Hw 2

XAI 2023/24 mimuw, 19.10.2023, Witold Drzewakowski

Task 1

I will start by calculating true positives, false positives, false negatives f
groups: red and blue.

$TP_b = 0.6, FP_b = 0.05, FN_b = 0.2$

$TP_r = 0.25, FP_r = 0.25, FN_r = 0.25$

Now let's calculate PPV and TPR.

$PPV_b = \frac{TP_b}{TP_b + FP_b} = \frac{0.6}{0.65} = 12/13$

$PPV_r = \frac{TP_r}{TP_r + FP_r} = \frac{0.25}{0.5} = 0.5$

$TPR_b = \frac{TP_b}{TP_b + FN_b} = \frac{0.6}{0.8} = 0.75$

$TPR_r = \frac{TP_r}{TP_r + FN_r} = \frac{0.25}{0.5} = 0.5$ (Alternatively

Now let's calculate probabilities of selecting a candidate from both grou

$R_r = P(selected|red) = 0.5$

$R_b = P(selected|blue) = 0.65$

The predictive rate parity coefficient:

$\frac{PPV_b}{PPV_r} = \frac{24}{13}$ — blue is more priviliged

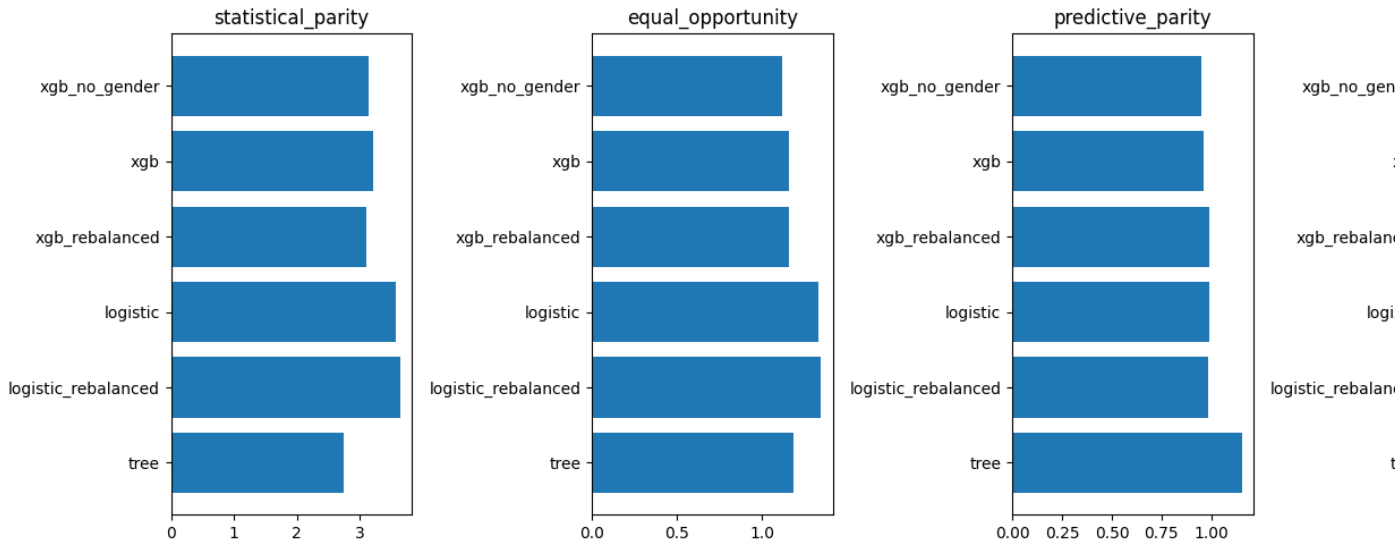
The equal opportunity coefficient:

$\frac{TPR_b}{TPR_r} = \frac{3}{2}$ — blue is more priviliged

Demographic parity coefficient:

$\frac{R_b}{R} = \frac{13}{10}$ — blue is more priviliged

I have trained logistic regression, decision tree and gradient boosted tr
income dataset. The protected attribute that I am working with is gende
tested two strategies for mitigating bias: (i) I removed the protected col
have tried rebalancing data. Results are presented in the following bar



We can see that there seems to be negative correlation between accur
predictive parity. We can see that decision tree, which has worst accura
best statistical parity. XGB with no gender column performs slightly bet
slightly better fairness metrics, suggesting a slight overfitting to gender

We can see that mitigating bias, by resampling the dataset to include e
of men and women does not help at all.

	logistic	logistic_rebalanced	tree	xgb	xgb_no_gender
statistical_parity	3.571922	3.645472	2.741528	3.218254	3.129965
equal_opportunity	1.328129	1.346642	1.186238	1.158213	1.118033