DBS
Live more, Bank less

CBG Analytics

Model Performance, Fairness and Explainability Report

Model ID:
Model Name:
Country:
Model Developer (Project Lead):
Date:

Live more, Bank less

I. Model Description

This is a supervised classification task ...

II. List of Prohibited Features

religion, nationality, birth place, gender, race

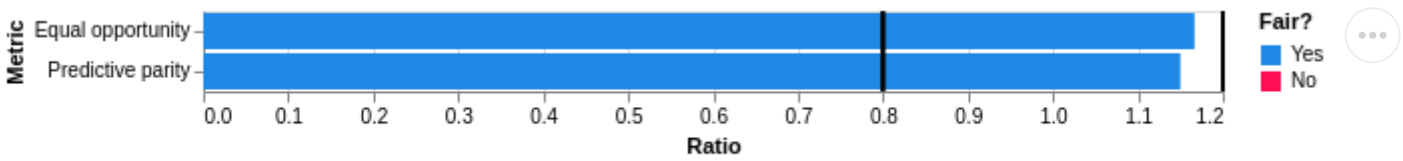
III. Algorithmic Fairness

Algorithmic fairness assesses the models based on technical definitions of fairness. If all are met, the model is deemed to be fair.

Fairness deviation threshold is set at **0.2**. Absolute fairness is 1, so a model is considered fair for the metric when the **metric is between 0.80 and 1.20**.

Prohibited Feature: `Sex=Male`

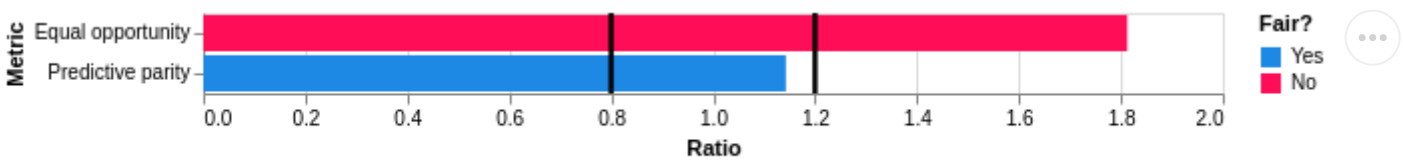
	Metric	Ratio	Fair?
0	Equal opportunity	1.166667	Yes
1	Predictive parity	1.150000	Yes



Overall: **Fair**

Prohibited Feature: `Race=White`

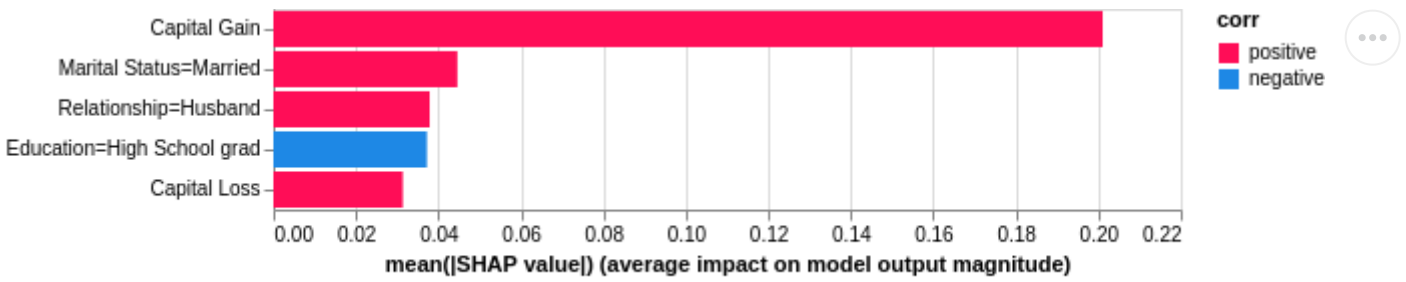
	Metric	Ratio	Fair?
0	Equal opportunity	1.812500	No
1	Predictive parity	1.142857	Yes

