

Bias Audit Report

Documentation

Getting Started

1. **Access the tool:** <https://bias-audit-report.netlify.app/>
2. **Select Dataset:** Use the dropdown interface to choose from pre-uploaded CSV datasets.
3. **Configure Audit:** Define your analysis parameters (Protected attributes, target variables).
4. **Execute:** Click "**Run Audit**" to generate the fairness report.
5. **Review:** Explore the interactive visualizations and automated metrics.

User Interface Overview

The application is logically structured into three functional stages:

1. Upload & Preview Section

- **Dataset Selection:** Dropdown menu containing curated CSV datasets.
- **Data Preview:** Live table showing the first few rows of the data.
- **Type Indicators:** Visual labels indicating column data types (Categorical, Numerical).

2. Configuration Panel

- **Protected Attribute:** Selection of the sensitive feature (e.g., Gender, Race).
- **Target Variable:** The outcome variable to be audited for bias.
- **Privileged Group:** Specification of the group historically advantaged in the dataset.

3. Report Dashboard

- **Metric Summary Cards:** High-level view of primary fairness ratios.
- **Comparative Charts:** Visual distribution showing privileged vs. unprivileged outcomes.
- **Drill-down Tables:** Exhaustive data points supporting the visualizations.
- **Status Indicators:** Color-coded alerts for easy interpretation (Fair, Warning, Biased).

Core Features

Audit Configuration Parameters

Refining the audit ensures accuracy in identifying disparate impact:

- **Attribute Selection:** Choose the sensitive feature for analysis.
- **Outcome definition:** Define the positive outcome (target variable).
- **Group Baseline:** Establishing the privileged reference point.

Report Visualization

Key Metrics

- **Disparate Impact Ratio:** Measures favourable outcome rates between groups.
- **Statistical Parity Difference:** Difference in overall selection rates.
- **Equal Opportunity Difference:** Measures equality in true positive rates.
- **Average Odds Difference:** Balance between FPR and TPR differences.
- **Theil Index:** Quantitative measure of economic/outcome inequality.

Understanding the Metrics

Disparate Impact Ratio

$(\text{Rate}_{\text{unprivileged}} / \text{Rate}_{\text{privileged}})$

- **Fair Range:** 0.80 — 1.25 (The "Four-Fifths Rule").
- Values below 0.8 indicate significant bias against the unprivileged group.
- Values above 1.25 indicate bias against the privileged group.

Statistical Parity Difference

$\text{Rate}_{\text{unprivileged}} - \text{Rate}_{\text{privileged}}$

- **Ideal Value:** 0.00
- Negative values reveal a selection rate disadvantage for the unprivileged group.

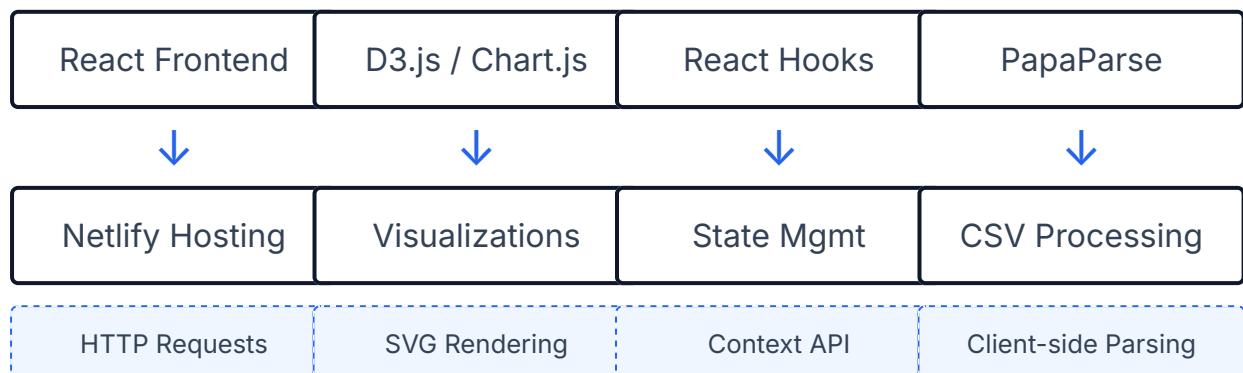
Equal Opportunity Difference

TPR_unprivileged - TPR_privileged

- **Fair Range:** -0.10 to +0.10
- Ensures that qualified candidates from both groups have the same chance of a positive outcome.

Technical Implementation

System Architecture



Tech Stack

React (v18) D3.js Chart.js Material-UI (MUI) PapaParse GitHub Actions
Netlify

Key Implementation Details

- **Responsiveness:** Uses Material-UI Grid system for desktop and mobile compliance.
- **Performance:** Client-side parsing via PapaParse ensures no data leaves the browser, maintaining privacy.
- **CI/CD:** Automated deployments through GitHub Actions workflows.