architecture

React Frontend	D3.js	React Hooks	PapaParse
HTTP Requests ↓	↓	↓	↓
Netlify Hosting	Visualizations	State Management	CSV Processing

Tech Stack

- *Framework*: React (v18).

- *Visualization*: D3.js, Chart.js.

- *UI Components*: Material-UI.

- *CSV Parsing*: PapaParse.

- *Hosting*: Netlify.

- *CI/CD*: GitHub Actions.