selected protected attribute - gender
privileged group is assumed to be male


Basic model

Fairness Check

Equal opportunity ratio   TP/(TP + FN)

Predictive parity ratio   TP/(TP + FP)

Predictive equality ratio   FP/(FP + TN)

Accuracy equality ratio   (TP + TN)/(TP + FP + TN + FN)

Statistical parity ratio   (TP + FP)/(TP + FP + TN + FN)

XGBClassifier
Subgroup: 0
Score: 0.26

label
■ XGBClassifier

score
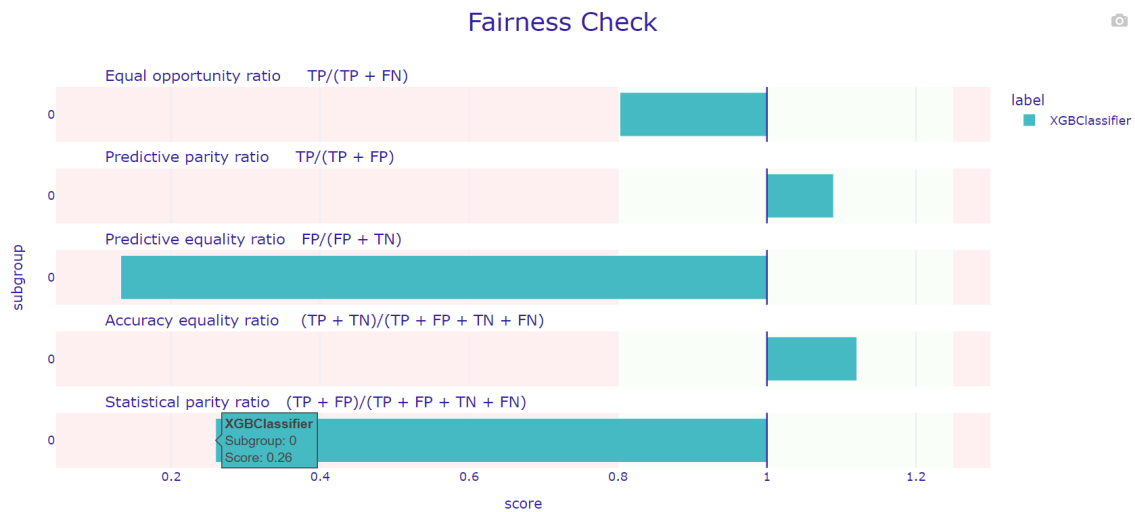
The statistical parity and predictive equality for the basic model are not good - they go well into the ranges of (ε, $1/\varepsilon$) and the equal opportunity statistic is almost in the range, which is good for the model.

---

another model - no gender included in the data

Fairness Check

Equal opportunity ratio   TP/(TP + FN)

Predictive parity ratio   TP/(TP + FP)

Predictive equality ratio   FP/(FP + TN)

Accuracy equality ratio   (TP + TN)/(TP + FP + TN + FN)

Statistical parity ratio   (TP + FP)/(TP + FP + TN + FN)

label
■ XGBClassifier

score
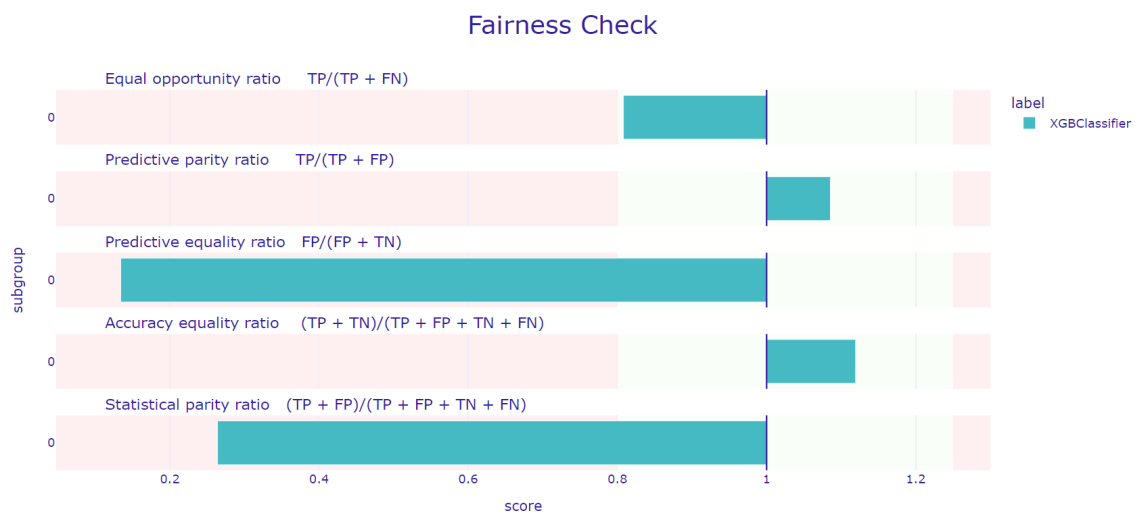
The fairness statistics are almost the same as in the original model which shows that the potential gender bias is hidden, and not only explicitly defined by the input data.

---

bias mitigation - resampling

Fairness Check

Equal opportunity ratio   TP/(TP + FN)

Predictive parity ratio   TP/(TP + FP)

Predictive equality ratio   FP/(FP + TN)

Accuracy equality ratio   (TP + TN)/(TP + FP + TN + FN)

Statistical parity ratio   (TP + FP)/(TP + FP + TN + FN)

label
- XGBClassifier
- XGBClassifier with Resample mitigation

Values of the Statistical parity, Equal opportunity and Predictive parity statistics are all higher for the model with resampling - the mitigation method did work in this case, however the advantage is not overwhelming. Other mitigation methods should be checked whether they give better results.