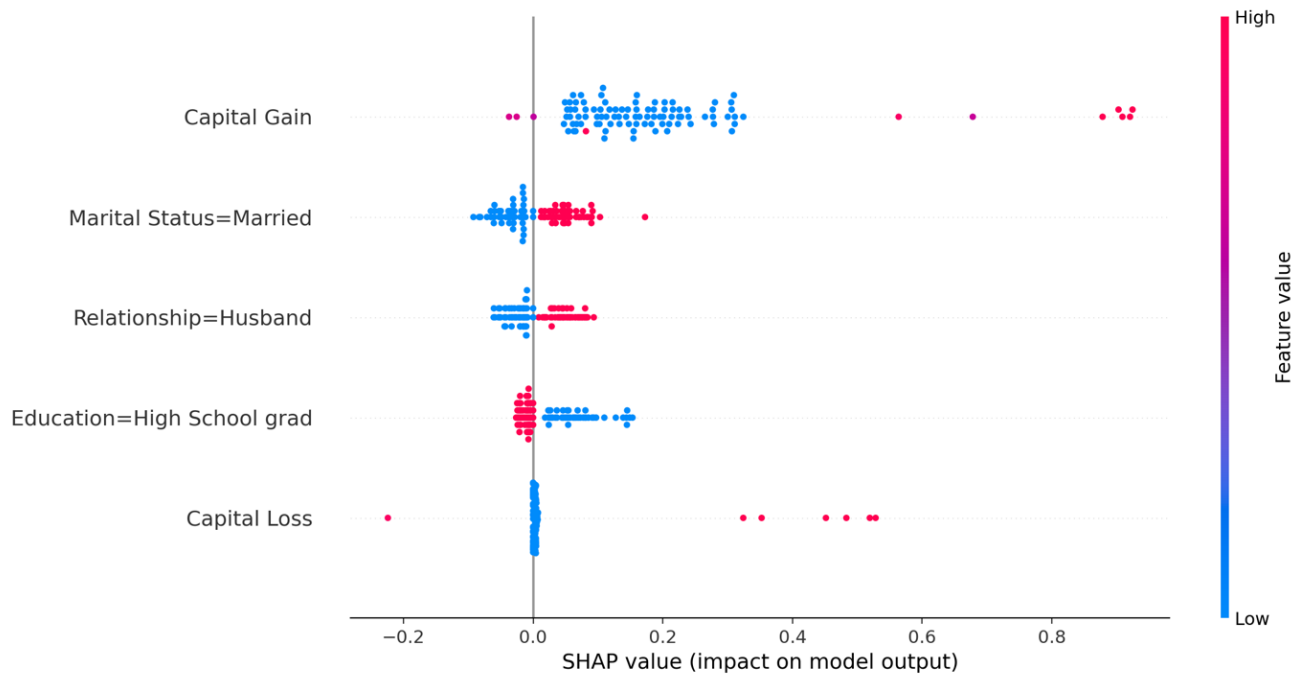


Overall: **Not Fair**

IV. Model Explainability

SHAP Summary Plots of Top Features





The top features are `Capital Gain` , `Marital Status=Married` , `Relationship=Husband` , `Education=High School grad` , `Capital Loss` .

[placeholder]

V. Model Performance

```
Model accuracy = 0.8800
Weighted Average Precision = 0.8800
Weighted Average Recall = 0.8800
```

	precision	recall	f1-score	support
0	0.8800	0.9565	0.9167	69
1	0.8800	0.7097	0.7857	31
accuracy			0.8800	100
macro avg	0.8800	0.8331	0.8512	100
weighted avg	0.8800	0.8800	0.8761	100

VI. Conclusion

Model performance: [placeholder]

Explainability: [placeholder]

The top features that have positive correlation with their model output are `Capital Gain`, `Marital Status=Married`, `Relationship=Husband`, `Capital Loss`.

The top features that have negative correlation with their model output are `Education=High School grad`.

Fairness: We consider the model to be fair if it is deemed to be fair for all metrics. From the table below, overall the model is considered **not fair**.

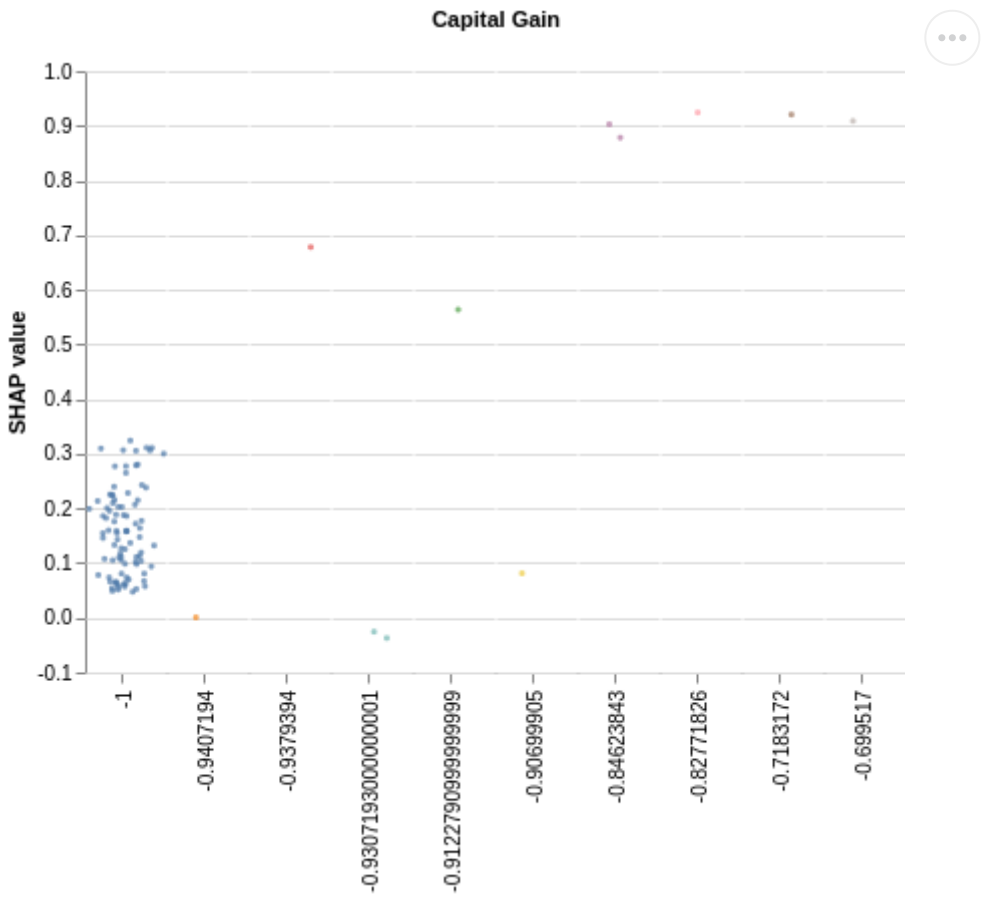
	Prohibited Variable	Fair?
0	Sex=Male-class1	Yes
1	Race=White-class1	No

