DBS

Live more, Bank less

CBG Analytics

Model Performance, Fairness and Explainability Report

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

DBS
Live more, Bank less

I. Model Description

This is a supervised classification task for credit default risk model. The objective is to use historical loan application data to predict whether or not an applicant will be able to repay a loan. The target is a 0 for the loan was repaid on time, or a 1 indicating the client had payment difficulties. There are over 750 features/input variables that includes CODE_GENDER, FLAG_OWN_CAR, AMT_INCOME_TOTAL, AMT_CREDIT, NAME_EDUCATION_TYPE, OCCUPATION_TYPE and NAME_HOUSING_TYPE.

II. List of Prohibited Features

religion, nationality, birth place, gender, race

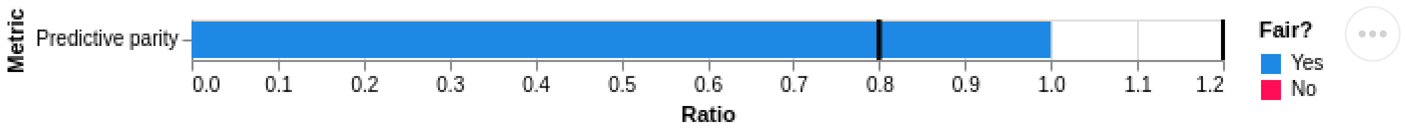
III. Algorithmic Fairness

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: `feat_126`

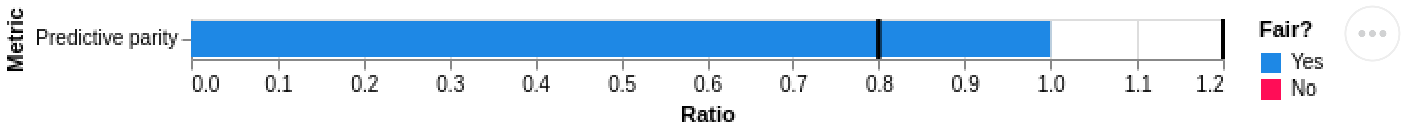
	Metric	Ratio	Fair?
0	Equal opportunity	inf	No
1	Predictive parity	1.000000	Yes



Overall: **Not Fair**

Prohibited Feature: `feat_180`

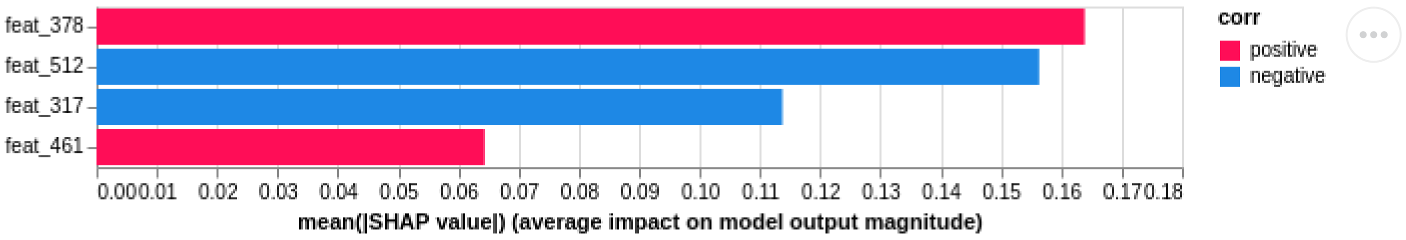
	Metric	Ratio	Fair?
0	Equal opportunity	inf	No
1	Predictive parity	1.000000	Yes

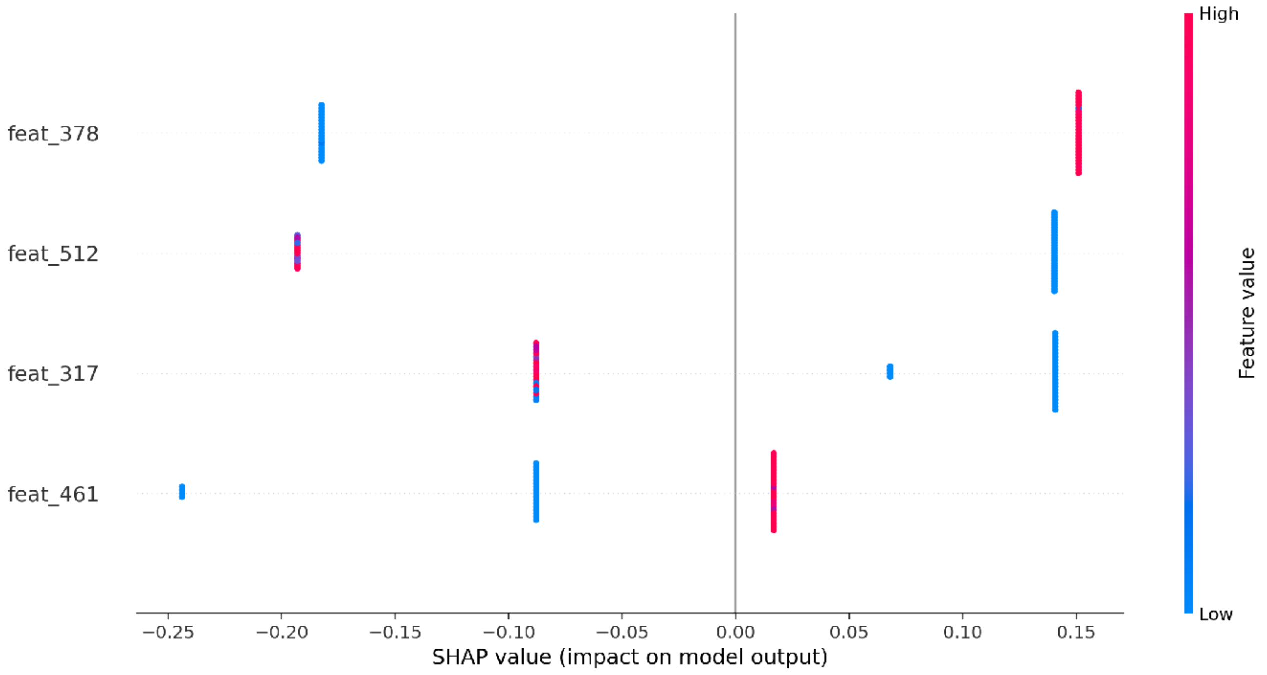


Overall: **Not Fair**

IV. Model Explainability

SHAP Summary Plots of Top Features





The top features are `feat_378` , `feat_512` , `feat_317` , `feat_461` .

