TRANSPARENCY AND FAIRNESS IN AI AND BIG DATA ALGORITHMS

Ricci Use Case

Yohan ISMAËL

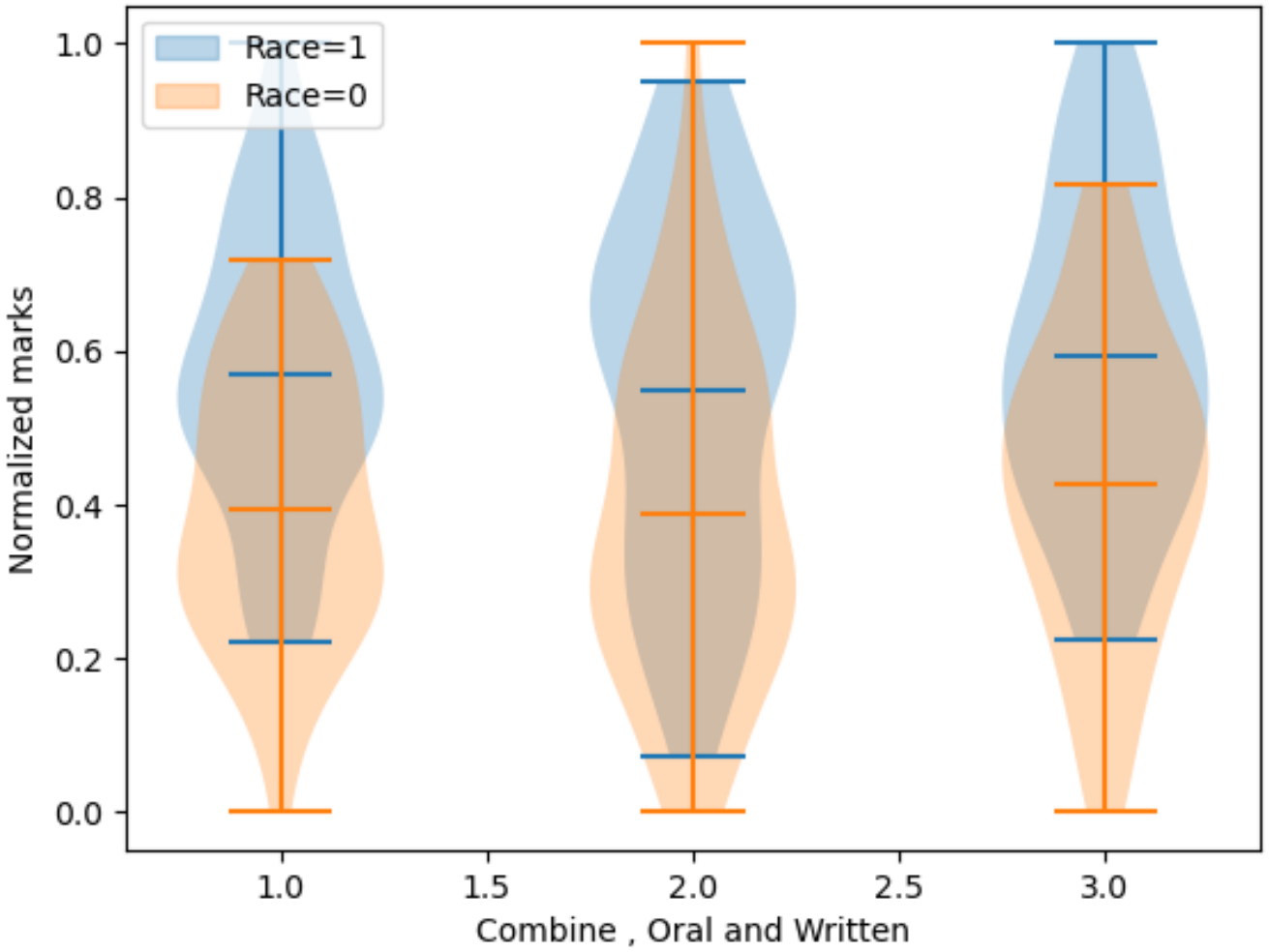
ENSEA
Ecole Nationale Supérieure de l'Electronique et de ses Applications

CERGY PARIS UNIVERSITÉ

The Ricci Dataset

- Exam to firefighters to have a promotion (Captain or Lieutenant)
- Features :
  - *Oral, Written and Combine score (Combine score is 60% Written and 40% Oral)*
  - *Race (White, Hispanic or Black)*
  - *Position for the promotion (Captain or Lieutenant)*

The Ricci Dataset



Race is the sensitive attribute

68 white participant (Race=1) with 41 promotion

50 black and hispanic participant (Race=0) with 15 promotion

Main Fairness Metrics

- Mean difference between privileged and unpriviliged group shows the fraction of privileged people which have more positive outocomes, We want him around 0 . For the original test set : 0.1
- Disparity impact shows the rate of positive outcomes over the unprivileged group divide by the rate of positive outcomes over the privileged group. We want it around 1. For the original test set : 0.78
- Average odd difference is the average of difference in false positive rates and true positive rates between unprivileged and privileged groups. We want it around 0. For the original test set : -0.169444
- Balanced accuracy is the mean between true positive rate and true negative rate. For the original test set : 0.872340

Reweight + Classifier

- Reweighing transforms the dataset to have more equity between the privileged and unpivileged groups
- Pre processing technique : we need a classifier after reweighting the dataset
- With Logistic regression:
  - *Mean difference* = 0.025
  - *Disparate impact =1.06*
- With Random Forest:
  - *Mean difference =1e-16*
  - *Disparate impact =1*

In processing Techniques

- In processing techniques are used to have a fair classifier without other processing techniques, we will use 2 of them.
- Prejudice Remover : Adds a discrimination-aware regularization term to the learning objective. Less accurate but fair classifier.
- Grid Search Reduction : Returning the deterministic classifier with the lowest empirical error subject to fair classification constraints among the candidates searched.

Results

	Original (Random Forest)	Reweight + Logistic Regression	Reweight + Random Forest	Grid Search Reduction	Meta Fair Classifier
Balanced accuracy	0.884	0.982	1.000	1.000	0.825
Mean difference	-0.261	-0.030	0.000	-0.326	-0.232
Disparate impact	0.490	0.945	1.000	0.505	0.706
Average odds difference	-0.012	-0.031	0.000	0.000	-0.021
Equal opportunity difference	-0.024	0.000	0.000	0.000	0.000
Theil index	0.128	0.005	0.000	0.000	0.044

Conclusion

- The best classfier is reweight + random forest. The problem is the lack of transparency in this method.
- Meta fair classifier is the best one in in processing techniques but the balanced accuracy decrease
- Grid search reduction maximise accuracy and minimize average odds difference but not the others metrics.