DBS
Live more, Bank less

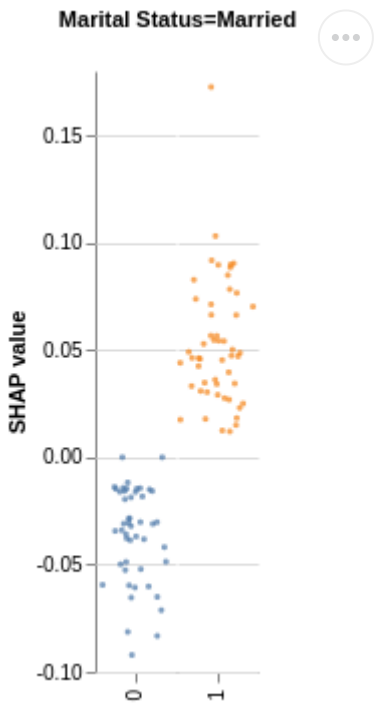
Appendix

Dependence Plots of Top Features

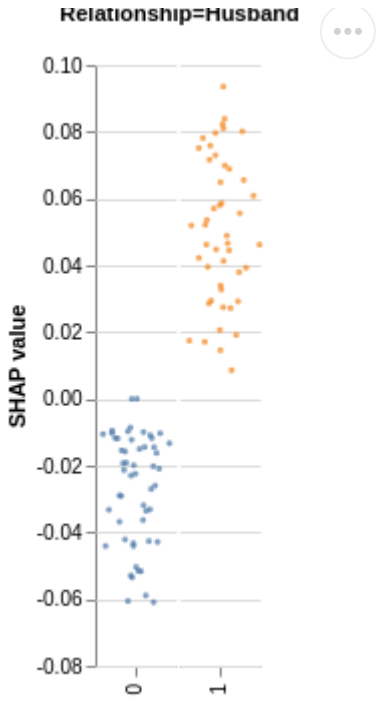
Feature: `Capital Gain`



Feature: `Marital Status=Married`

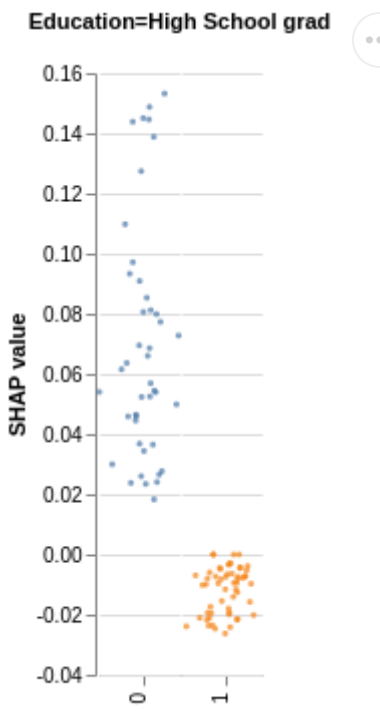


Feature: `Relationship=Husband`

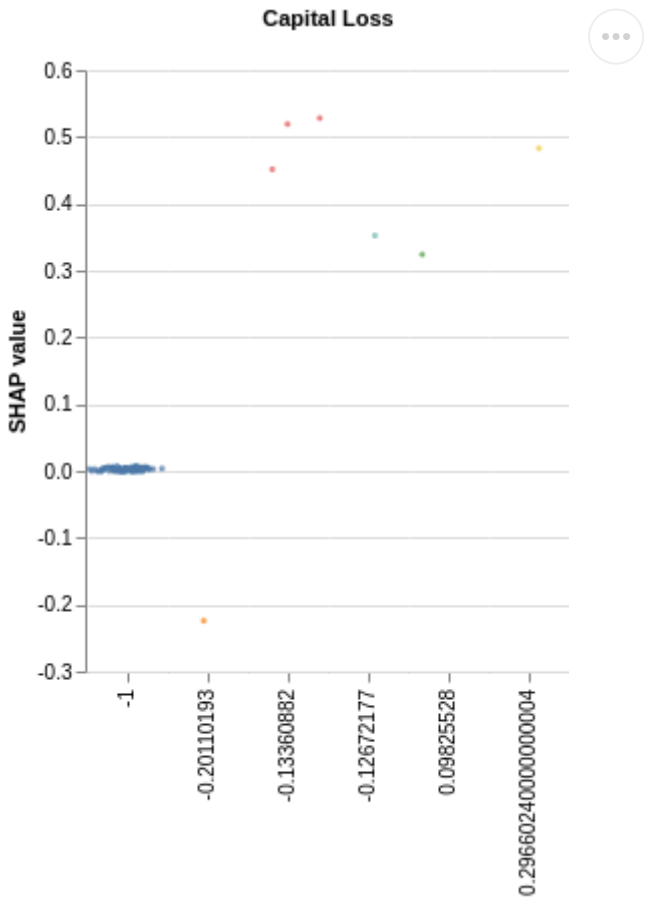


Source file changed. Rerun Always rerun

