

| <b>Metric</b>              | <b>Equation</b>                                                 | <b>Definition</b>                                                                                                      |
|----------------------------|-----------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------|
| Equalized Odds             | $P(\hat{Y} = 1   S = a, Y = y) = P(\hat{Y} = 1   S = b, Y = y)$ | Both groups should have equal true positive and false positive rates.                                                  |
| Equal Opportunity          | $P(\hat{Y} = 1   S = a, Y = 1) = P(\hat{Y} = 1   S = b, Y = 1)$ | Both groups should have equal true positive rates. A relaxed version of equalized odds.                                |
| Predictive Equality        | $P(\hat{Y} = 1   Y = 0, S = a) = P(\hat{Y} = 1   Y = 0, S = b)$ | The rate of false positives (negative events categorized as positives) should be independent of the sensitive feature. |
| False Negative Rate Parity | $P(\hat{Y} = 0   Y = 1, S = a) = P(\hat{Y} = 0   Y = 1, S = b)$ | The rate of false negatives (positive events categorized as negatives) should be independent of the sensitive feature. |
| Predictive Parity          | $P(Y = 1   \hat{Y} = 1, S = a) = P(Y = 1   \hat{Y} = 1, S = b)$ | Model precision should be the same for both groups.                                                                    |
| Demographic Parity         | $P(\hat{Y} = 1   S = a) = P(\hat{Y} = 1   S = b)$               | The prediction or decision should be independent of the sensitive feature.                                             |

**Table 1:** Commonly used fairness metrics, adapted from (1, 6).  $\hat{Y}$  = Prediction/decision,  $Y$  = Observed data, and  $S$  = Sensitive feature, in the case of  $S$  being a multi-group variable where a comparison with a reference or privileged group is meaningful.  $\hat{Y}$  and  $Y$  are binary variables.