CERGY PARIS UNIVERSITÉ

ENSEA
Ecole Nationale Supérieure de l'Electronique et de ses Applications

ProPublica recidivism Use Case (Compas risk assessment tool)

Transparency & Fairness in AI and Big Data Algorithms

LYOUSFI Youssef

Dimitris KOTZINOS
Vassilis CHRISTOPHIDES

M2 Data Science & Machine Learning

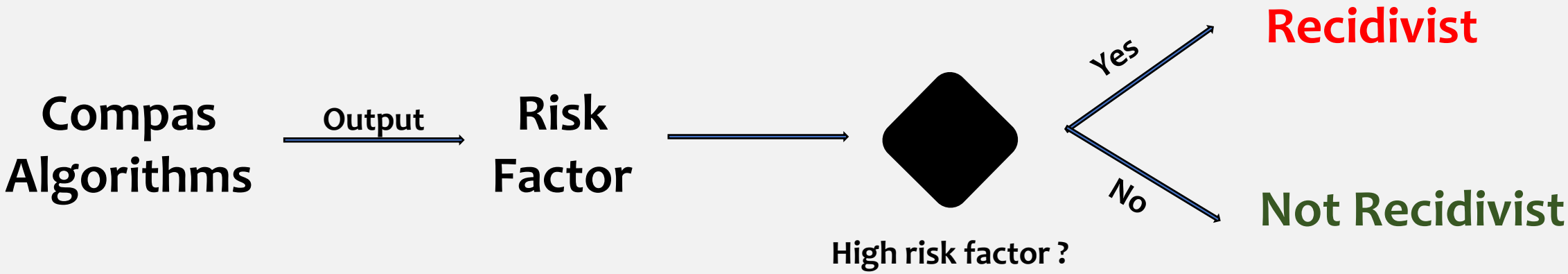
03/03/2023

Overview

- Compas algorithm
- Bias Detection & Mitigation
  - Reweighing preprocessing
  - Optimized preprocessing
  - Adversarial Inprocessing
  - Calibrated Odds PostProcessing

Compas

Correctional Offender Management Profiling for Alternative Sanctions



The higher the risk factor, the more you become a recidivist

(predicted to reoffend, but don't)

False Positive

Black: 44.9%

White: 23.5%

(predicted not to reoffend, but did)

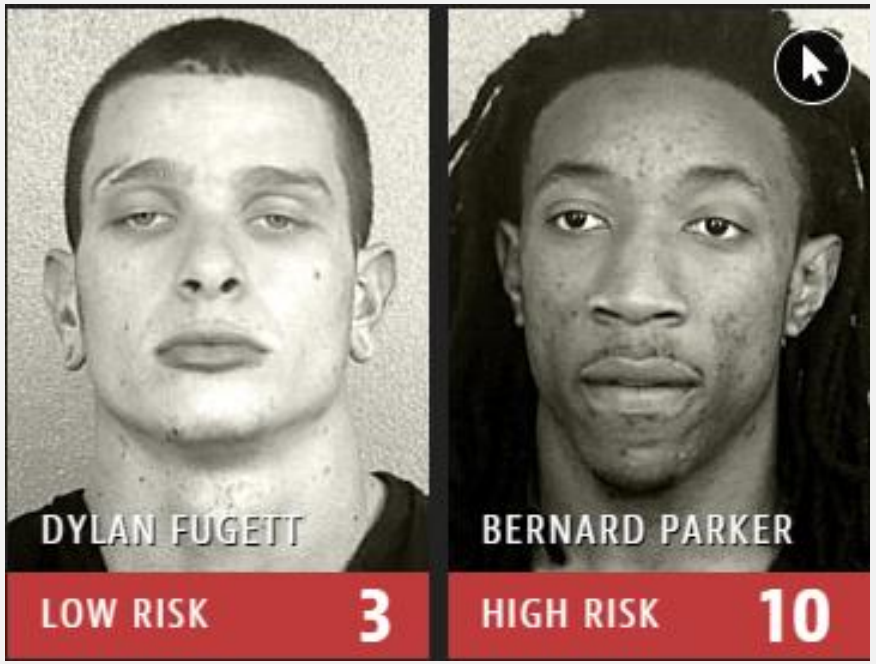
False Negative

Black: 28.1%

White: 47.7%

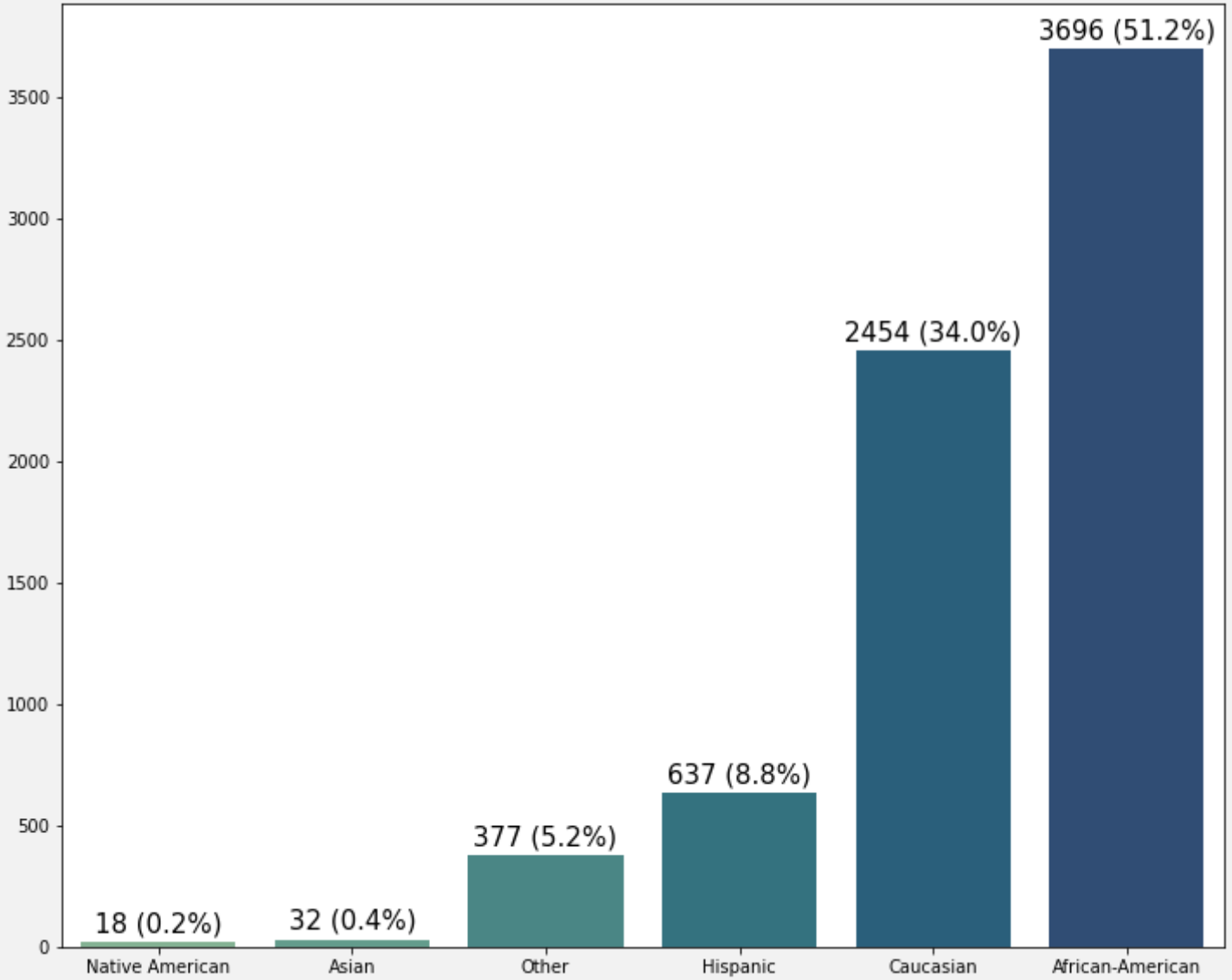
If you are black, you are more likely to be assessed as high risk than if you were white