Feature: `Education=High School grad`

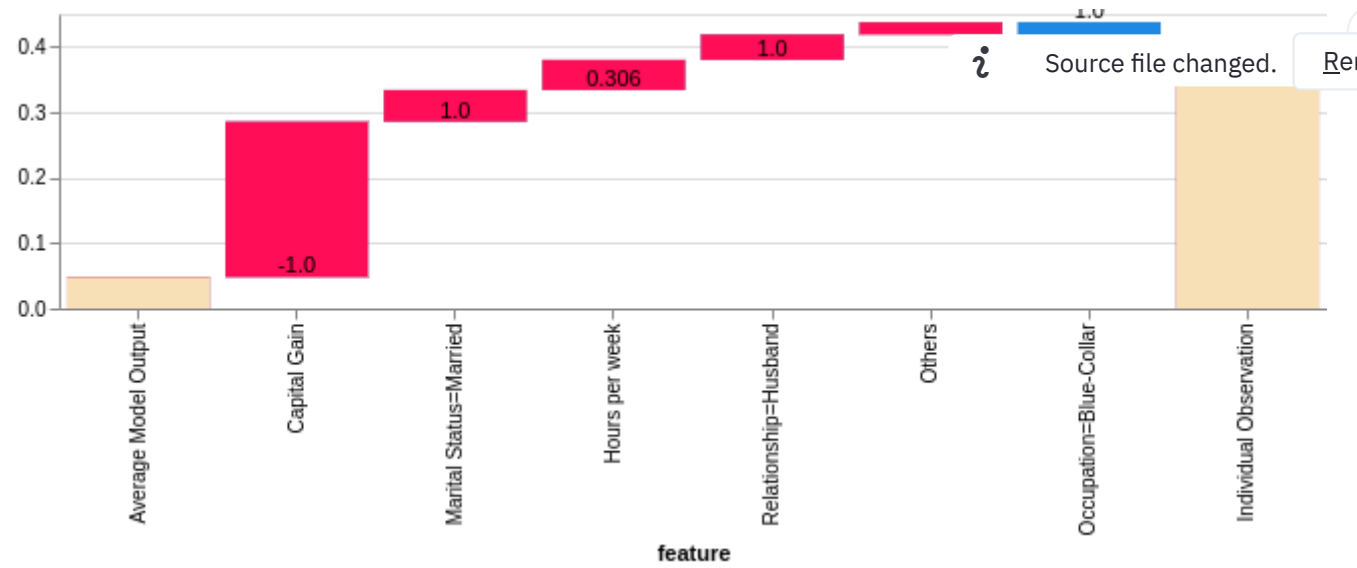


Feature: `Capital Loss`



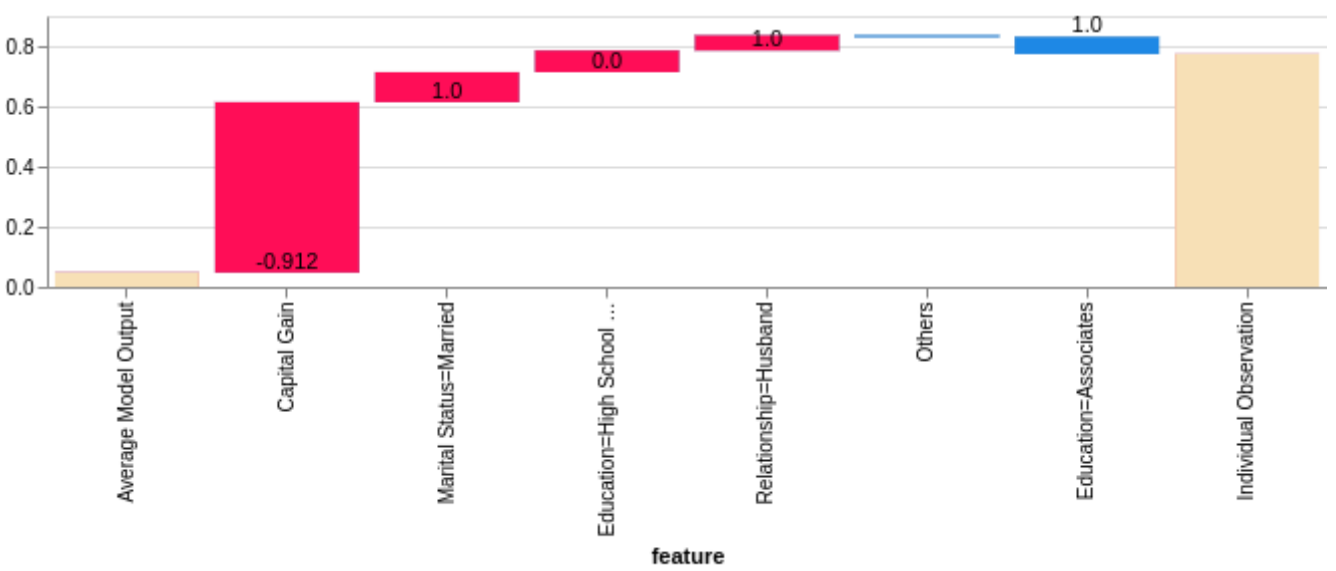
Sample Individual Explainability

Sample from Class=0: SHAP Contribution to Model Prediction



Source file changed. Rerun Always rerun

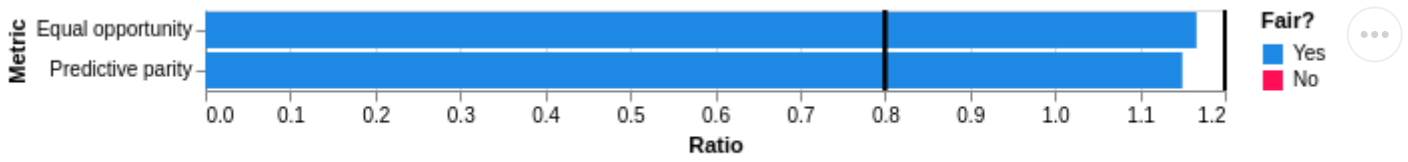
Sample from Class=1: SHAP Contribution to Model Prediction



