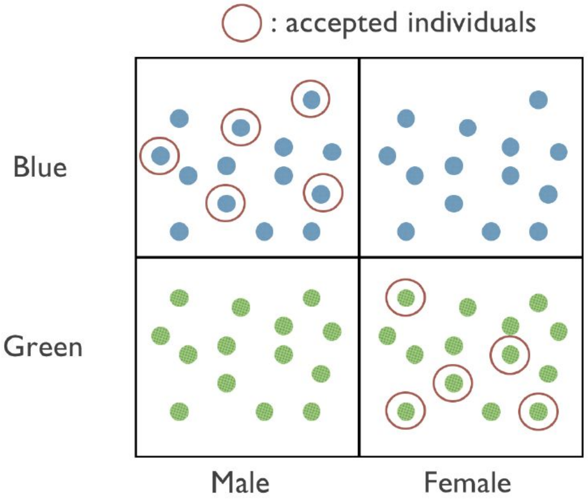
At Risk Student Prediction

Rui Chen, James Fantin, Kun Yi

Fairness

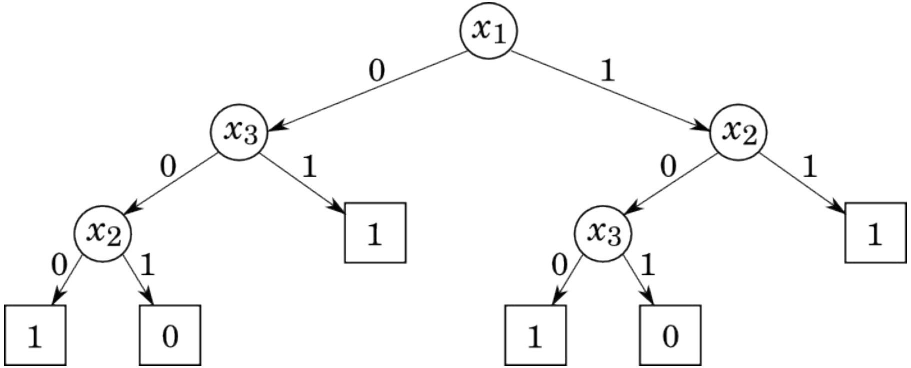
Fairness between groups means that each group is treated equally

Intersectional bias means that the model does discriminate based on several attributes of a student



Distributed Fair Random Forest

- Developed a fair random forest algorithm based on randomly generated decision trees



Distributed Fair Random Forest

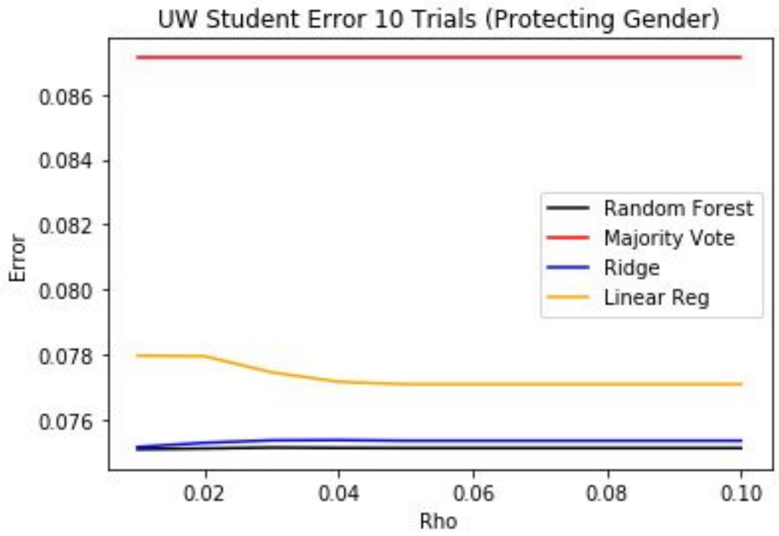
Our Distributed Fair Random Forest Algorithm

