

- [Bias Detection Dashboard](#)
  - [Features](#)
  - [Project Structure](#)
  - [Installation](#)
  - [Usage](#)
  - [Sample Dataset](#)
  - [Dependencies](#)
  - [Contributing](#)
  - [License](#)

# Bias Detection Dashboard

---



A comprehensive tool to analyze datasets for bias, privacy issues, and machine learning readiness. This project provides an interactive, visually appealing interface to perform data analysis, visualize insights, and generate detailed PDF reports.

## Features

---

- **Multi-Page Interface** : Separate pages for each analysis step (Upload, Visualizations, Bias Analysis, etc.).
- **Interactive Visualizations** : Includes histograms, box plots, violin plots, pair plots, and fairness metric plots.
- **Bias Analysis** : Calculates fairness metrics (Disparate Impact, Equalized Odds) and overall bias percentage.
- **Privacy Check** : Detects Personally Identifiable Information (PII) with actionable recommendations.
- **ML Readiness** : Assesses dataset suitability for machine learning with a detailed gauge and sample predictions.
- **PDF Report** : Generates a comprehensive report with visualizations, fairness metrics, PII analysis, statistical summaries, and ML readiness conclusion.
- **Theming** : Supports light and dark themes for better user experience.
- **Bias Mitigation** : Offers advanced bias mitigation techniques with downloadable mitigated datasets.

# Project Structure

---

```
bias_detection_dashboard/
├── pages/
│   ├── 1_upload.py           # Page for dataset upload and cleaning
│   ├── 2_visualizations.py   # Page for data visualizations
│   ├── 3_bias_analysis.py    # Page for bias and fairness analysis
│   ├── 4_privacy_check.py    # Page for PII detection
│   ├── 5_statistical_analysis.py # Page for statistical analysis
│   ├── 6_recommendations.py  # Page for recommendations and bias mitigation
│   ├── 7_ml_readiness.py     # Page for ML readiness and prediction
│   └── 8_generate_report.py  # Page for generating PDF report
├── app.py                   # Main app with navigation
├── data_processor.py        # Data loading and cleaning logic
├── bias_analyzer.py         # Bias detection and mitigation logic
├── privacy_checker.py       # PII detection logic
├── visualizer.py            # Visualization logic
├── pdf_generator.py         # PDF report generation logic
├── ml_predictor.py          # ML readiness and prediction logic
├── style.css                # Custom CSS for styling
├── requirements.txt          # Python dependencies
├── generate_hiring_data.py  # Script to generate sample dataset
└── assets/
    ├── flower.gif           # Animation for success
    ├── error.gif            # Animation for errors
    └── logo.png              # Logo for dashboard and report
```

## Installation

---

### 1. Clone the Repository :

```
git clone <repository-url>
cd bias_detection_dashboard
```

### 1. Set Up a Virtual Environment (optional but recommended):

```
python -m venv venv
source venv/bin/activate # On Windows
```

### 2. Install Dependencies :

```
pip install -r requirements.txt
```

### WeasyPrint Dependencies (required for PDF generation):

- **Ubuntu/Debian** : `bash sudo apt-get install libpango-1.0-0 libpangoft2-1.0-0`
- **macOS** (using Homebrew):

```
brew install pango
```

- **Windows** :
  - Install GTK from [GTK for Windows](#).
  - Follow the installation instructions provided there.

#### 1. Generate Sample Dataset :

```
python generate_hiring_data.py
```

This generates `hiring_data.csv`, a sample dataset for testing.

#### 1. Run the Dashboard :

```
streamlit run app.py
```

Open your browser and navigate to `http://localhost:8501`.

## Usage

---

#### 1. Upload Dataset :

- Navigate to the "Upload" page.
- Upload a CSV or Excel file (e.g., `hiring_data.csv`).
- View the dataset overview and download the cleaned dataset if needed.

#### 1. Explore Visualizations :

- Go to the "Visualizations" page.
- Explore distributions, relationships, box plots, violin plots, and data flow diagrams.

#### 1. **Perform Bias Analysis :**

- Visit the "Bias Analysis" page.
- Select a binary target column (e.g., `shortlisted`) to calculate fairness metrics.
- View the overall bias percentage and comparative fairness plots.

#### 1. **Check Privacy :**

- Navigate to the "Privacy Check" page.
- Review detected PII columns and read recommendations.

#### 1. **Statistical Analysis :**

- Go to the "Statistical Analysis" page.
- View correlation heatmaps, statistical summaries, skewness, and kurtosis.

#### 1. **Get Recommendations :**

- Visit the "Recommendations" page.
- Review actionable recommendations and apply bias mitigation if needed.

#### 1. **Assess ML Readiness :**

- Navigate to the "ML Readiness" page.
- Check the ML readiness score and run sample predictions.

#### 1. **Generate PDF Report :**

- Go to the "Generate Report" page.
- Click "Generate Report" to download a PDF with all analysis results, including fairness metrics, PII analysis, visualizations, and ML readiness conclusion.

## Sample Dataset

---

The `generate_hiring_data.py` script generates a sample dataset (`hiring_data.csv`) with the following columns:

- `name`, `email`, `phone`: PII columns
- `Age`, `Gender`: Sensitive columns

- `DailyRate`, `MonthlyIncome`: Numerical features
- `shortlisted`: Binary target column (0 or 1)

## Dependencies

---

Listed in `requirements.txt`:

- `streamlit==1.25.0`
- `pandas==2.0.3`
- `plotly==5.15.0`
- `fairlearn==0.9.0`
- `scikit-learn==1.3.0`
- `numpy==1.25.0`
- `weasyprint==59.0`
- `seaborn==0.12.2`
- `matplotlib==3.7.2`
- `networkx==3.1`
- `scipy==1.10.1`

## Contributing

---

Feel free to submit issues or pull requests to enhance the dashboard. Contributions are welcome!

## License

---

This project is licensed under the MIT License.

---

Built with  by the Satyajit Nayak...