Algorithmic Fairness

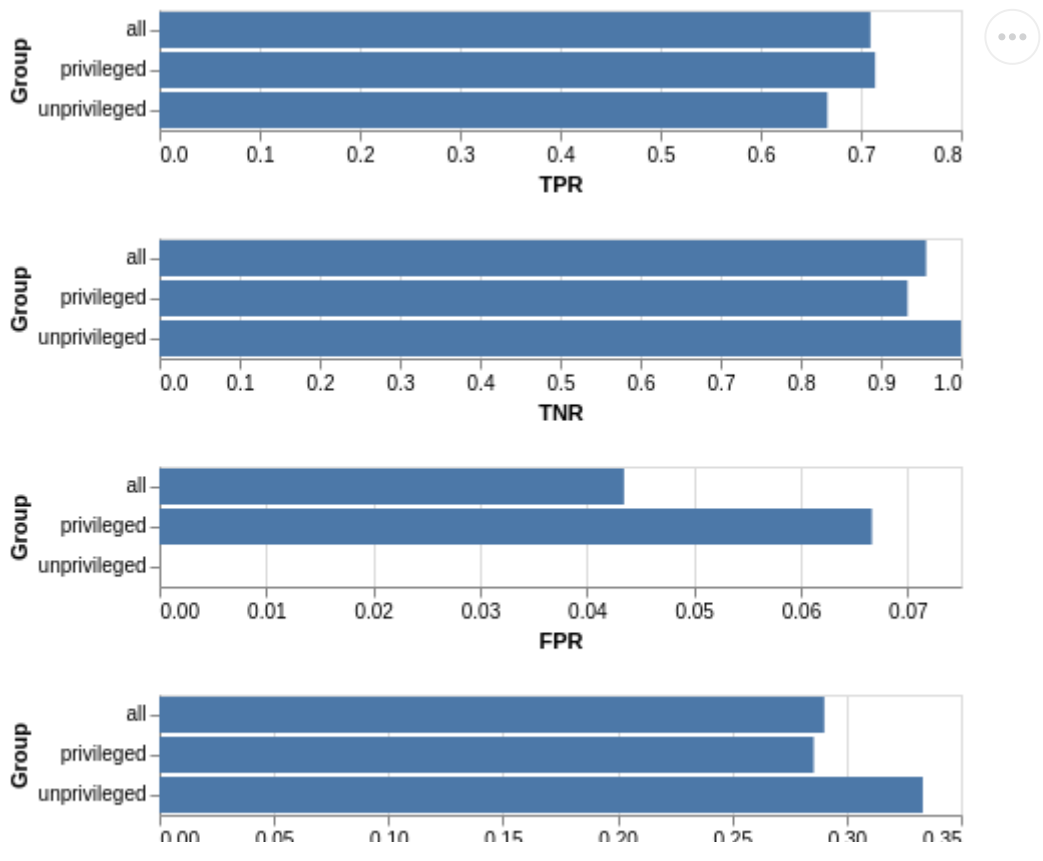
Prohibited Feature: `Sex=Male`

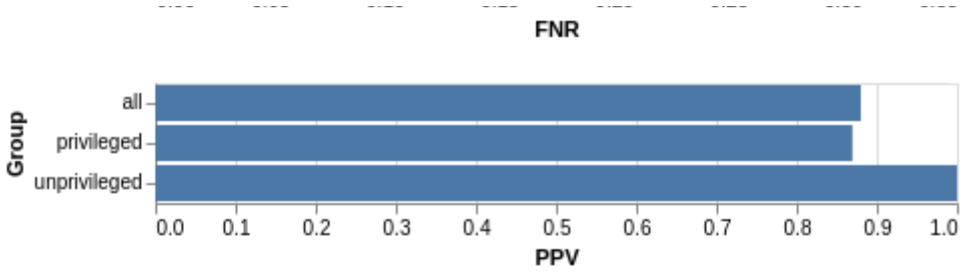
Fairness is when **ratio is between 0.80 and 1.20**.

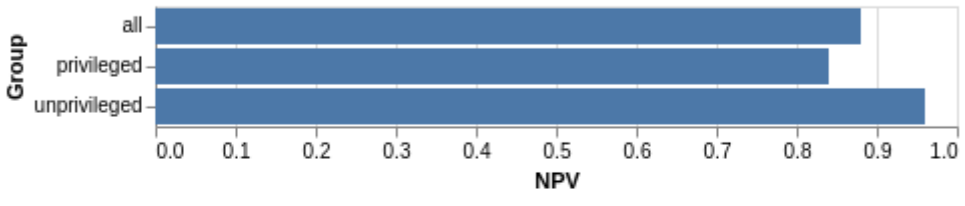


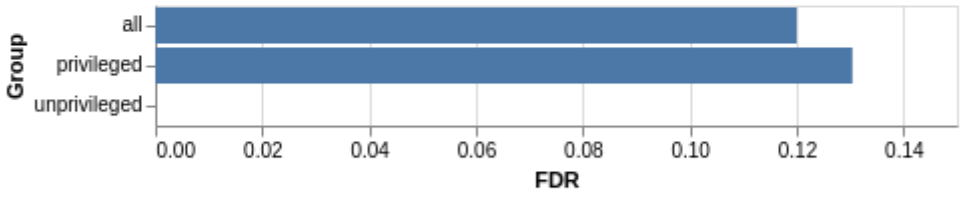
	Metric	Unprivileged	Privileged	Ratio	Fair?
0	Equal opportunity	0.333333	0.285714	1.166667	Yes
1	Predictive parity	1.000000	0.869565	1.150000	Yes

