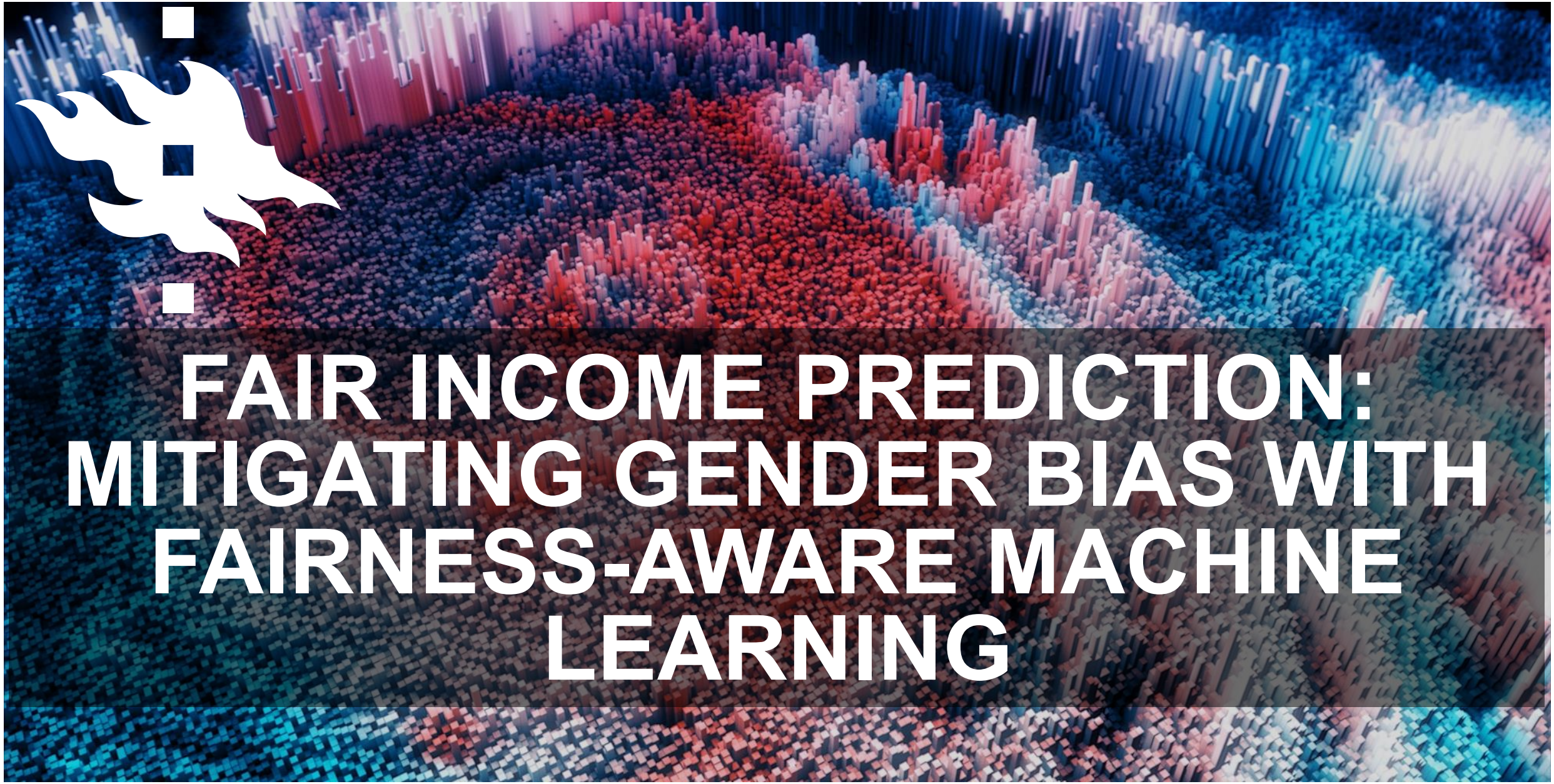


FAIR INCOME PREDICTION: MITIGATING GENDER BIAS WITH FAIRNESS-AWARE MACHINE LEARNING

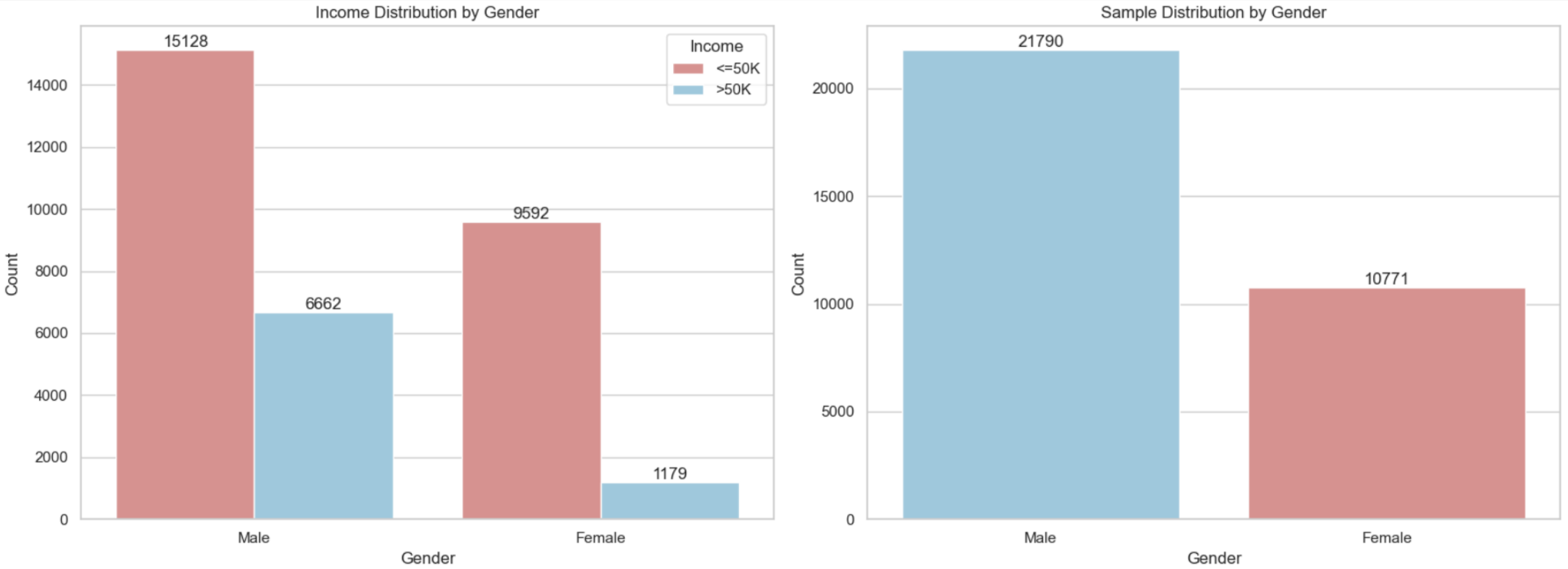
WHY ADDRESS GENDER BIAS IN INCOME PREDICTION?

- AI affects hiring, promotions, and salaries (e.g., Amazon recruitment tools).
- Historical data shows men often earn more than women (Adult dataset).
- Models can amplify existing inequalities if trained on biased data.
- Ethical & legal risks: biased decisions may violate anti-discrimination laws.
- Fairness-aware models promote equitable decisions and build trust in AI.

ADULT / CENSUS INCOME DATASET

- Task: Predict whether an individual earns > $50K per year
- Number of instances: 48,842 records
- Number of features: 14 attributes (numerical and categorical)
- Target variable: income (binary: ≤50K or >50K)
- Sensitive attribute: Gender (Male / Female)
- Key features:
  - Age, Workclass, Education level,
  - Marital status, Occupation, Relationship,
  - Race, Hours per week, Native country
- Mostly clean data but contains some missing values (represented as ?)