The accuracy, fairness, and limits of predicting recidivism | Science Advances

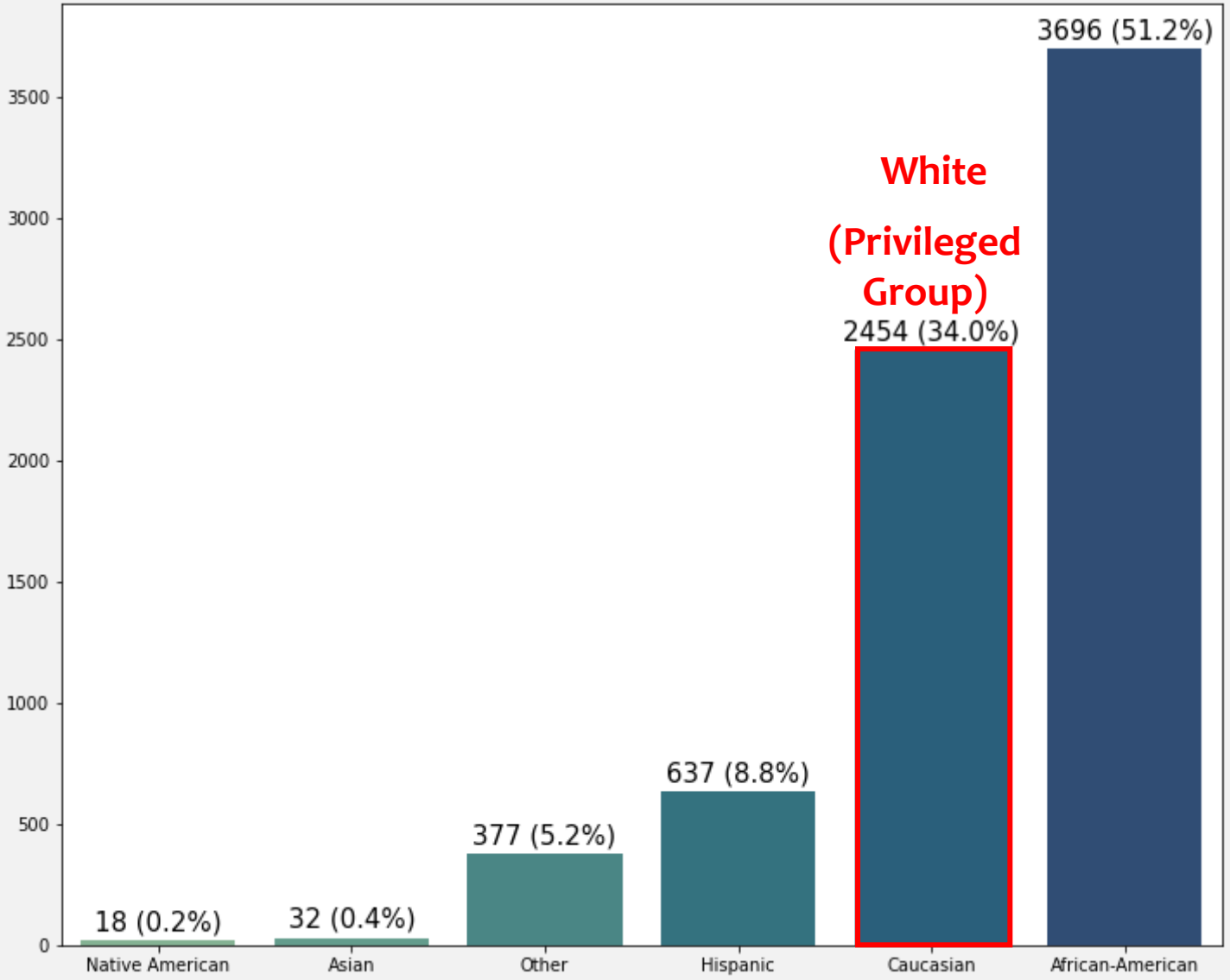


Fugett was rated low risk after being arrested with cocaine and marijuana. He was arrested three times on drug charges after that

