Algorithmic Fairness

📊 Fairness Definitions

1 Demographic Parity

$P(\hat{Y}=1|A) = P(\hat{Y}=1|B)$

Equal positive prediction rates

2 Equalized Odds

Equal TPR, FPR

Equal accuracy for true positive/negative

3 Predictive Parity

Equal PPV by group

Equal accuracy of positive predictions

4 Individual Fairness
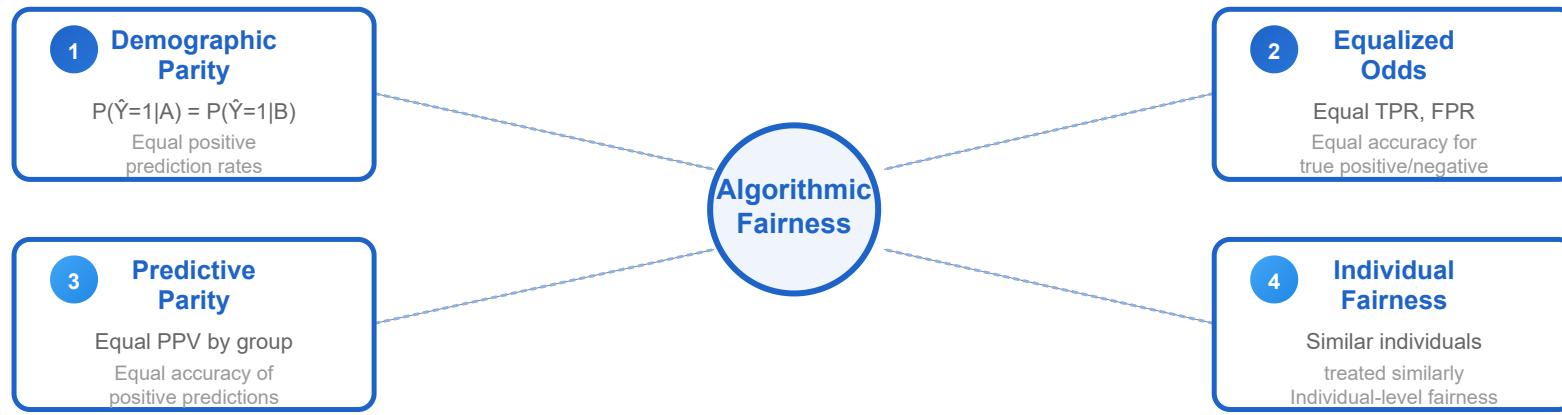
Similar individuals treated similarly Individual-level fairness

Algorithmic Fairness

⚠ **Impossible to satisfy all fairness criteria simultaneously**

⚠️ Trade-offs

- Fairness criteria cannot be satisfied simultaneously
- Accuracy vs fairness trade-offs
- Group fairness vs individual fairness
- Short-term vs long-term fairness

🛠️ Implementation Methods

- Pre-processing: Data resampling
- In-processing: Fairness-constrained learning
- Post-processing: Result adjustment
- Hybrid approaches