

How Fair Is Your ML Model? A Visualization Approach

PER2024-030

Students: Adam DIENES (EIT M2) and Bence Zsolt BEREGLI (EIT M2)
Supervisors: Mireille Blay-Fornarino (I3S) and Anne-Marie Pinna-Dery (I3S)

Type: Research
Year: 2025

As AI increasingly influences critical decisions, ensuring model fairness is paramount. Our research focuses on systematically **visualizing and justifying fairness throughout the ML lifecycle**. We aim to bridge the gap between complex algorithms and real-world impact, empowering AI developers and integrators with interactive fairness assessments — without relying on sensitive attributes.



Problem

1

- Lack of systematic support for evaluating dataset fairness (e.g., Gender pay gap).
- **Difficulty in visualizing and justifying** fairness across the ML lifecycle.
- Growing need for tools to assess and enhance model **fairness and transparency**.

Methodology

2

- Development of **interactive visualizations** for pre-processing and post-processing stages.
- Creation of tailored visualizations for model integrators and AI developers.
- Evaluating the **impact of sensitive attributes** on model outcomes using the What-If Tool.

Solution

3

Our tool empowers users to detect, visualize, and mitigate fairness issues in classification data and models through an intuitive, user-friendly interface.

Key Features

- **Seamless Data Upload & Setup:** Easily upload CSV datasets and specify sensitive attributes, prediction columns and other relevant features.
- **Automated Metric Calculation:** Calculates key fairness metrics: Statistical Parity, Equality of Odds, Equal Opportunity, Disparate Impact and MinDiff.
- **Interactive Visualizations:** Gain insights through clear, interactive visual representations tailored to each fairness metric.
- **Actionable Recommendations:** Provides contextual guidance on techniques to reduce bias in pre-processing and post-processing stages.
- **Results Documentation:** Export analysis results as a comprehensive fairness report for auditing and compliance purposes.

PER2024-030: FairnessLens

☐ Show source code

Please select all attributes from the uploaded dataset according to the characteristics. Click the green Run button below to continue the execution.

Select All Numerical Attributes*

duration x amount x installment_rate x
age x number_credits x people_liable x

Select All Categorical Attributes*

status x credit_history x purpose x
savings x employment_duration x
personal_status_sex x other_debtors x
present_residence x property x
other_installment_plans x housing x
job x telephone x foreign_worker x

Select The Prediction Attribute*

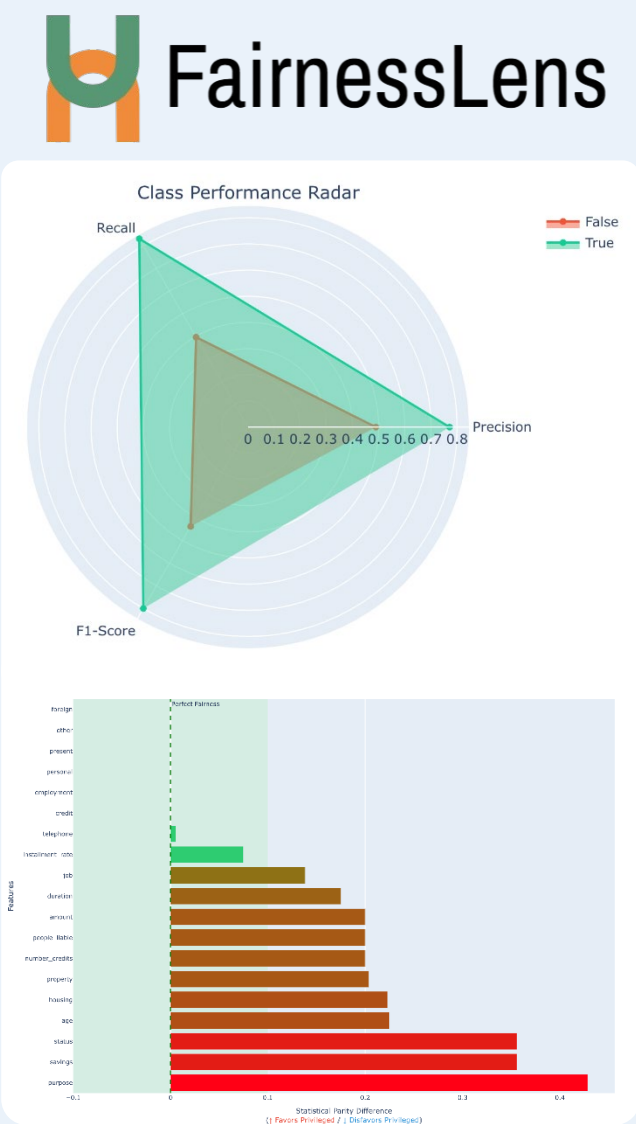
credit_risk

Select Protected Attributes

personal_status_sex x age x

Not necessary to provide any protected attribute. Later, you can choose to rerun the tool.

Run



Upload dataset as CSV

Set types & sensitive attributes

Process and evaluate the data

Interactive visualizations

Change parameters

Action based on recommendations

Result export

Perspectives

- Potential to revolutionize fairness assessment
- Implications for improving trust and accountability
- Future work: Extending the **tool use cases** (e.g., for regression model) and exploring **new fairness metrics**

Contact

- Adam DIENES / adam.dbence-zsoltienes@etu.univ-cotedazur.fr
- Bence Zsolt BEREGLI / bence-zsolt.beregi@etu.univ-cotedazur.fr
- Mireille Blay-Fornarino / mireille.blay@univ-cotedazur.fr
- Anne-Marie Pinna-Dery / anne-marie.pinna@univ-cotedazur.fr
- Nicolas Lacroix / nicolas.lacroix@univ-cotedazur.fr