Machine Bias — ProPublica

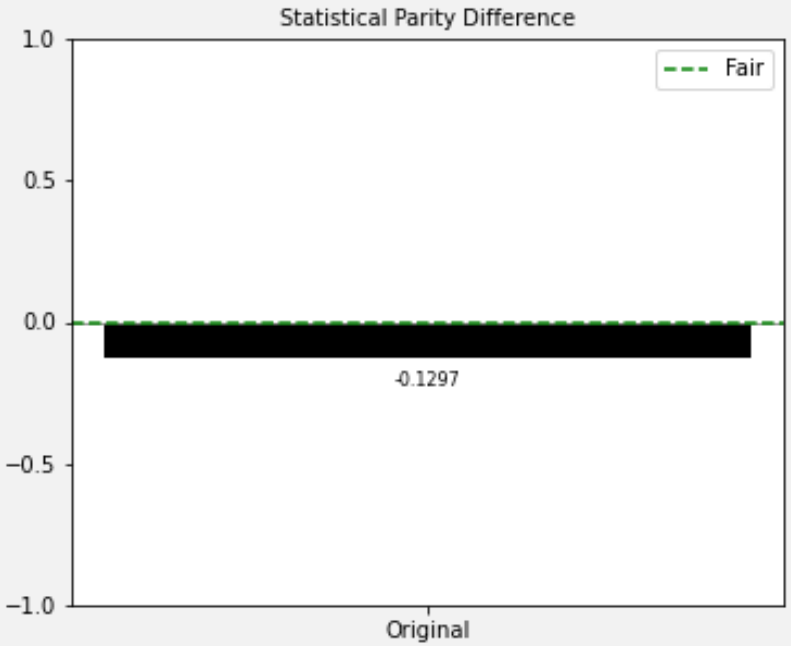


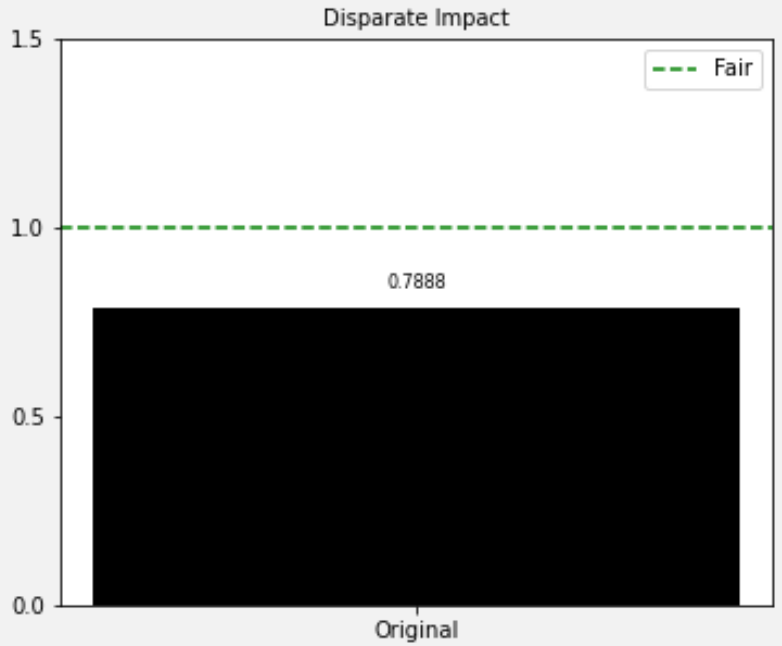
Race attribute distribution



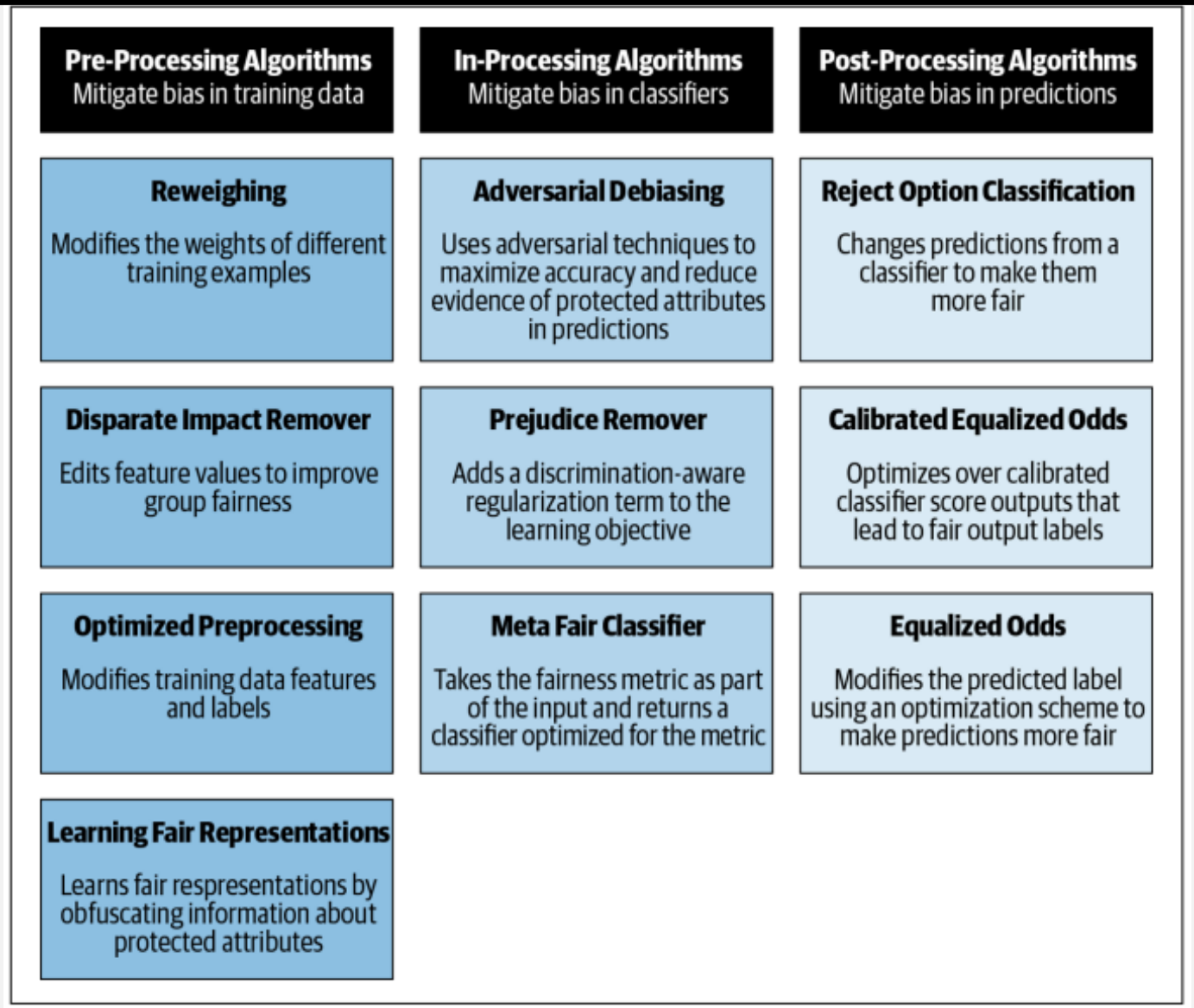
Race attribute distribution

Bias Detection & Mitigation