EXT_SOURCE_2, EXT_SOURCE_3, EXT_SOURCE_1 are scores derived from past records of the client transactions with our bank. Thus, they are important predictors, which are correctly reflected in the feature importance plots.

V. Model Performance

```
Model accuracy = 0.9565

              precision    recall  f1-score   support

         0.0     0.9048    1.0000    0.9500        19
         1.0     1.0000    0.9259    0.9615        27

    accuracy                         0.9565        46
   macro avg     0.9524    0.9630    0.9558        46
weighted avg     0.9607    0.9565    0.9568        46
```

VI. Conclusion

Model performance: As the precision and recall values are above 70%, the model is considered performing well.

Explainability: Having reviewed the plots in Section 4, we have assessed that the prohibited features do not contribute significantly to the model performance. The direction of the feature impact is also as expected. We are confident that if asked to explain the key factors in the model, a clear explanation can be given to stakeholders and customers.

The top features that have positive correlation with their model output are `feat_378` , `feat_461` .

The top features that have negative correlation with their model output are `feat_512` , `feat_317` .

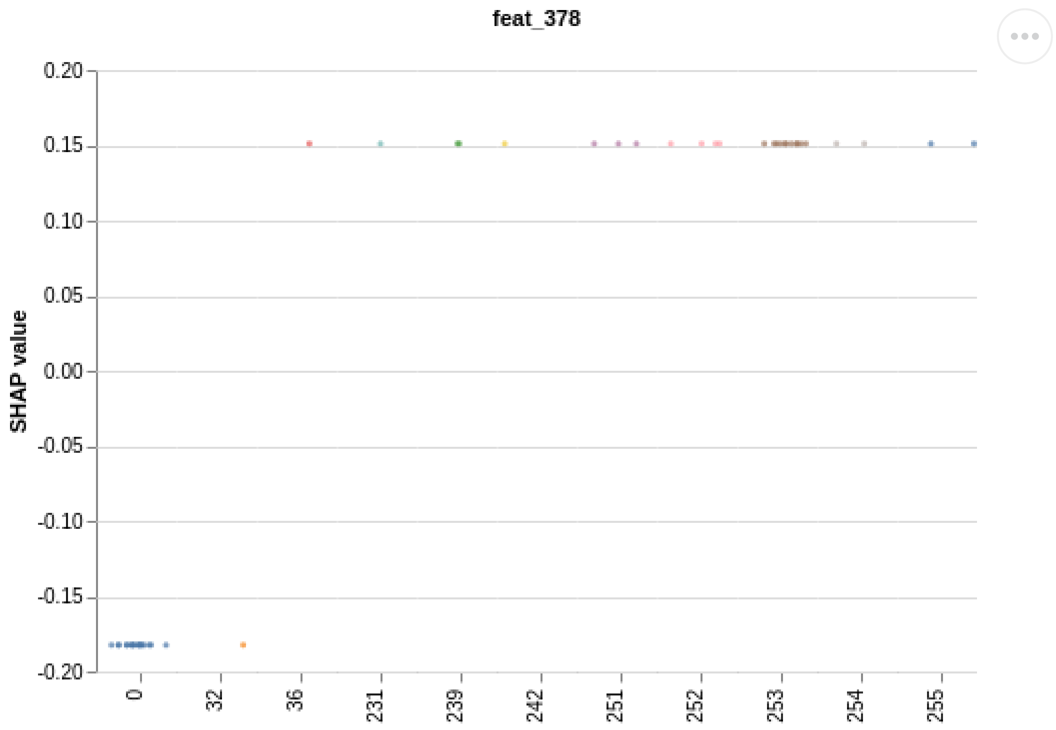
Fairness: We consider the model to be fair if it is deemed to be fair for all metrics. From the table below, overall the model is considered **not fair**.