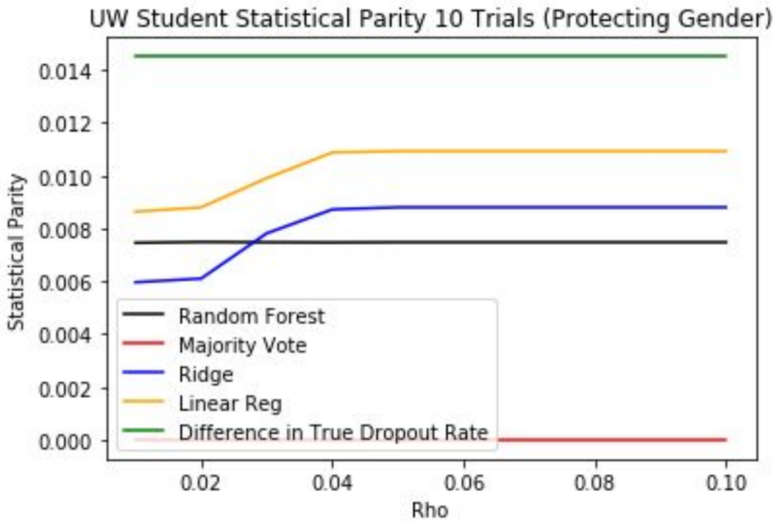
- Assume a third party hold private demographic data
- A data center holds remaining data and builds a model
    - Build completely random decision trees
    - Third party determines which ones are fair
    - The fair trees are weighted based on accuracy to make decisions

University Data Set

- Total dropout rate after cleaned data set is around 8.8%.
- Female dropout rate is slightly lower than male dropout rate.
- 20,175 student records
- Included ACT, GPA, Major, Gender, Race, etc.
- All records were anonymized in our data set.

Current Results





Current Results On UW Data

DFRF is our Distributed Fair Random Forest Model

FRF, FDT, FAHT are competing fair algorithms

Model	Error Rate	Statistical Parity
DFRF	0.075172969	**0.005964156**
FRF	**0.074395516**	0.046874894
FDT	0.119231385	0.075656425
FAHT	0.084257619	0.026245917

Other Results

Table 1: Performance on the Community Crime Data Set

Method	Statistical Parity	Classification Error
FAHT	0.0943 ± 0.4405	0.1482 ± 0.0117
FRF	0.0991 ± 0.3755	0.1263 ± 0.0741
FDT	0.1730 ± 0.1914	0.1395 ± 0.0392
DFRF	0.0153 ± 0.0430	0.1413 ± 0.0770

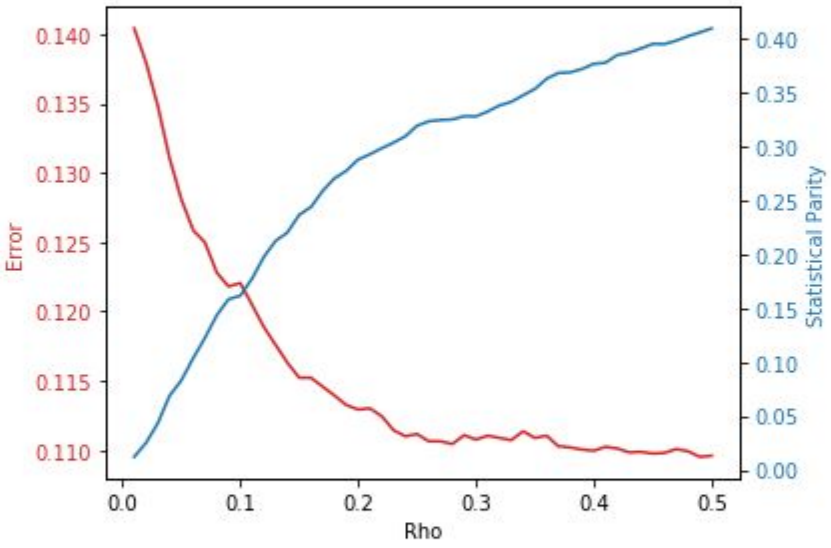
Table 2: Performance on the COMPAS Data Set

Method	Statistical Parity	Classification Error
FAHT	0.0008 ± 0.0037	0.2296 ± 0.0031
FRF	0.0000 ± 0.0000	0.2281 ± 0.0176
FDT	0.0019 ± 0.0019	0.2302 ± 0.0164
DFRF	0.0037 ± 0.0023	0.2296 ± 0.0104