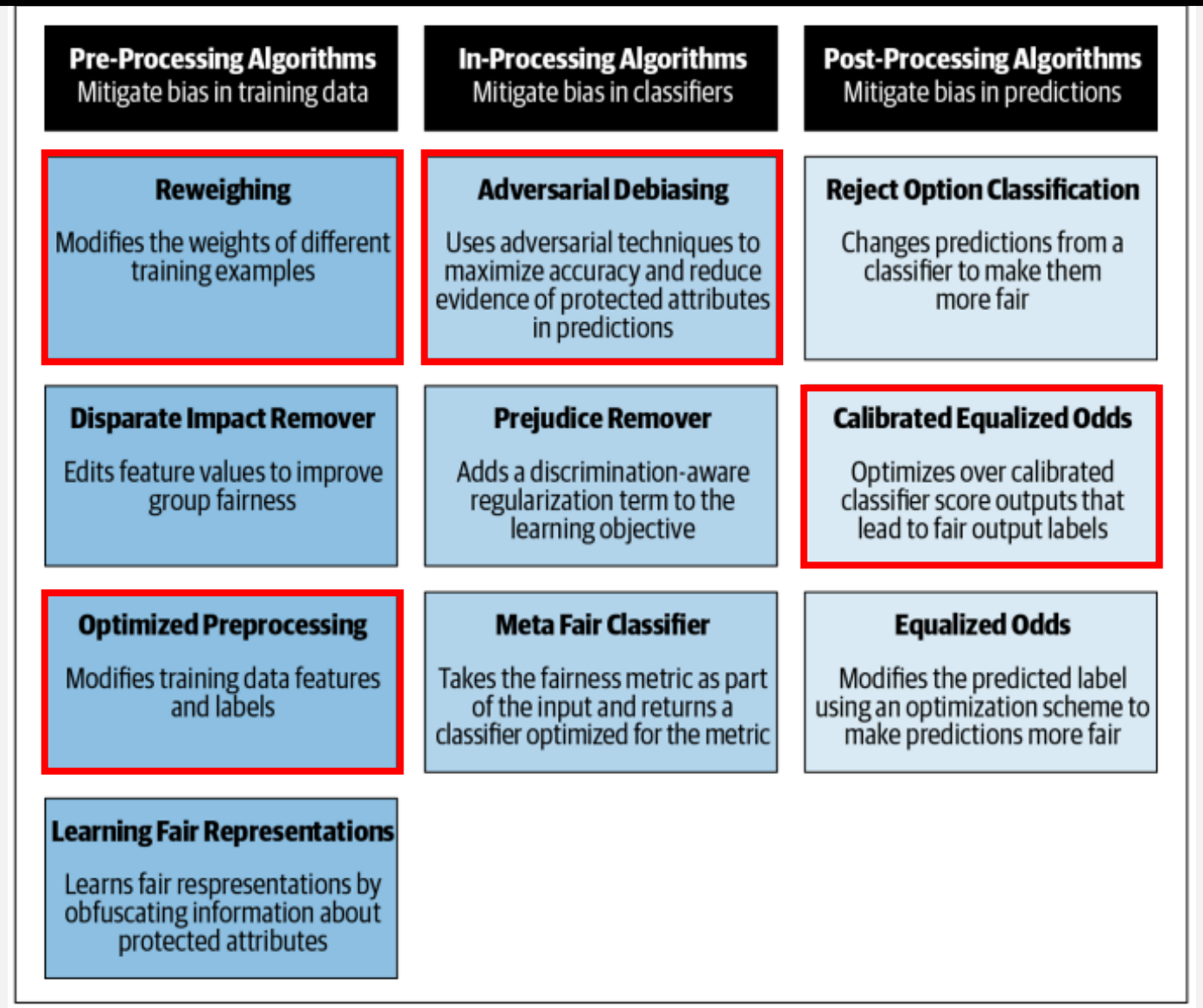




Fairness Metrics on Original Dataset

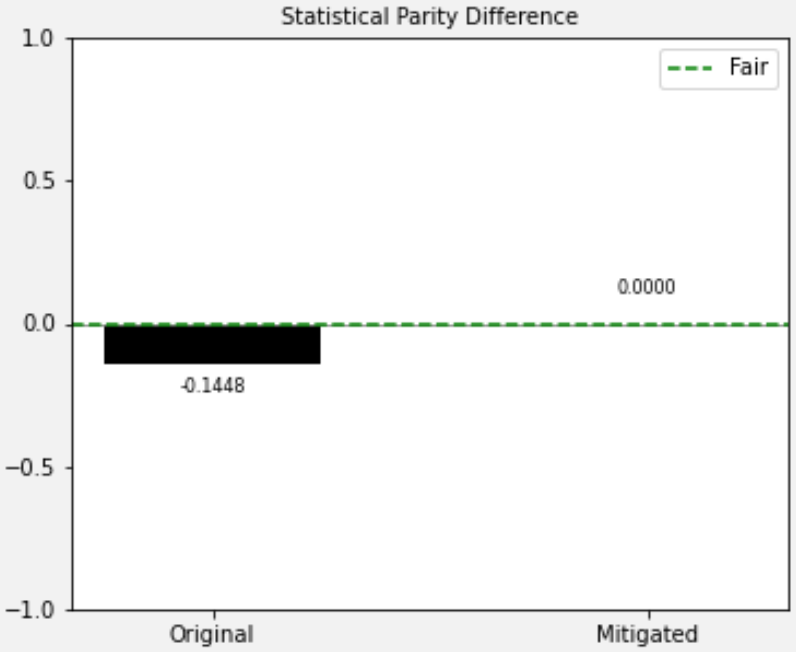


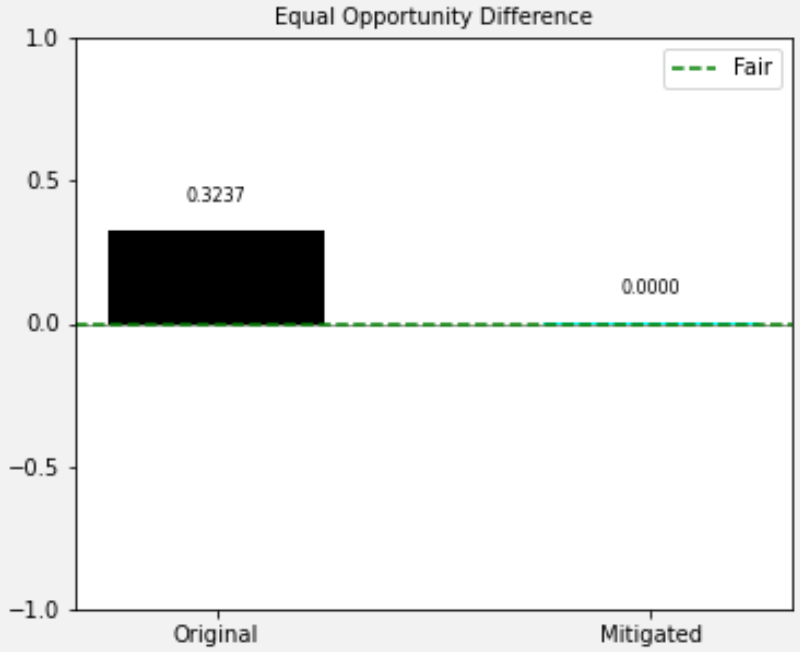
Bias Mitigation Algorithms

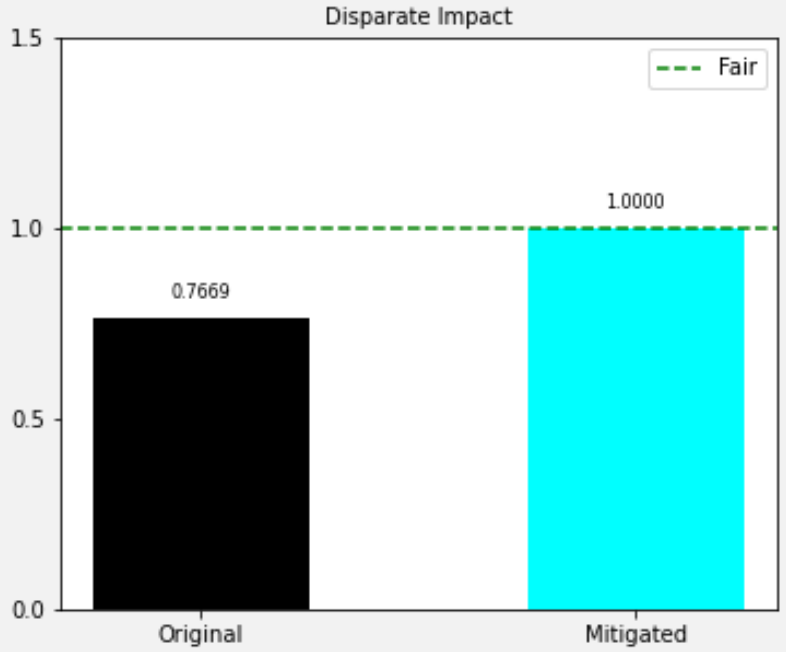