	Prohibited Variable	Fair?
0	feat_126-class1.0	No
1	feat_180-class1.0	No

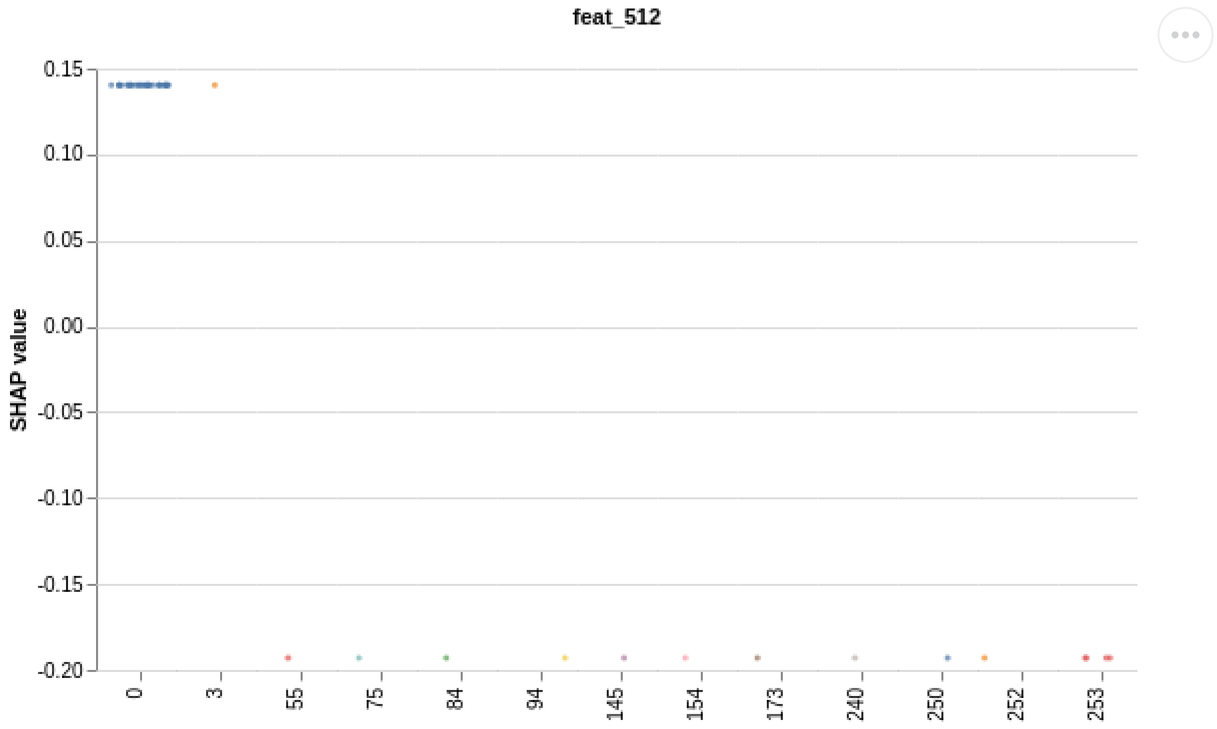
Appendix

Dependence Plots of Top Features

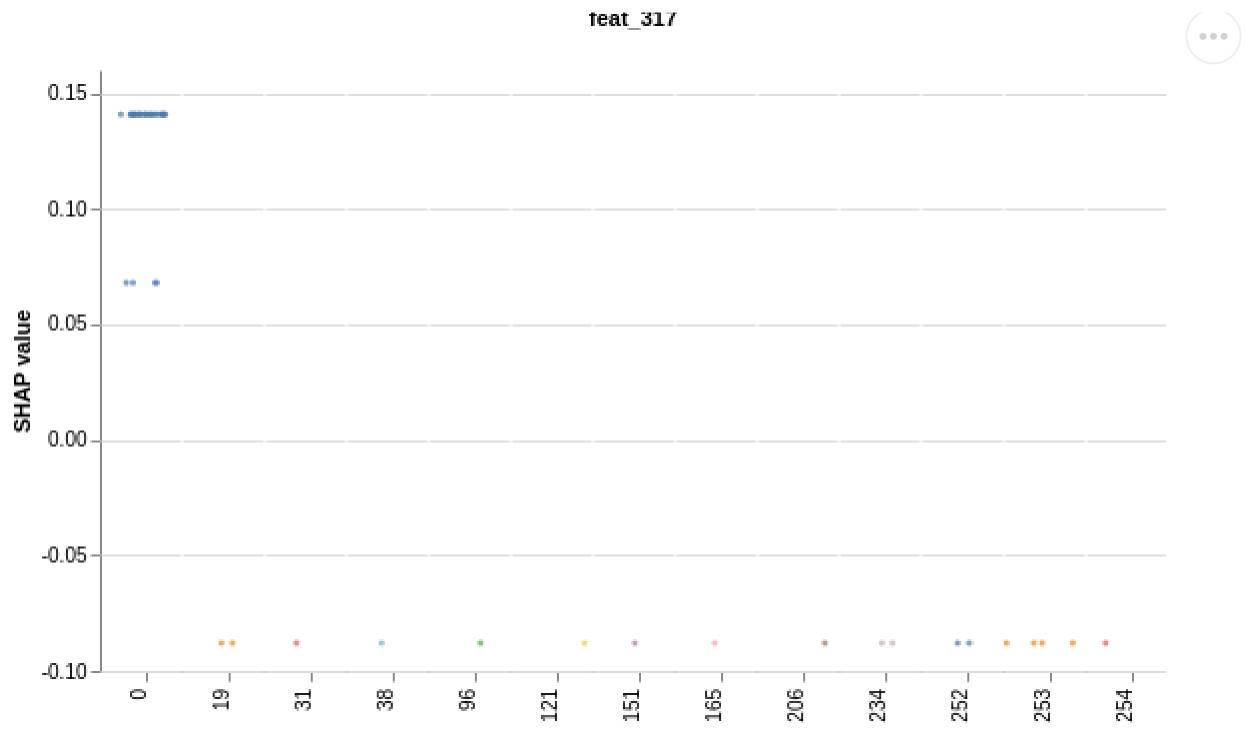
Feature: feat_378



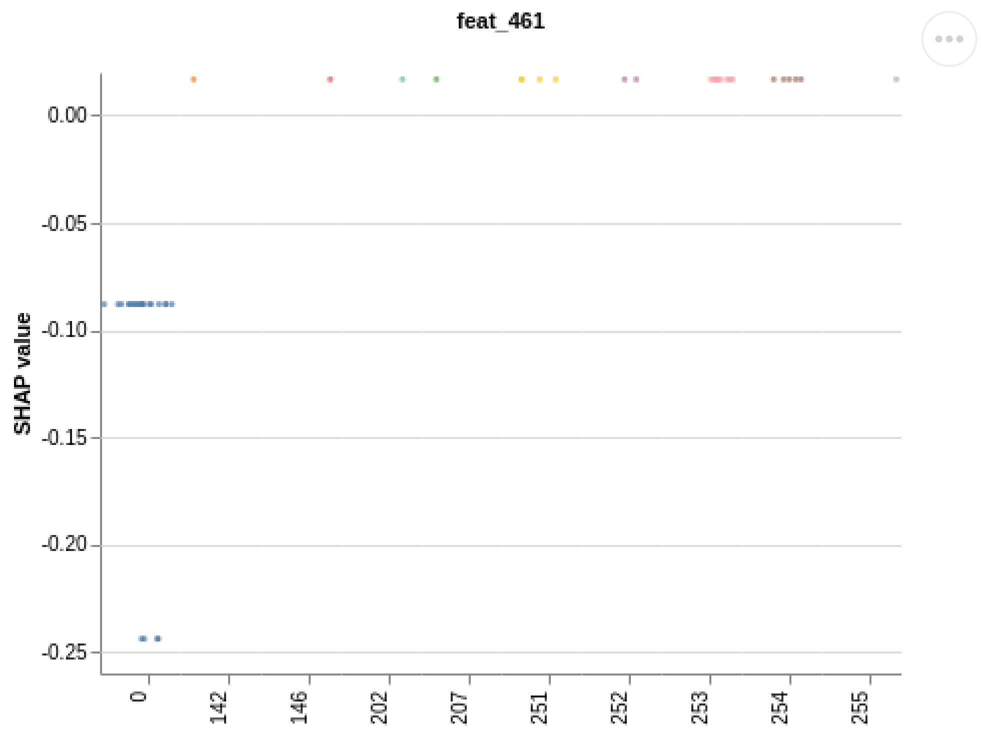
Feature: feat_512



Feature: feat_317

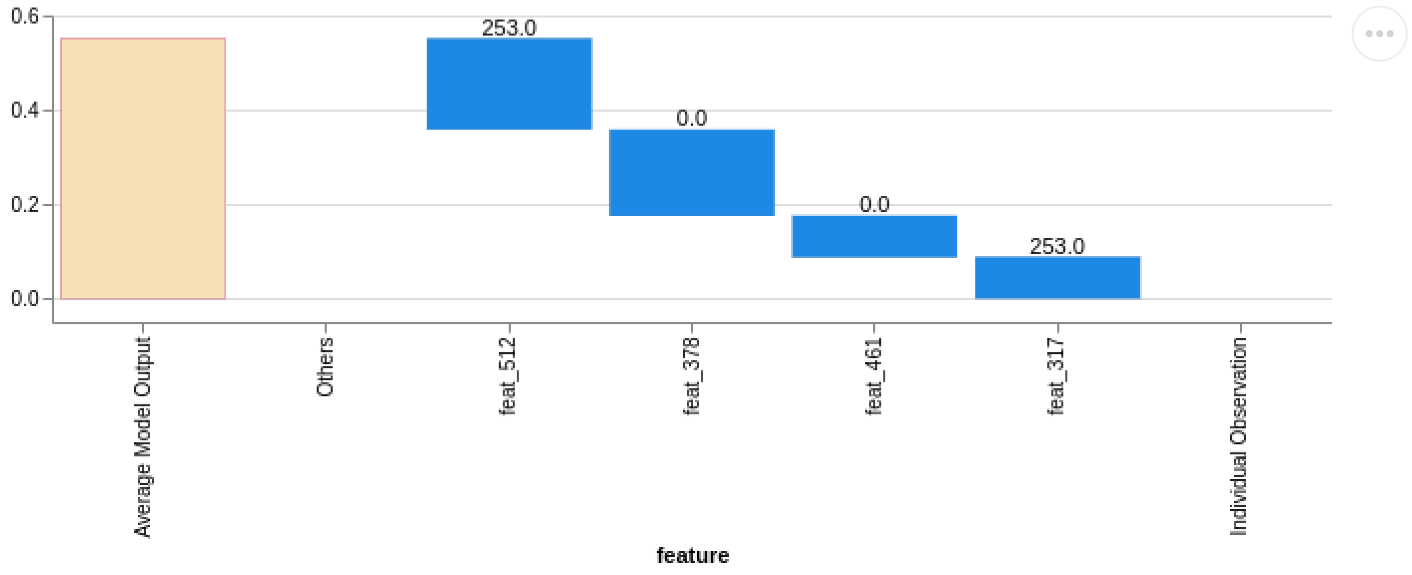


Feature: `feat_461`

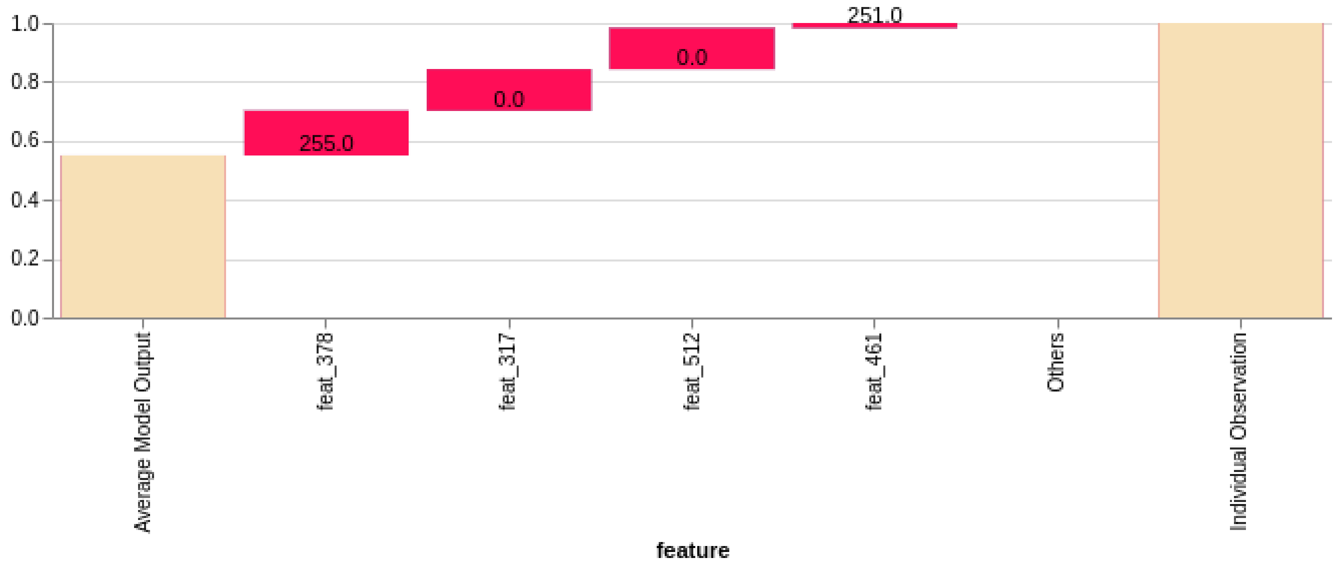