Table 3: Performance on the Credit Card Data Set

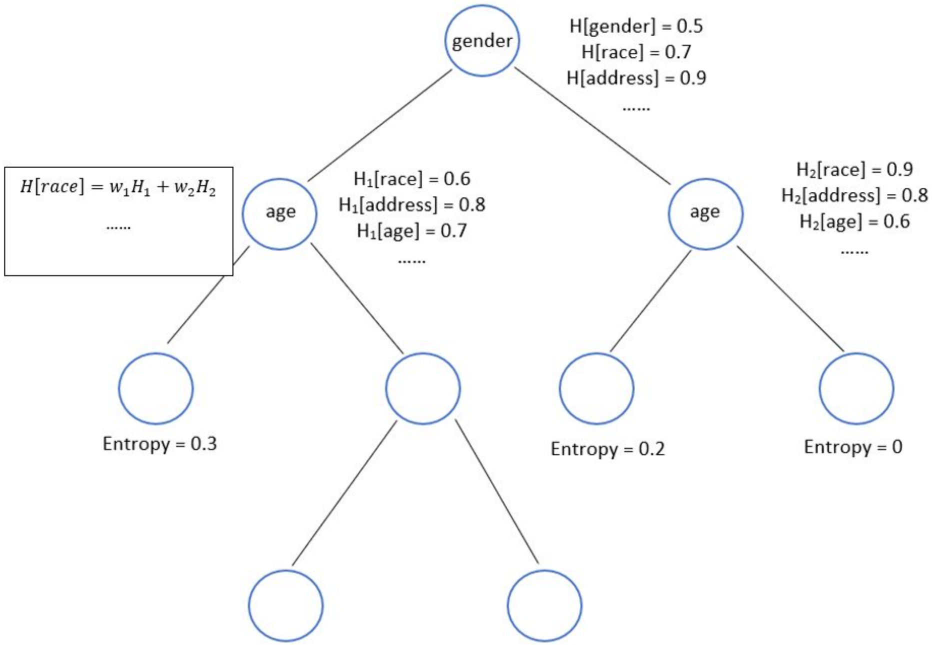
Method	Statistical Parity	Classification Error
FAHT	0.0020 ± 0.0050	0.2309 ± 0.0013
FRF	0.0626 ± 0.0436	0.1882 ± 0.0124
FDT	0.0461 ± 0.0289	0.1924 ± 0.0138
DFRF	0.0123 ± 0.0540	0.2055 ± 0.0183

Fairness vs Error (Community Crime Set)



Decision Tree Based Detector

- Build intersectional bias detector based on standard decision tree
- Split each parent node by the feature with best purity (lowest entropy or impurity)
  - Compute level entropy by weighted addition
- Stop split until max depth or reach entropy threshold
- Output is a set of features



Detector Result Hypothesis

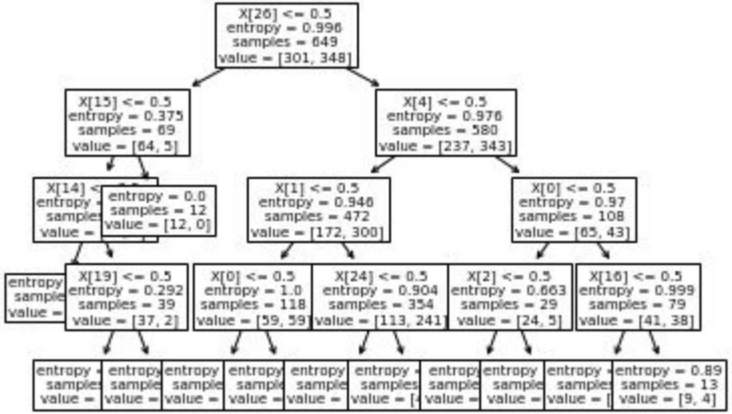
The top several features in Decision Tree Detector is more likely to show intersectional bias than bottom several features

$$
\begin{aligned}
&D(f_i, f_j) \geq D(f_a, f_b) \\
&if \\
&i, j \leq a, b
\end{aligned}
$$

where *f* is feature and *D* is disparity between two features

Standard decision tree VS. decision tree detector

- Standard decision tree
  - Use weighted entropy for splitting criteria
  - Different feature in one level
  - Features can be reused
  - Result:
- Decision tree detector
  - Weighted entropy for splitting criteria
  - Same feature in one level
  - Features cannot be reused
  - Result top 10 features: [26, 4, 0, 24, 1, 9, 14, 22, 2, 19]