The Algorithms used

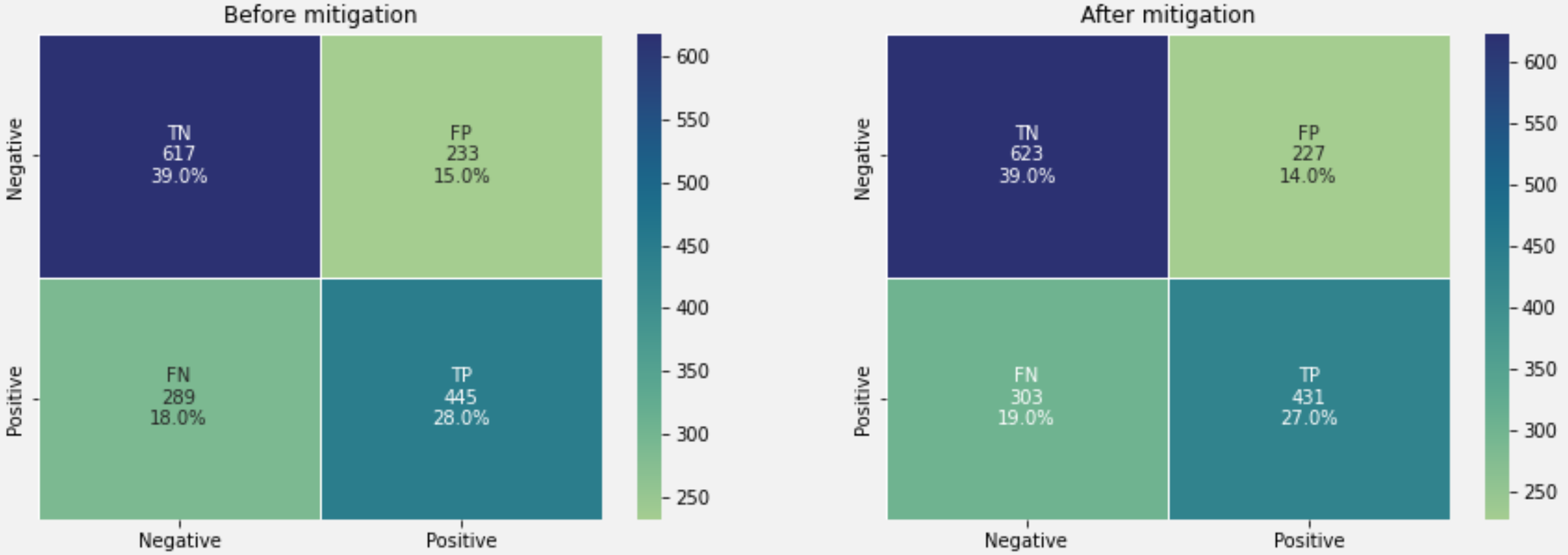
Reweighing Preprocessing





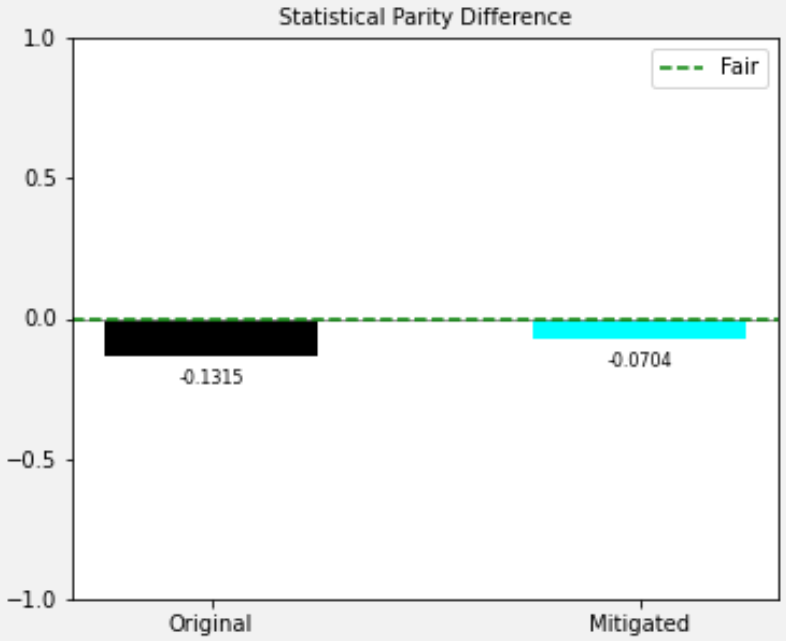


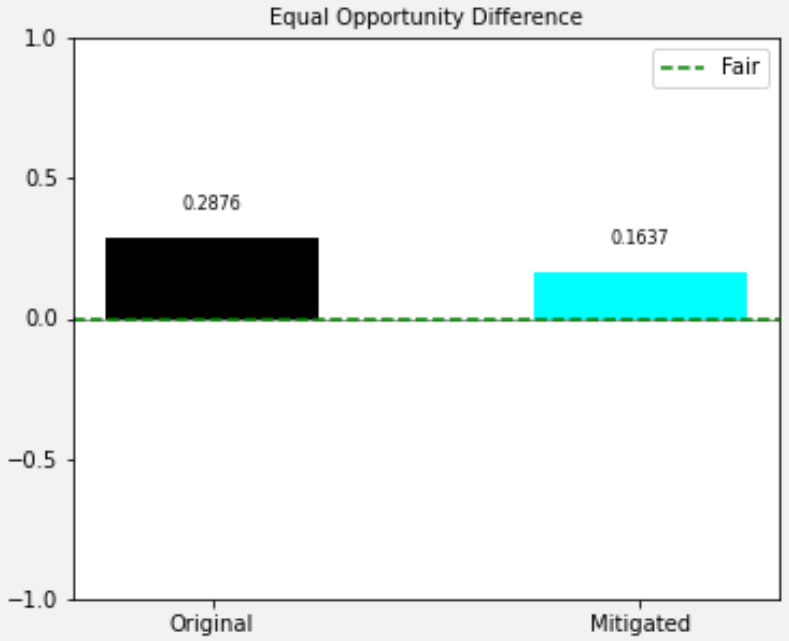
Fairness Metrics after Reweighing technique