Sample Individual Explainability

Sample from Class=0: SHAP Contribution to Model Prediction



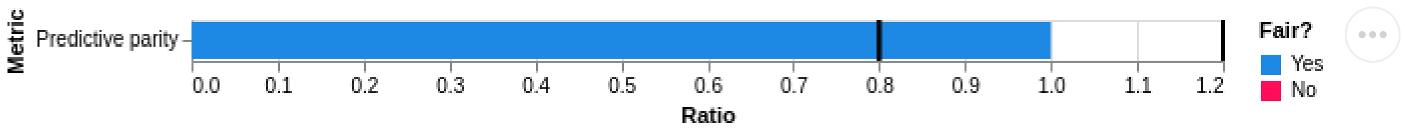
Sample from Class=1: SHAP Contribution to Model Prediction



Algorithmic Fairness

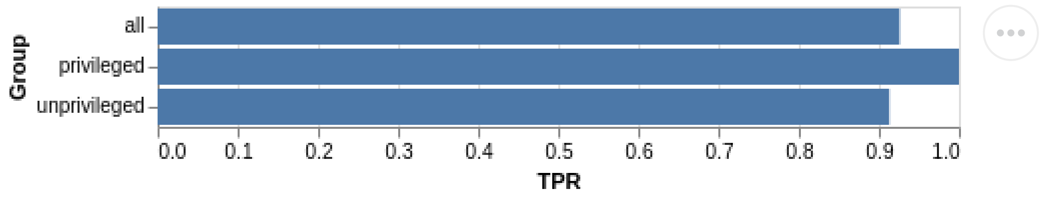
Prohibited Feature: `feat_126`

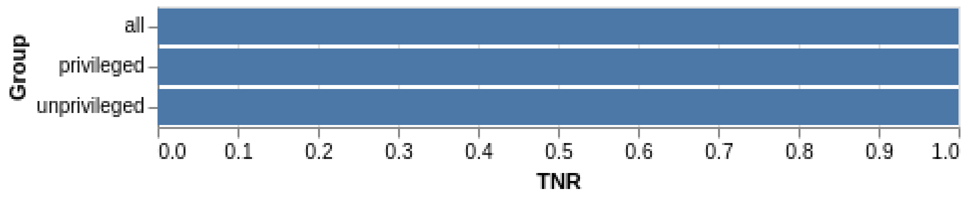
Fairness is when **ratio is between 0.80 and 1.20**.