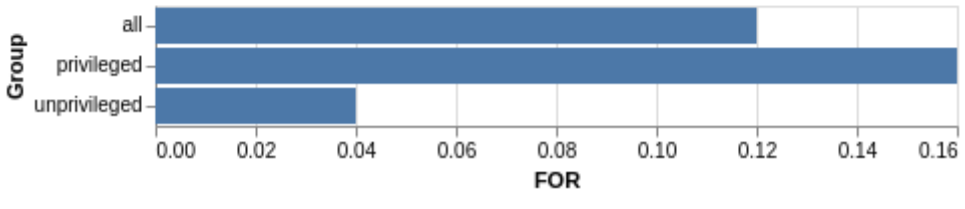
Performance Metrics

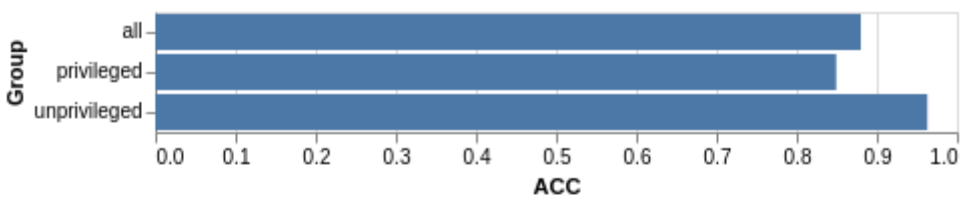


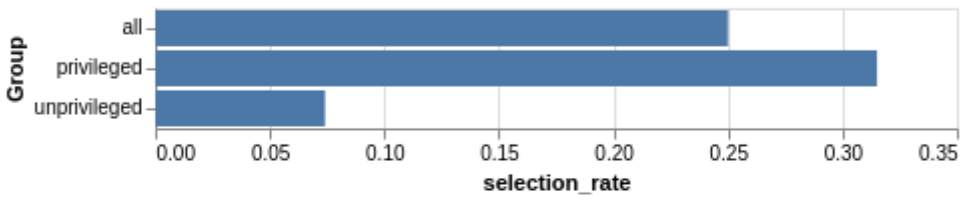


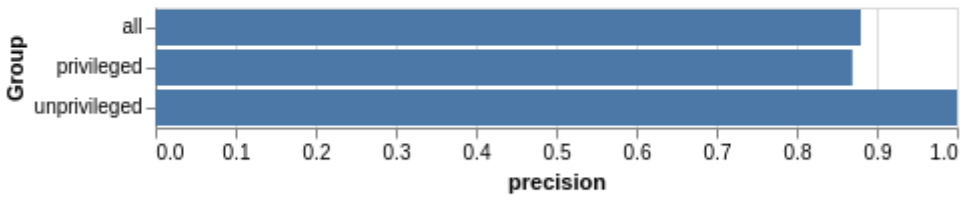


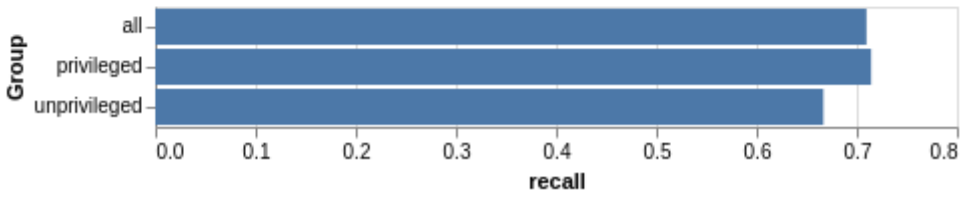


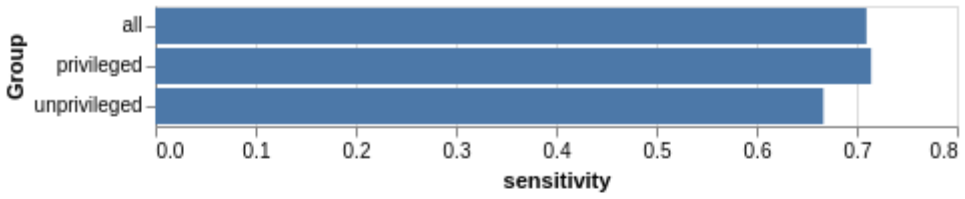


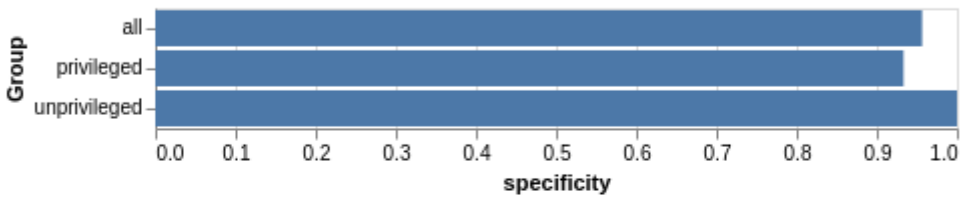


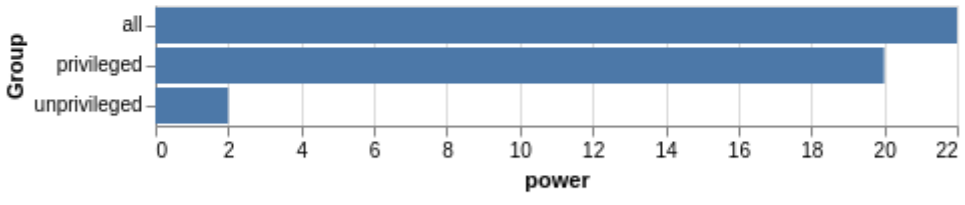


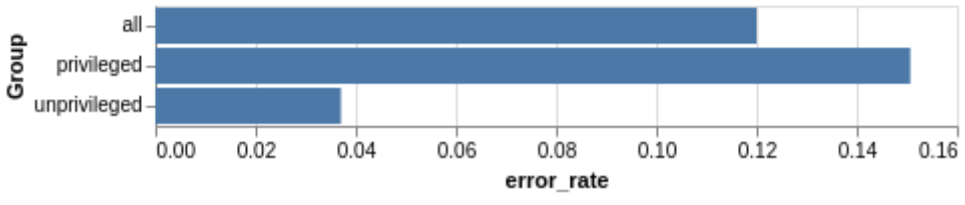




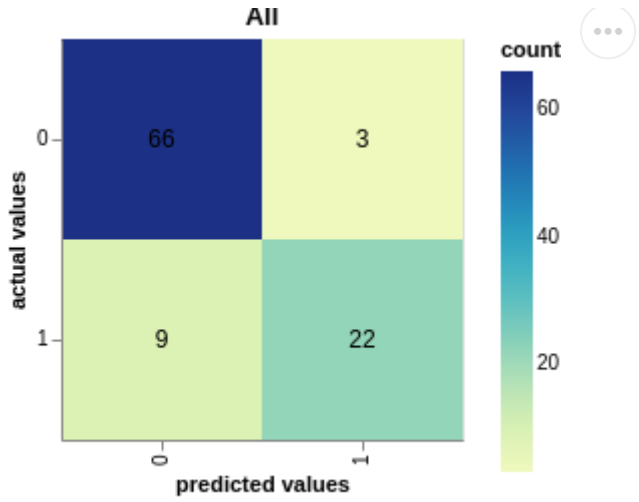




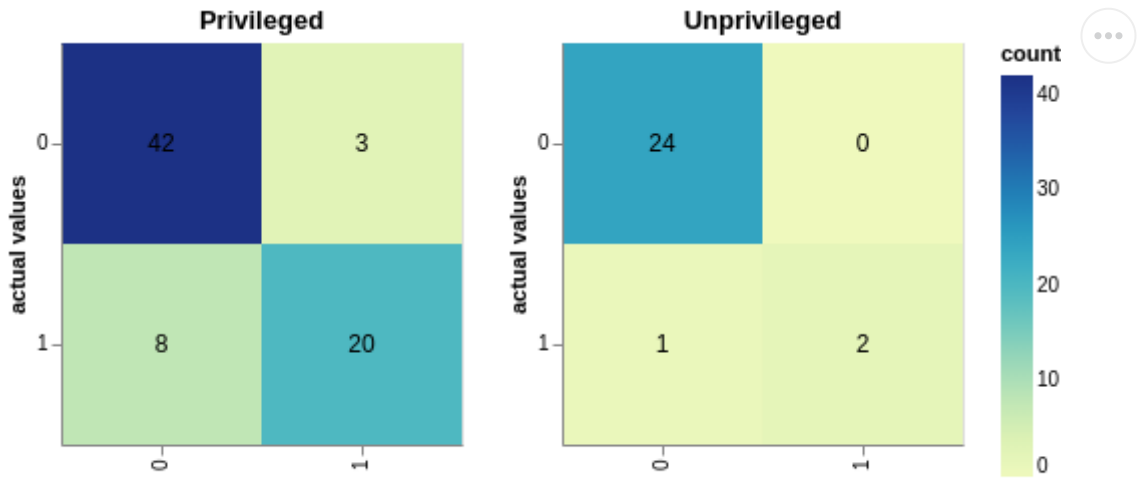