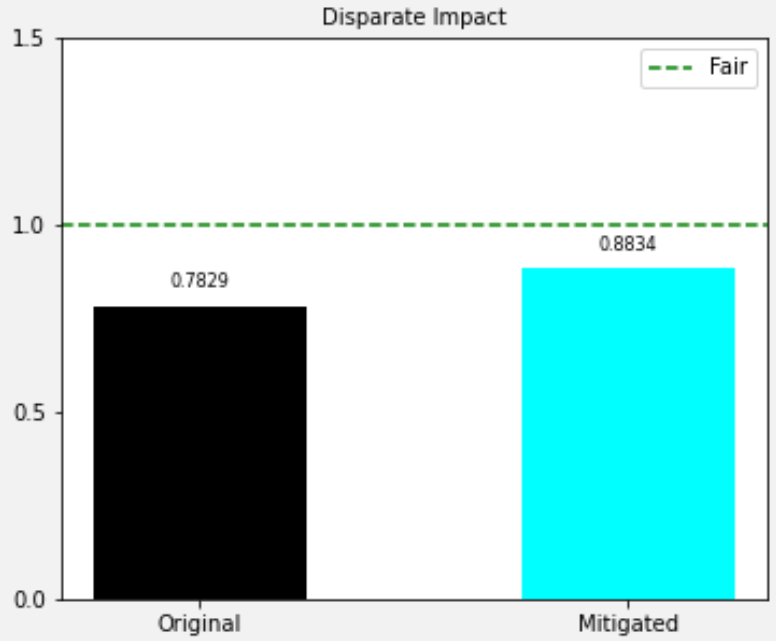


Confusion Matrix for Random Forest

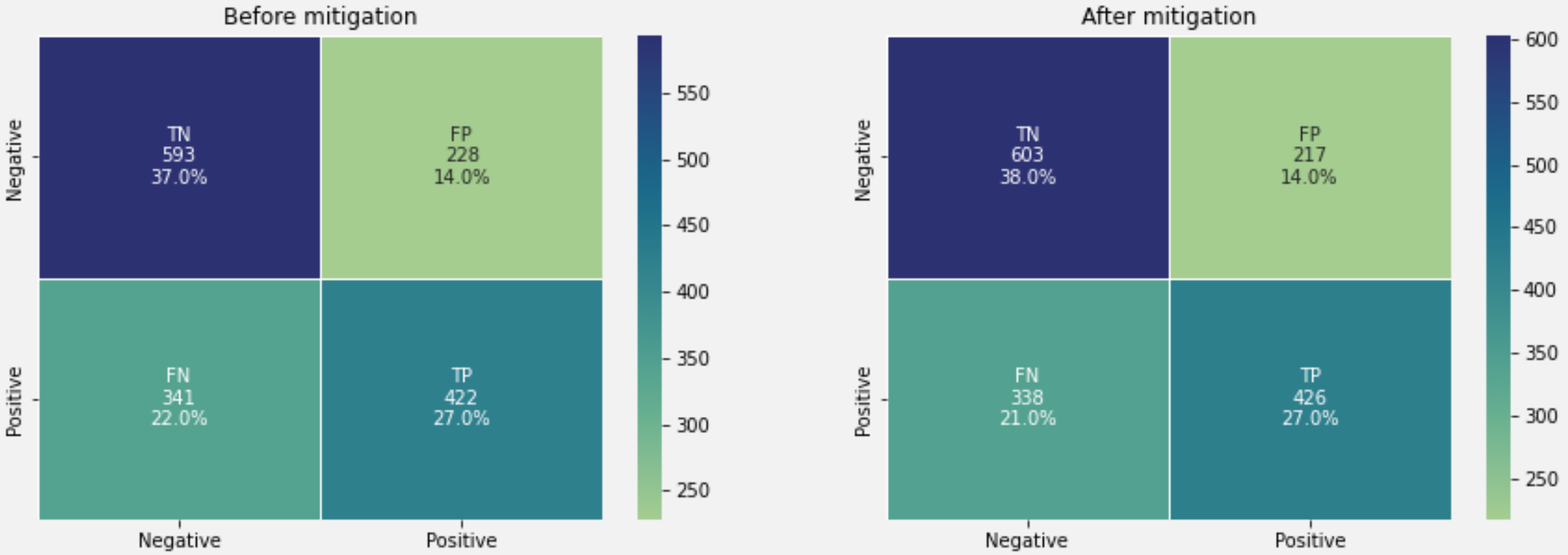
Optimized Preprocessing





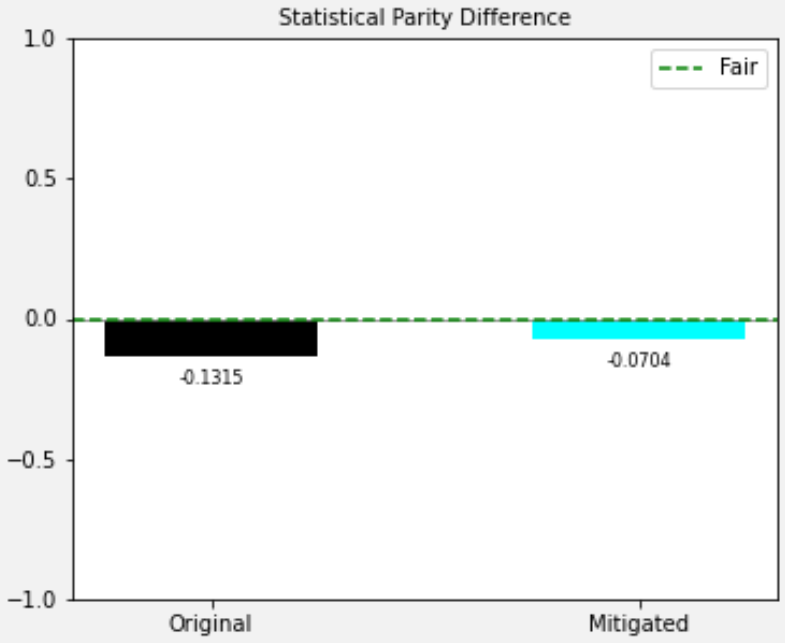


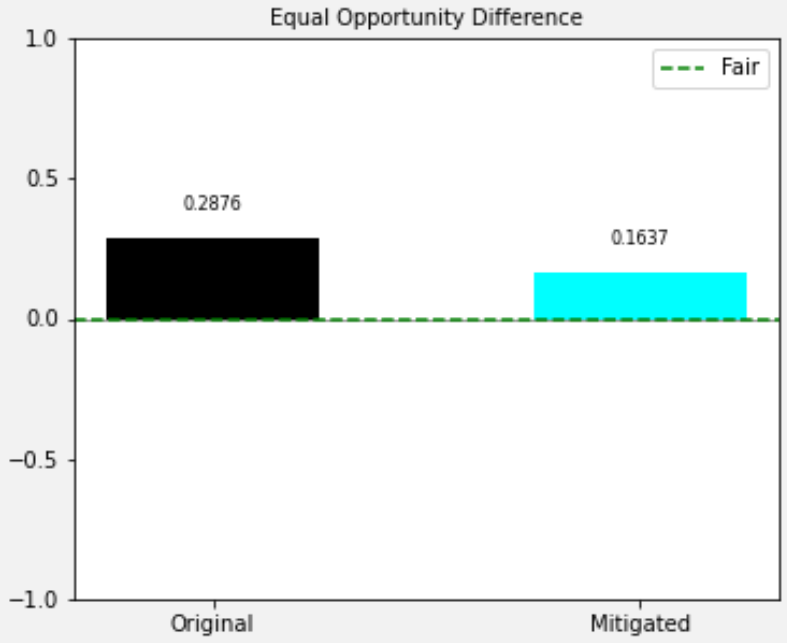
Fairness Metrics after Optimized technique