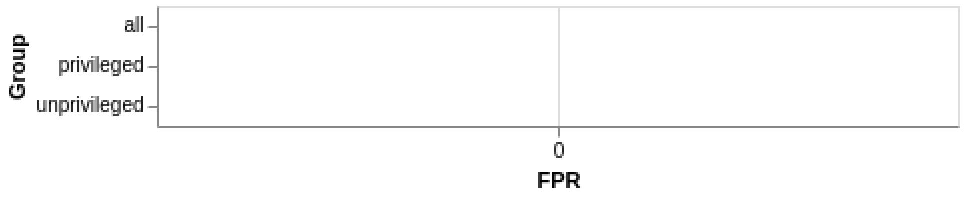


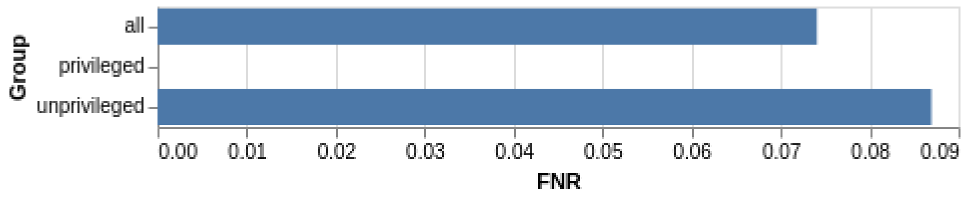
	Metric	Unprivileged	Privileged	Ratio	Fair?
0	Equal opportunity	0.086957	0.000000	inf	No
1	Predictive parity	1.000000	1.000000	1.000000	Yes

