Hw 2

XAI 2023/24 mimuw, 19.10.2023, Witold Drzewakowski

Task 1

I will start by calculating true positives, false positives, false negatives for both 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.

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

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

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

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

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

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

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

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

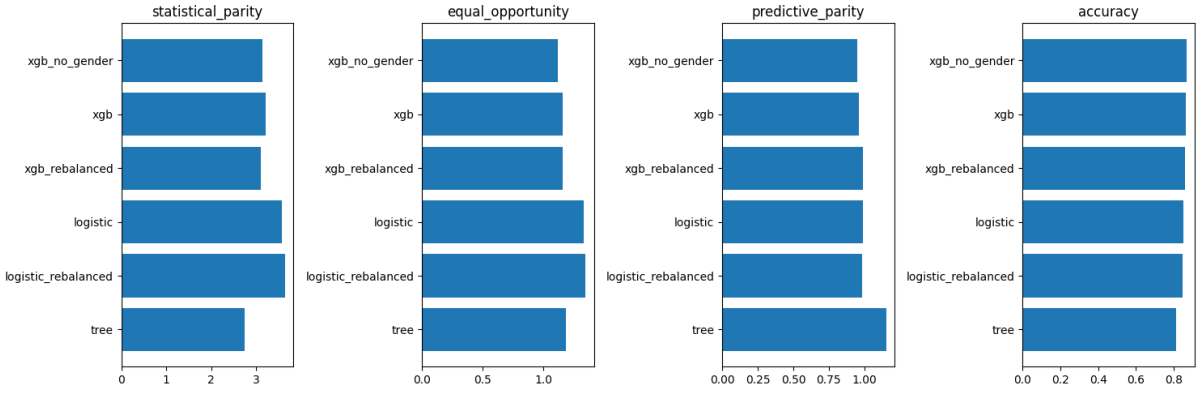
Demographic parity coefficient:

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

If in the red group, 0.65 of people were randomly admitted, clearly Demographic parity coefficient would equal 1, TPR_r would increase leading to equal opportunity coefficient being closer to 1, and PPV_r would stay the same, so predictive rate parity coefficient would not change. So two out of three metrics would strictly improve.

Task 2

I have trained logistic regression, decision tree and gradient boosted tries on adult income dataset. The protected attribute that I am working with is gender. I have also tested two strategies for mitigating bias: (i) I removed the protected column and (ii) I have tried rebalancing data. Results are presented in the following bar plot:



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

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

	logistic	logistic_rebalanced	tree	xgb	xgb_no_gender	xgb_rebalanced
statistical_parity	3.571922	3.645472	2.741528	3.218254	3.129965	3.110655
equal_opportunity	1.328129	1.346642	1.186238	1.158213	1.118033	1.156220
predictive_parity	0.990896	0.984437	1.153105	0.959087	0.951929	0.990554
accuracy	0.850906	0.848519	0.815106	0.864804	0.869297	0.862839