DISTRIBUTION OF INCOME AND GENDER IN THE DATASET



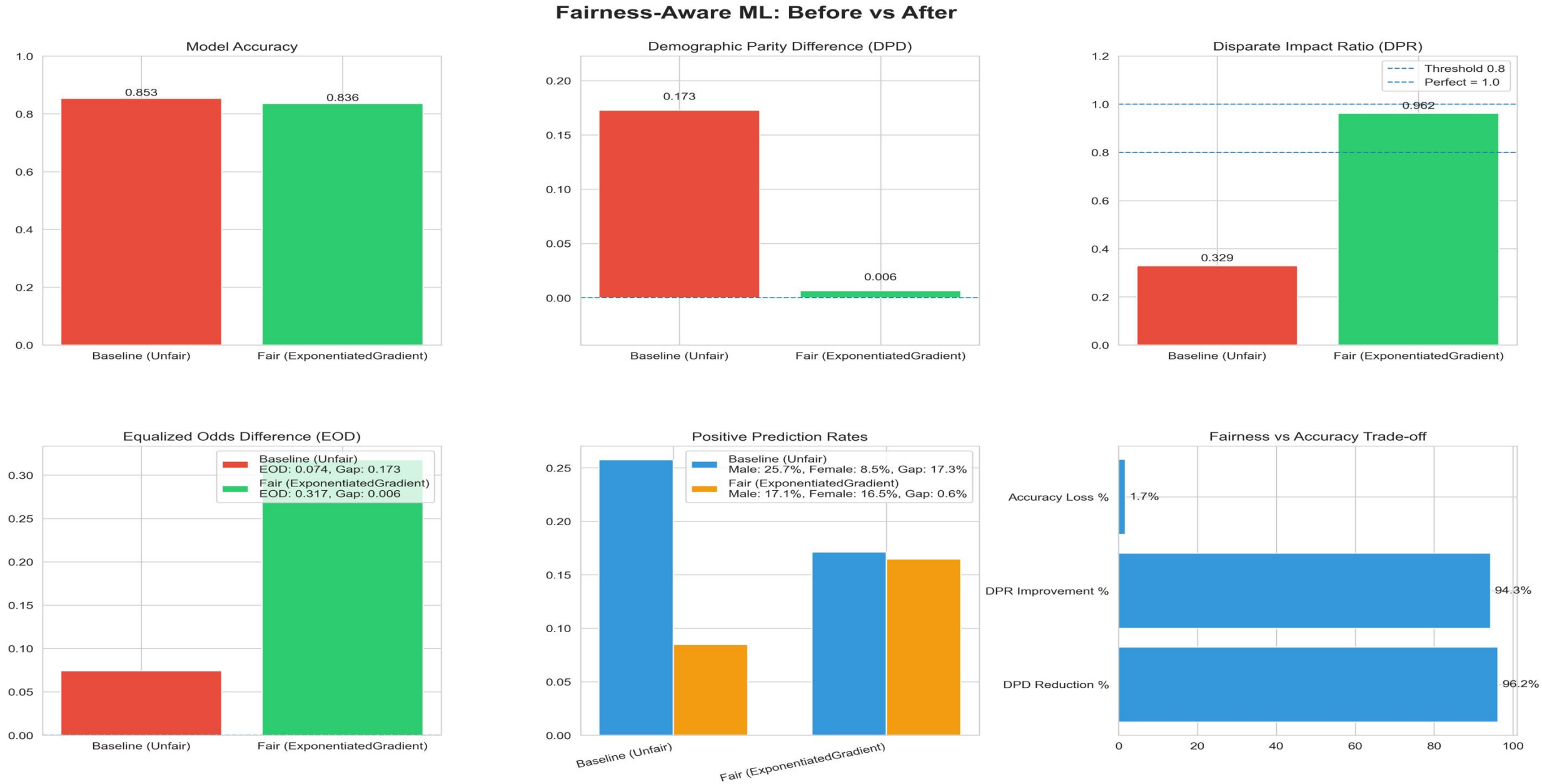
OUR APPROACH

- Train baseline models (Logistic Regression and Random Forest) without constraints.
- Apply Exponentiated Gradient (EG) to enforce Demographic Parity (DP) during training.
- Demographic Parity (DP): aim for similar positive prediction rates for men and women.
- Evaluate models on both accuracy and fairness metrics (DPD, DPR, EOD, EOR).
- Compare baseline vs fairness-constrained models to understand trade-offs.