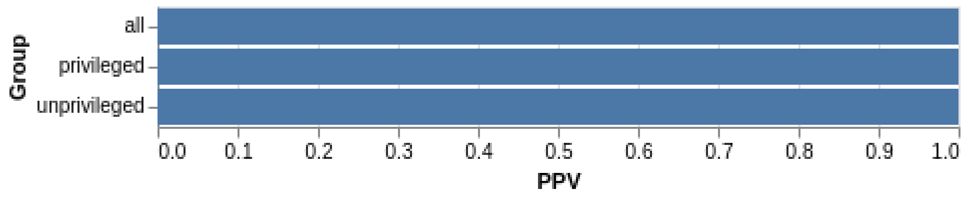
Performance Metrics

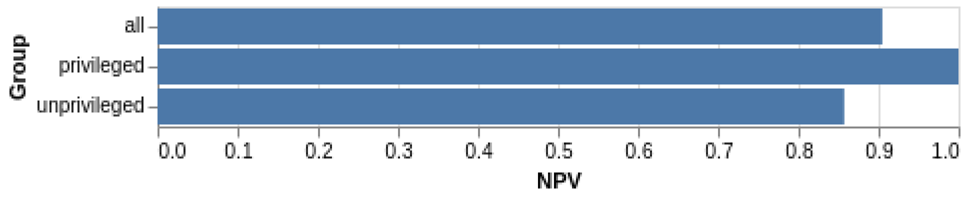


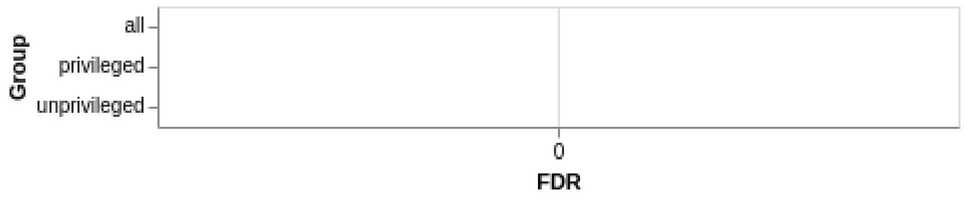


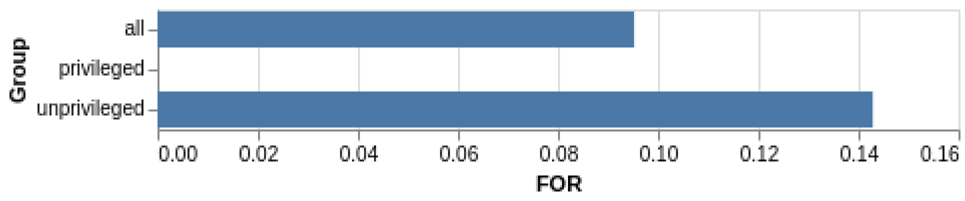


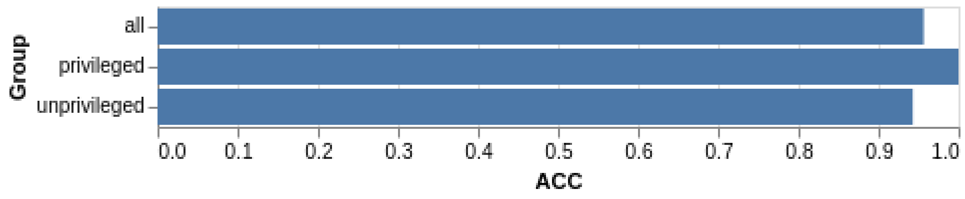


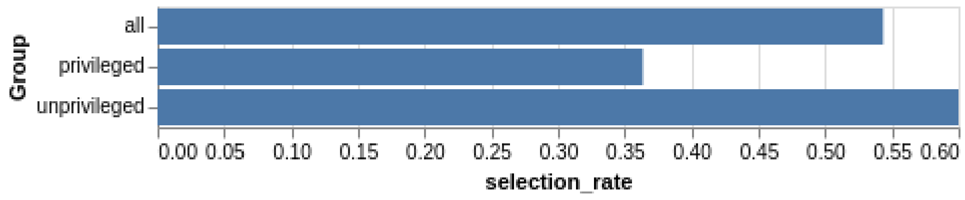


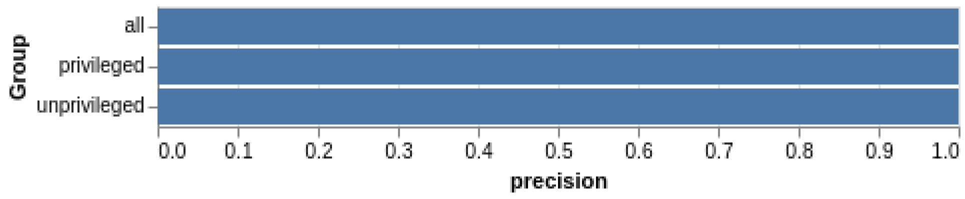


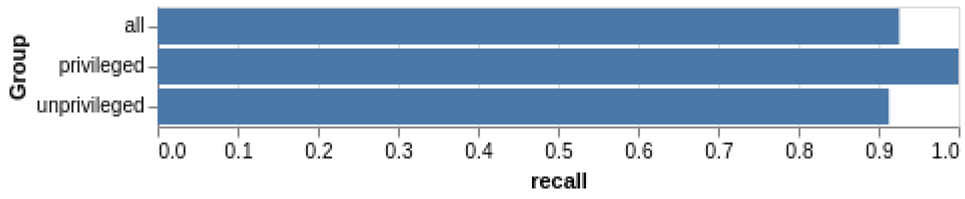


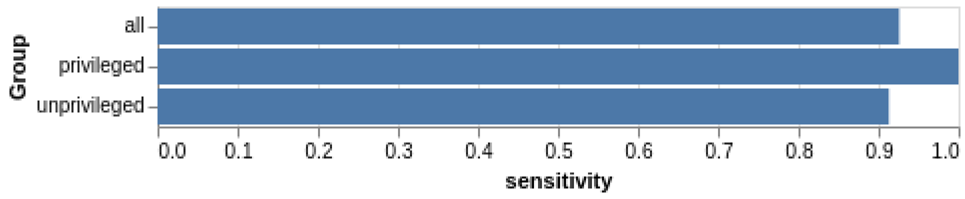


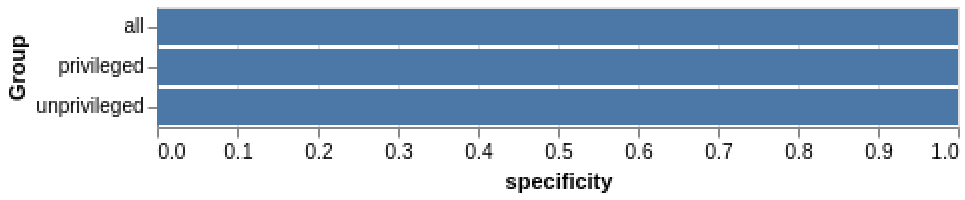


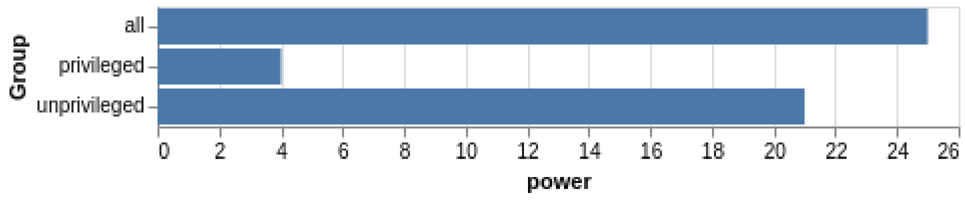


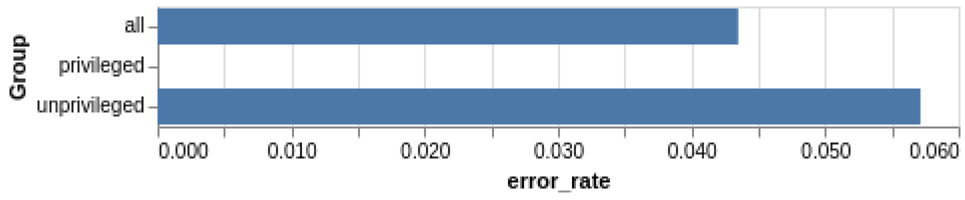