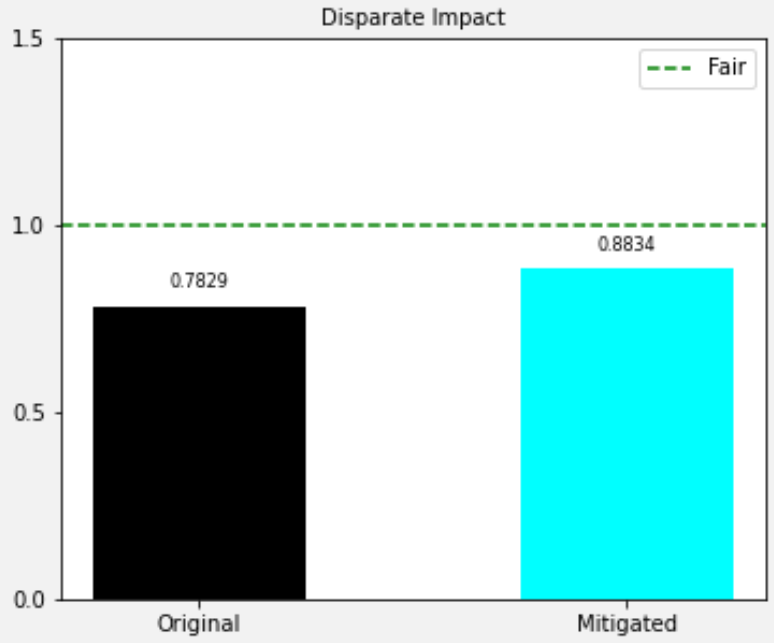


Confusion Matrix for Random Forest

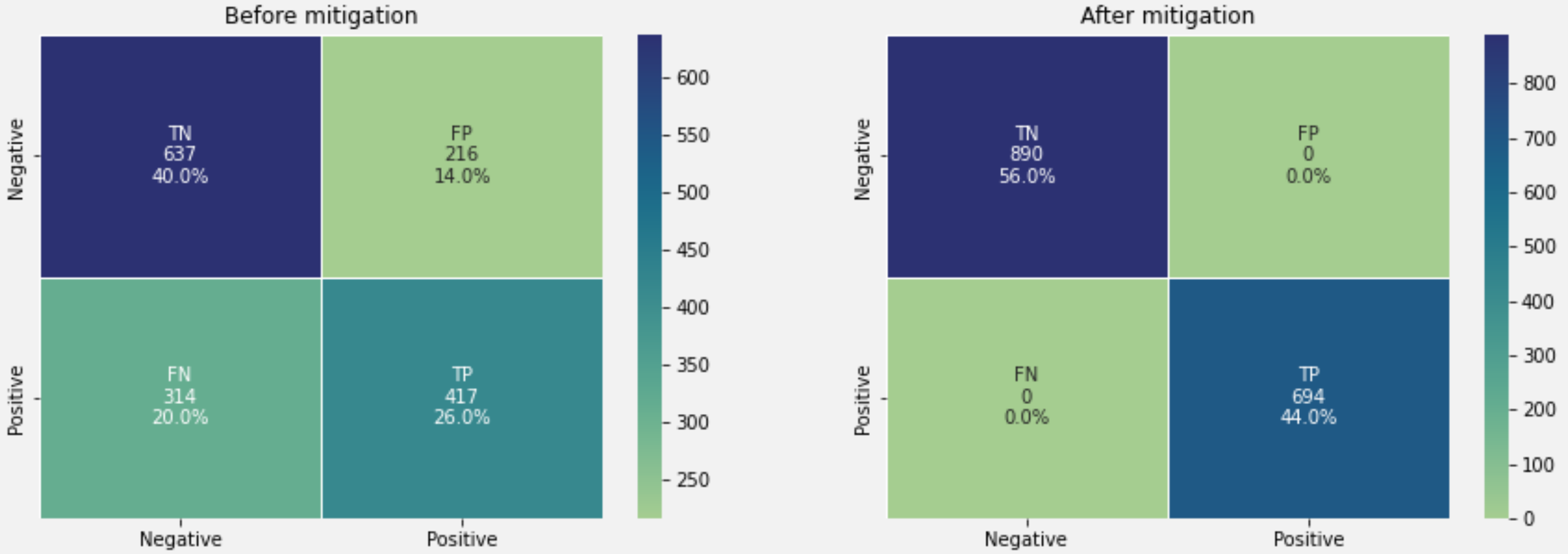
Adversarial Inprocessing





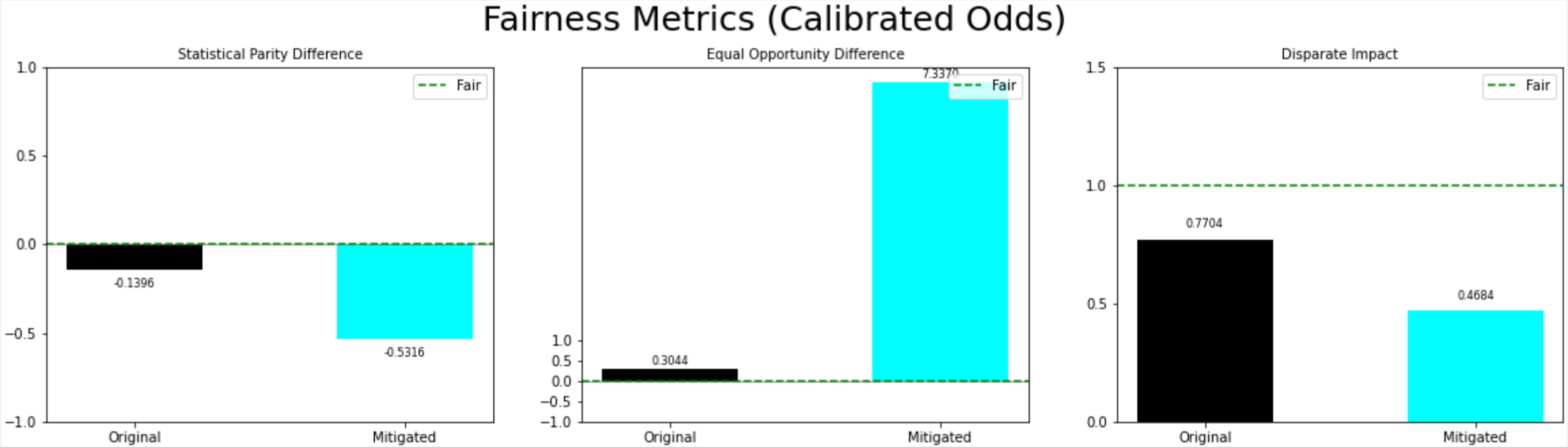


Fairness Metrics after Adversarial technique