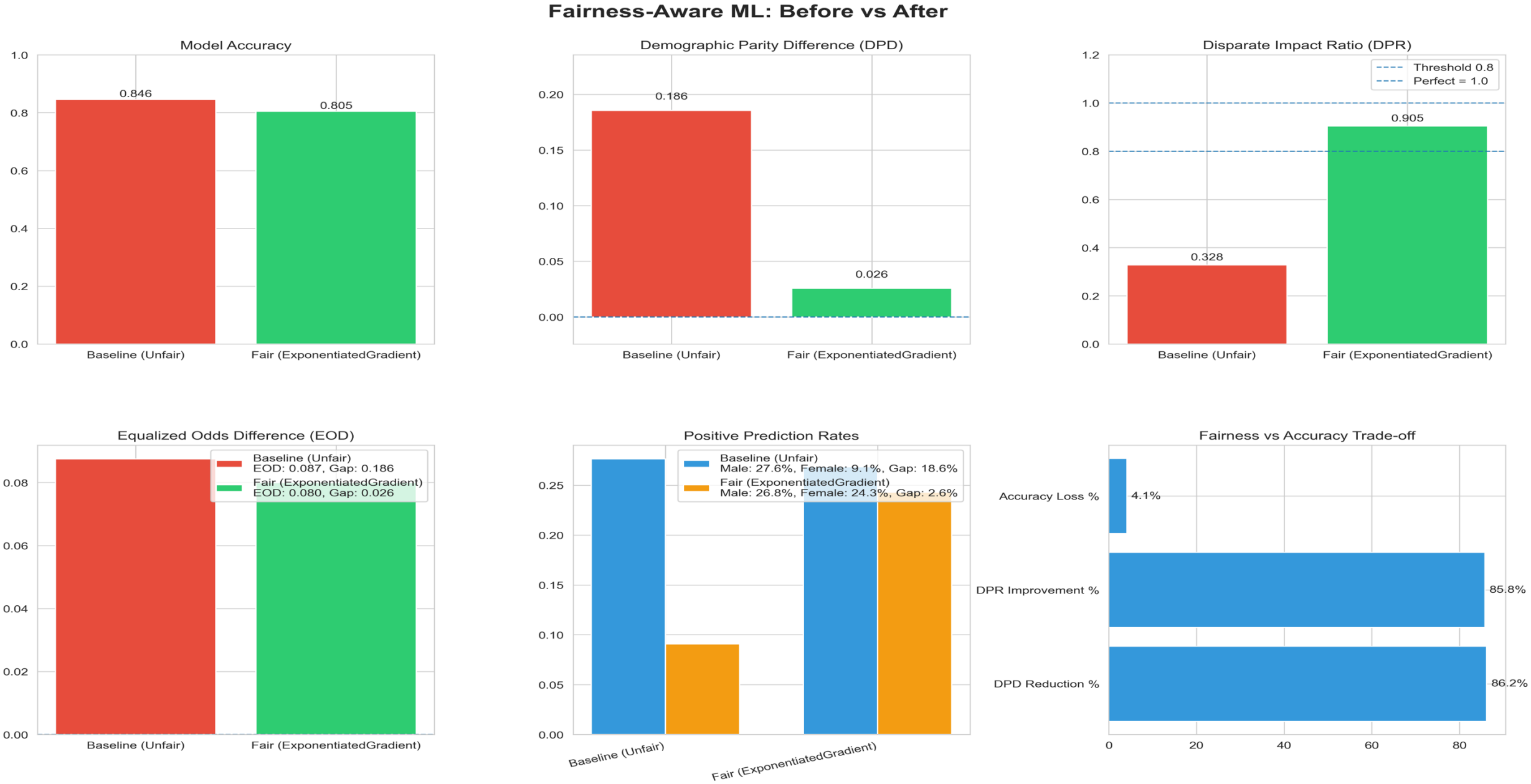
EXPONENTIATED GRADIENT

- Optimization method used to train models **while enforcing fairness**
- Used with **Demographic Parity** constraints in our project
- Learns several models and combines them as a weighted ensemble
- Adjusts model weights to reduce bias at each iteration
- Final outcome = model with **balanced accuracy & fairness**

LOGISTIC REGRESSION



RANDOM FOREST



CONCLUSION

- Baseline models (LR & RF) showed **clear gender disparities** in predictions
- Applying **Exponentiated Gradient with Demographic Parity** reduced bias significantly
- Fair models achieved **better balance between male & female prediction rates**
- Some reduction in accuracy occurred — but fairness improved meaningfully
- Demonstrates that **mitigating bias is achievable without destroying performance**
- Fairness-aware methods make ML systems **more ethical, trustworthy and deployable**

THANK YOU

- Any questions?