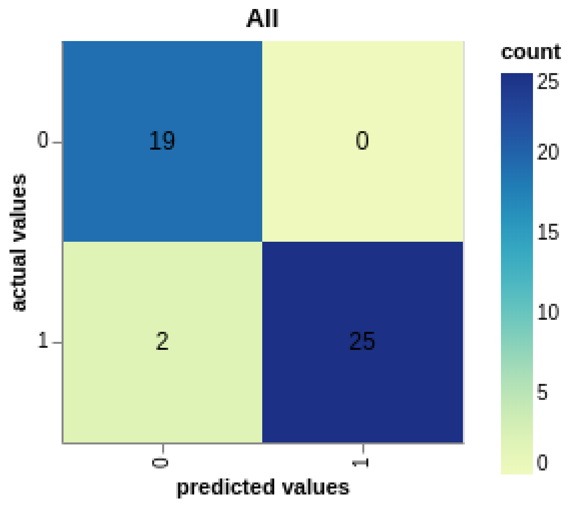


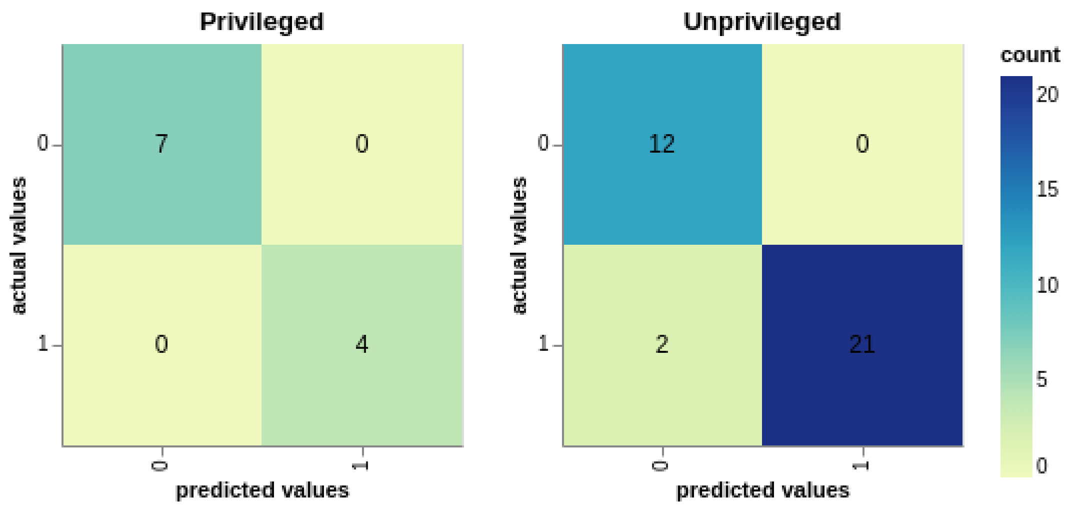






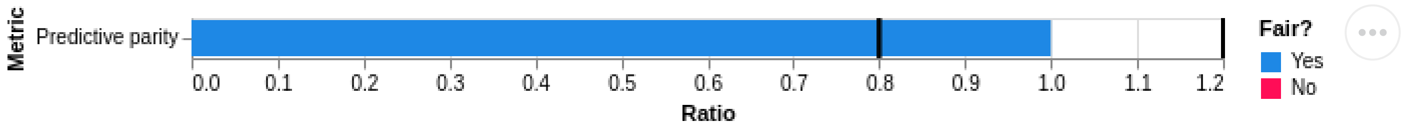
Confusion Matrices





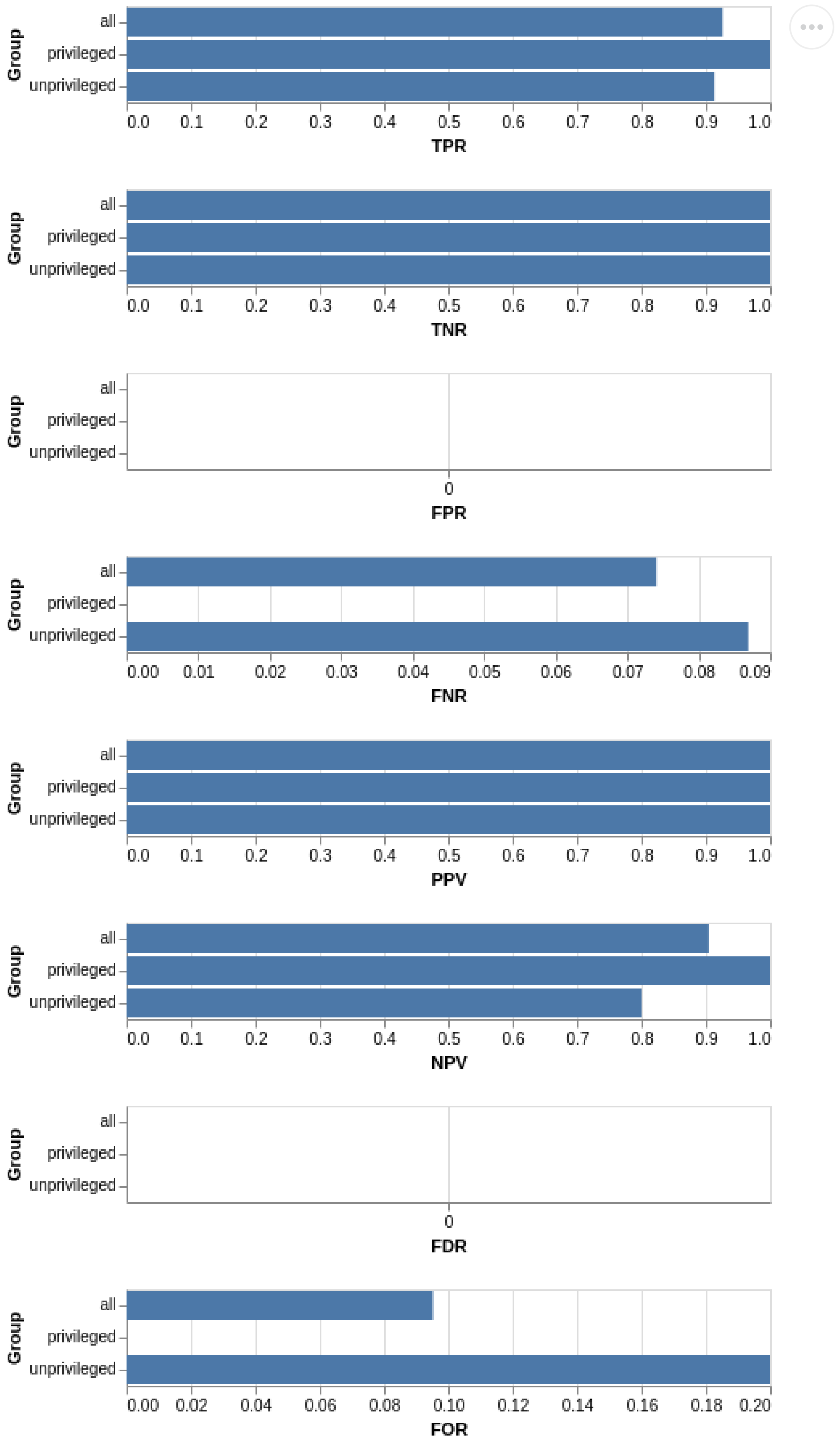
Prohibited Feature: `feat_180`

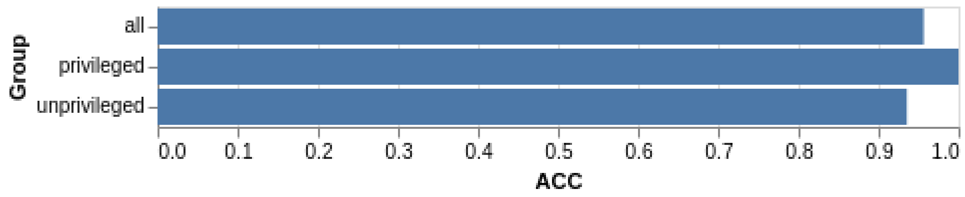
Fairness is when **ratio is between 0.80 and 1.20**.

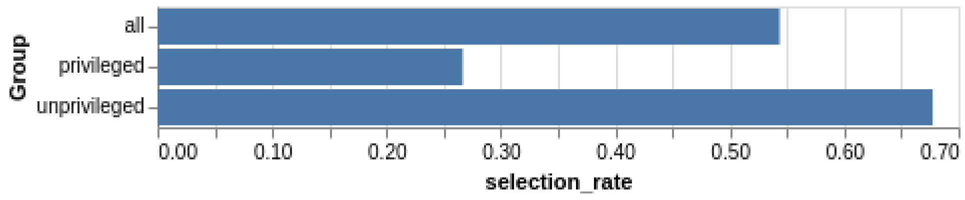


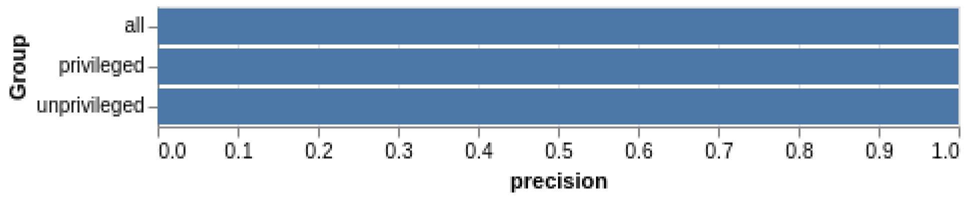
	Metric	Unprivileged	Privileged	Ratio	Fair?
0	Equal opportunity	0.086957	0.000000	inf	No
1	Predictive parity	1.000000	1.000000	1.000000	Yes