Confusion Matrices

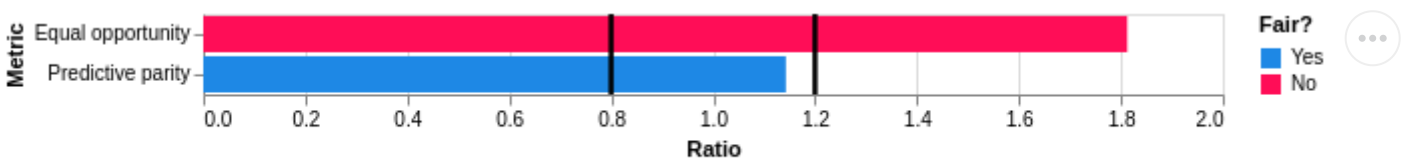


Source file changed. Rerun Always rerun



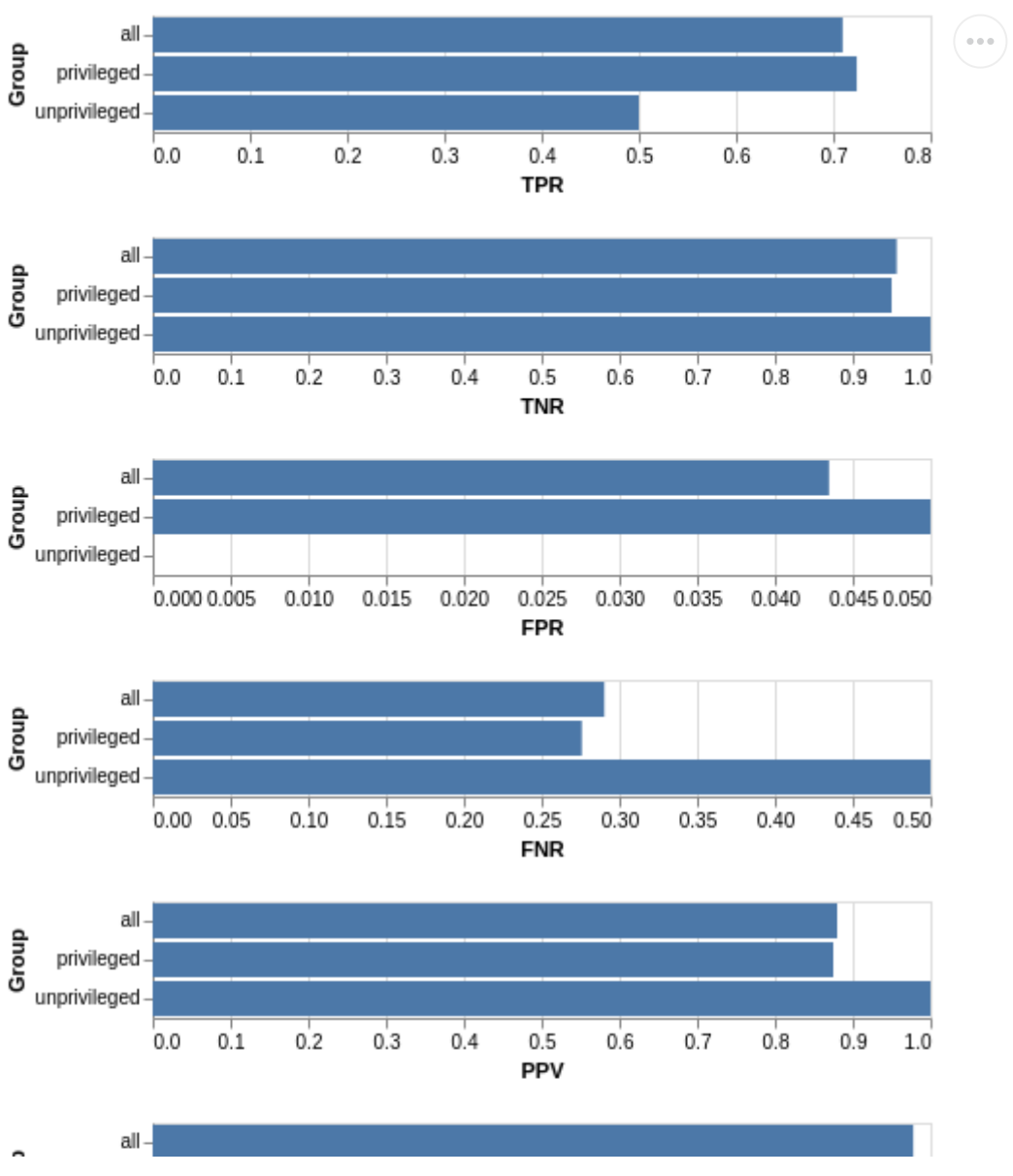
Prohibited Feature: `Race=White`

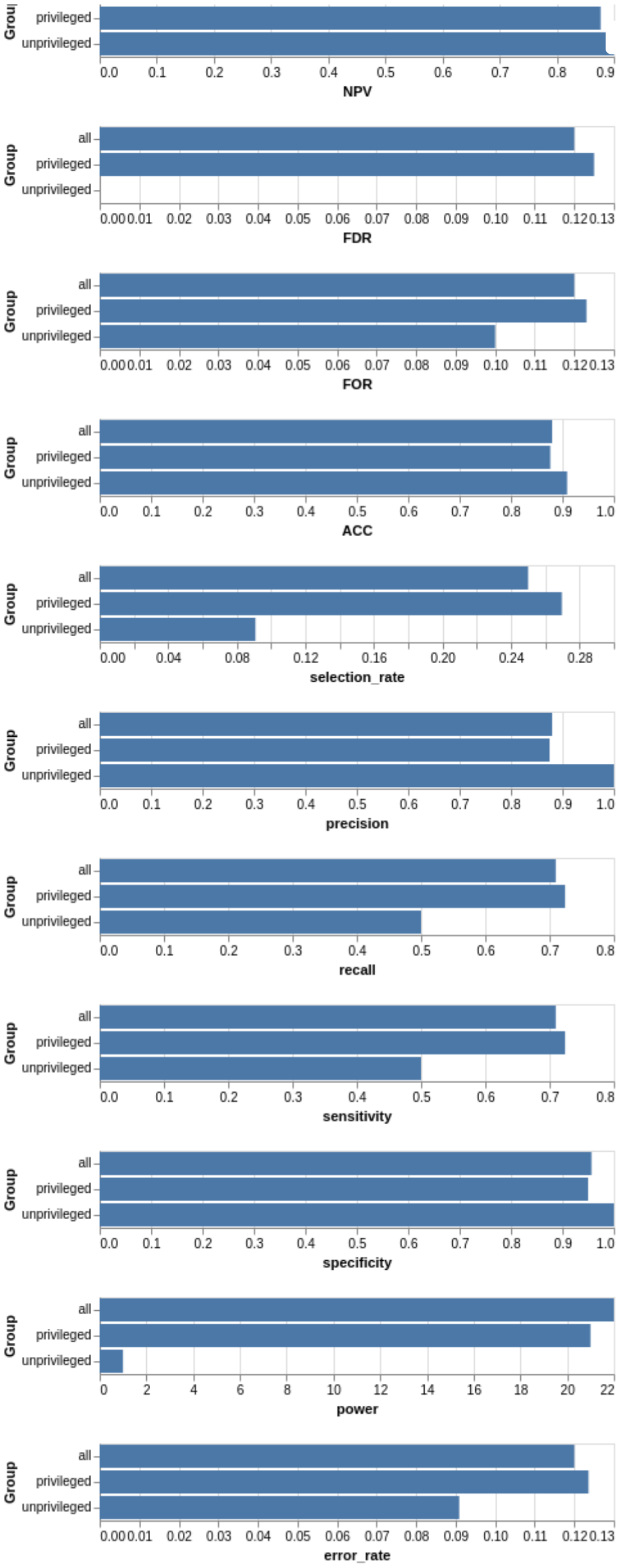
Fairness is when **ratio is between 0.80 and 1.20**.



	Metric	Unprivileged	Privileged	Ratio	Fair?
0	Equal opportunity	0.500000	0.275862	1.812500	No
1	Predictive parity	1.000000	0.875000	1.142857	Yes

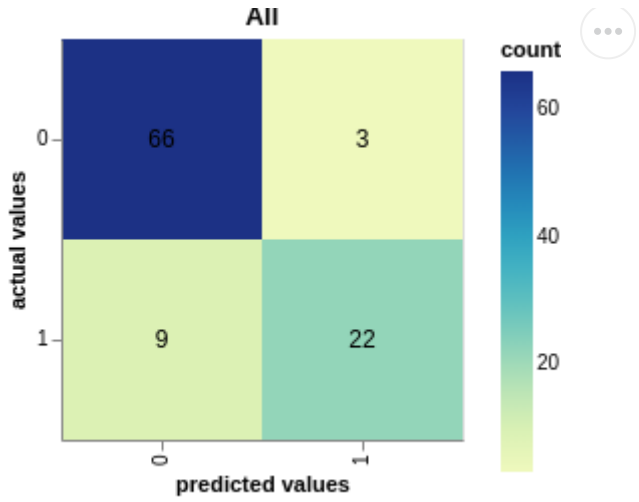
Performance Metrics