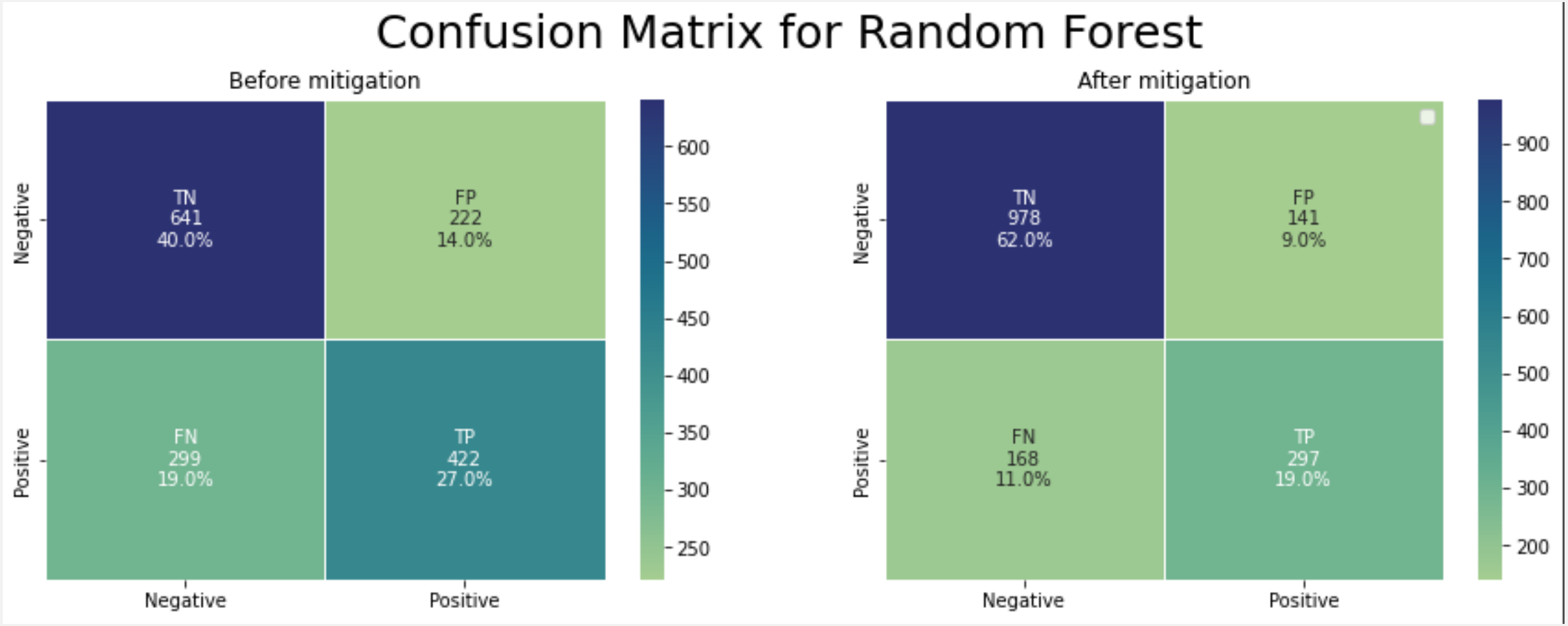


Confusion Matrix for Random Forest

Calibrated Odds post-processing



Fairness Metrics after Calibrated Odds technique



Confusion Matrix for Random Forest