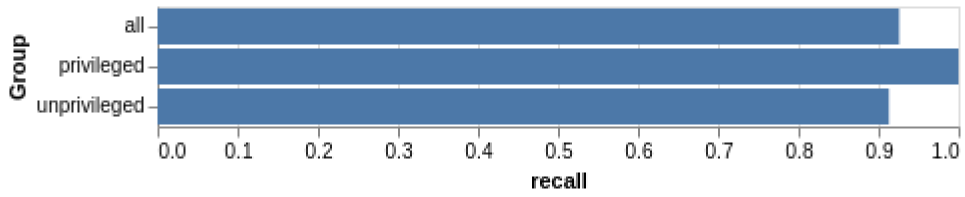
Performance Metrics

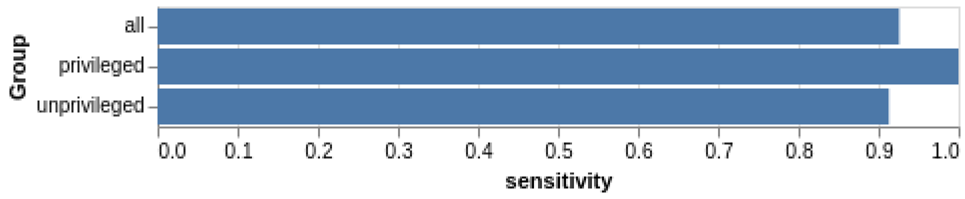


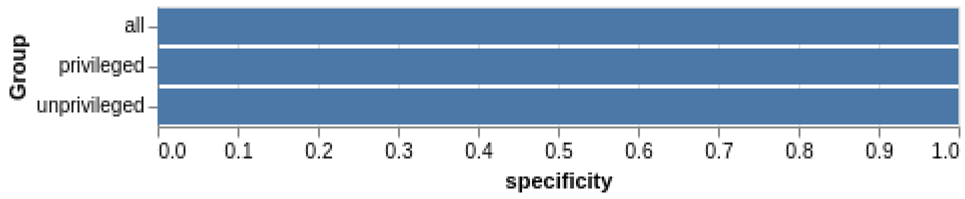


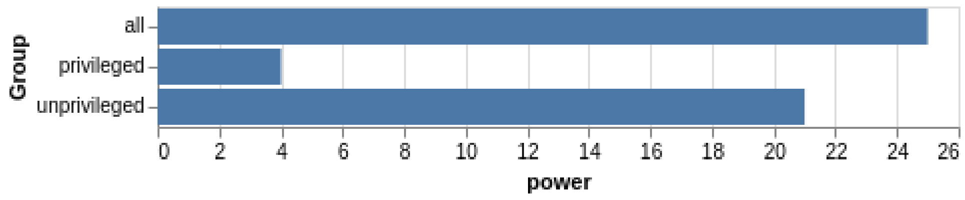


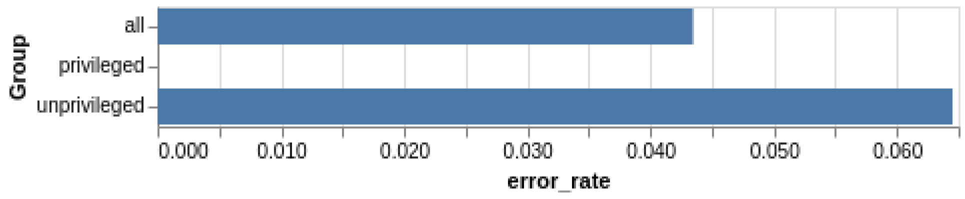




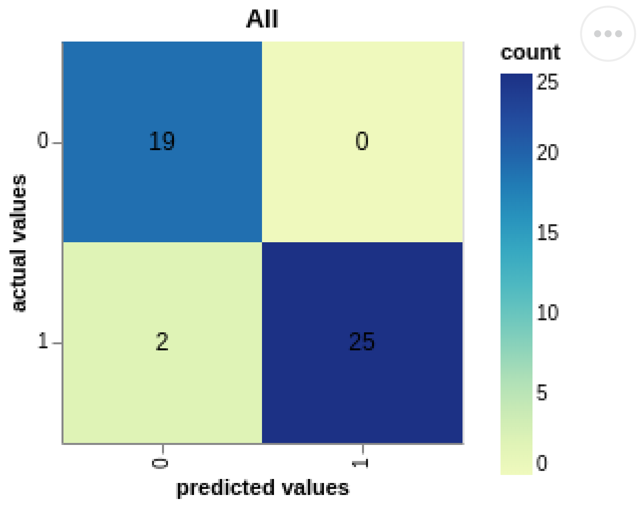


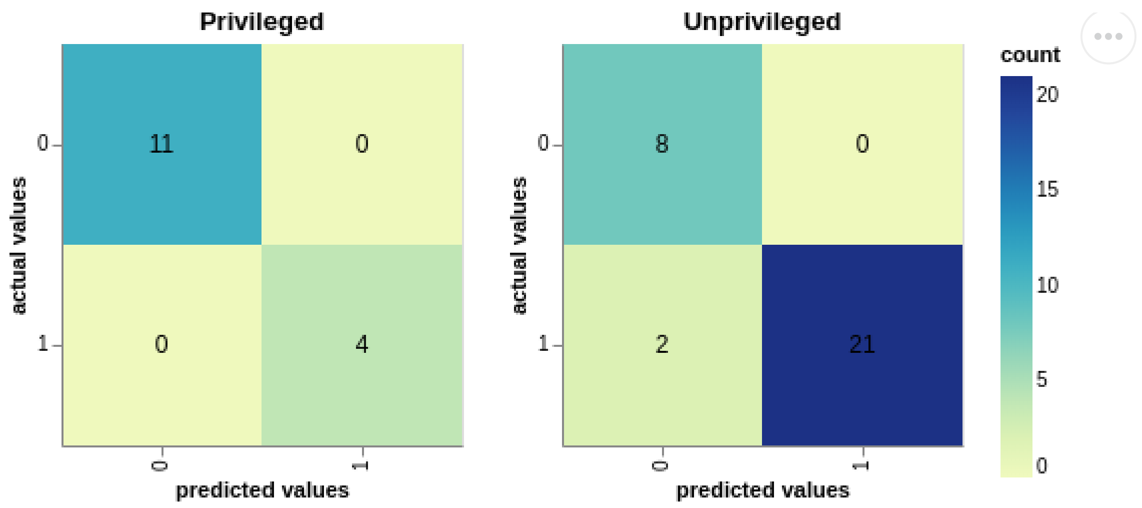






Confusion Matrices





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:

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