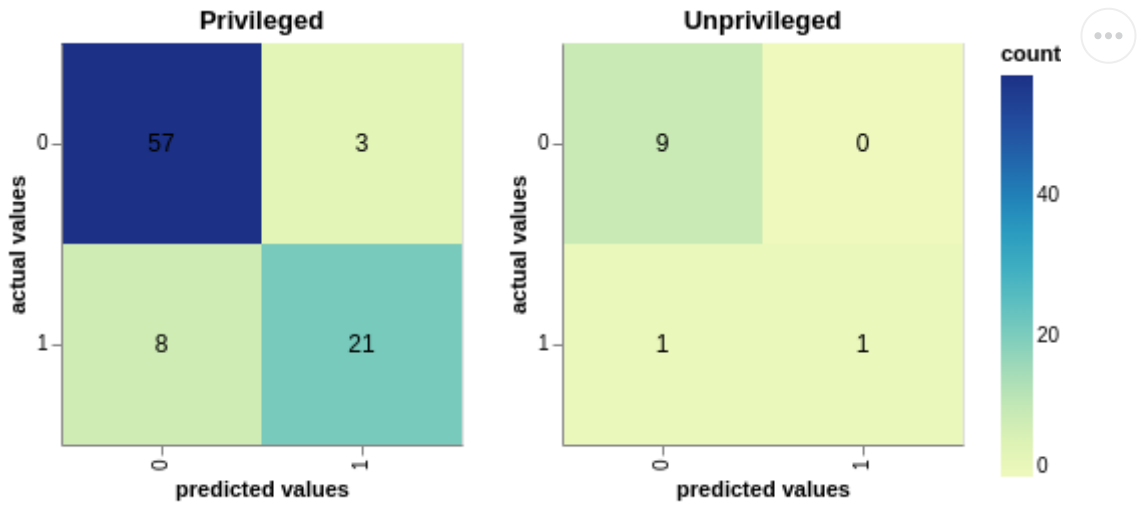


Source file changed. Rerun Always rerun

Confusion Matrices



Source file changed. Rerun Always rerun



Notes

Equal opportunity:

$$\frac{\text{FNR}(D = \text{unprivileged})}{\text{FNR}(D = \text{privileged})}$$

Predictive parity:

$$\frac{\text{PPV}(D = \text{unprivileged})}{\text{PPV}(D = \text{privileged})}$$

Statistical parity:

$$\frac{\text{Selection Rate}(D = \text{unprivileged})}{\text{Selection Rate}(D = \text{privileged})}$$