Current Result

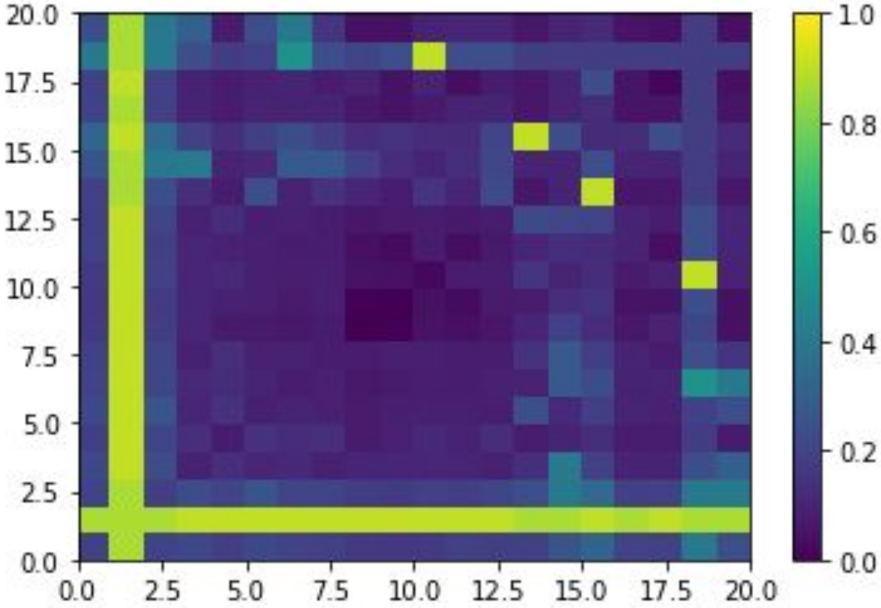
University of Wyoming Datasets

- 1572 students for one term
- 9.7% dropout rate

Testing

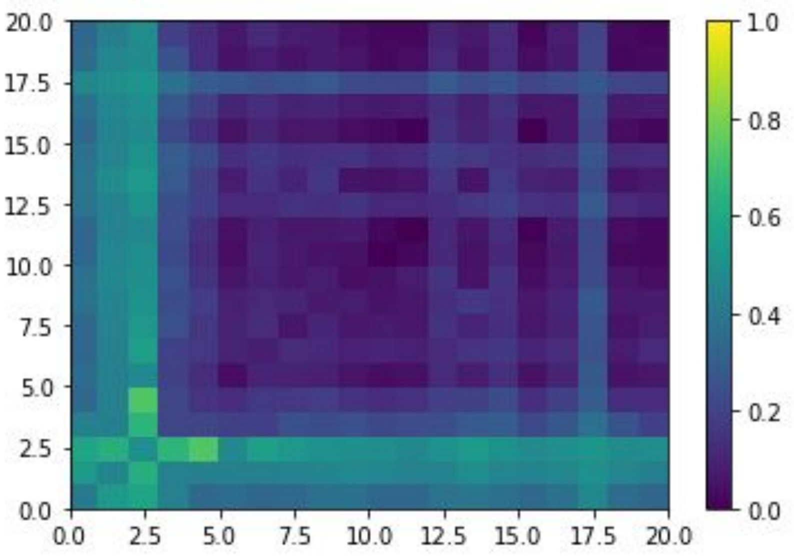
- Test first 20 features
- Entropy threshold = 0
- Light color: high disparity
- Dark color: low disparity

Colormap with first 20 features



Current Result

- Crime Community
    - 1993 instances
    - 40% high crime rate
- Testing
    - Test first 20 features
    - Entropy threshold = 0



Current Result

Colormap result from Standard Decision Tree:

Feature order from top to down, left to right