<u>Programming Assignment 3 Report</u>
Group 8

Suraj Ashikar
Shreya Pampattiwar
Aditi Ghatge

Model Choice: SVM
Algorithm Choice: Equal Opportunity
Secondary Optimization Criteria: Cost

<u>With total Data:</u>

--------------------EQUAL OPPORTUNITY RESULTS--------------------

Accuracy for Hispanic: 0.5968028419182948
Accuracy for Caucasian: 0.6268840356813288
Accuracy for African-American: 0.6347748815165877
Accuracy for Other: 0.6005917159763313

FPR for Hispanic: 0.513595166163142
FPR for Caucasian: 0.5032297754537065
FPR for African-American: 0.512542372881356
FPR for Other: 0.4975609756097561

FNR for Hispanic: 0.24568965517241378
FNR for Caucasian: 0.2430021531836358
FNR for African-American: 0.25092056812204105
FNR for Other: 0.24812030075187969

TPR for Hispanic: 0.7543103448275862
TPR for Caucasian: 0.7569978468163642
TPR for African-American: 0.749079431877959
TPR for Other: 0.7518796992481203

TNR for Hispanic: 0.48640483383685795
TNR for Caucasian: 0.4967702245462935
TNR for African-American: 0.48745762711864404
TNR for Other: 0.5024390243902439

Threshold for Hispanic: 0.05
Threshold for Caucasian: 0.08
Threshold for African-American: 0.11
Threshold for Other: 0.03

Total cost: $-757,737,810
Total accuracy: 0.6269598292977085

With market model:
Accuracy on training data: 0.6184790169255738
Accuracy on test data: 0.6358569437993498
Cost on training data: $-613,922,792
Cost on test data: $-145,688,224

Total cost by adding and testing data: -$759,611,016

TPR for Caucasian: 0.8571972581873573
TPR for Hispanic: 0.855
TPR for Other: 0.8598130841121495
TPR for African-American: 0.8638778220451527

FPR for Caucasian: 0.6294003868471953
FPR for Hispanic: 0.6431372549019608
FPR for Other: 0.4838709677419355
FPR for African-American: 0.6719798657718121

FNR for Caucasian: 0.1428027418126428
FNR for Hispanic: 0.145
FNR for Other: 0.14018691588785046
FNR for African-American: 0.13612217795484727

TNR for Caucasian: 0.37059961315280465
TNR for Hispanic: 0.35686274509803917
TNR for Other: 0.5161290322580645
TNR for African-American: 0.3280201342281879

Threshold for Caucasian: 0.08
Threshold for Hispanic: 0.06
Threshold for Other: 0.09
Threshold for African-American: 0.09

The COMPAS system was being used by many **Justice Departments across the country** to deny or accept bails. If the risk predicted by the system was high then the bail would be denied and if the risk was low then bail would be granted. Critically, race was not one of the factors used by COMPAS. **That meant, this sensitive attribute, was not being considered while making decisions, which may have made the model unfair.**
However, when the algorithm was incorrect it tended to skew very differently for each of these groups - White defendants and Black Defendants. **The reason behind this was disparity in the training data which caused the Machine Learning Model to become unfair. Even after not using the sensitive attribute, some other attribute related to the sensitive attribute started behaving as a proxy, causing the unfairness.**
In real-world problems, sensitive attributes will often be a feature that carries legal ramifications for decisions based around it – for example, race, age, gender, or sexual orientation.
There is no bias present in the algorithm, however due to the disparity present in the input data, our machine learning model/algorithm may propagate this unfairness further.

The COMPAS system was making decisions that impacted individuals and thus it was expected from the machine Learning model to not propagate this disparity further.
Hence the motivation of creating a new model to replace COMPAS which will be fair as well as bias free.

Our Proposed Model:

Algorithm Choice: *Equal Opportunity*
Secondary Optimization Criteria: *Cost*

In our model, along with the secondary optimization metric we are also considering another criteria for fairness selection - TPR.

The fairness criteria - equal True Positive Rates can lead to all possible outcomes -improvement, stagnation, and decline in natural parameter regimes. There are a class of settings where equal selection rates cause decline, whereas equal true positive rates do not.
Impact of our model is that, along with minimized cost, including TPR as fairness selection criteria would ensure improvement and stagnation.

Comparison with other alternatives:

Equal opportunity vs Demographic parity:
TPRs and FNRs are the same across all groups but if we apply secondary optimization on the metric COST, the Equal opportunity algorithm clearly stands out.

Equal opportunity vs Single threshold:
TPRs and FNRs are not the same across all the groups in case of Single threshold and hence that model cannot be considered as a fair model in comparison to Equal Opportunity.

Equal opportunity vs Predictive Parity:
As compared to Equal Opportunity it can be observed that TPRs and FNRs are not the same across all groups.
For example FPR for African-American is 0.8149 while for other groups it is, 0.5392(Caucasian) 0.2054(Hispanic), 0.0878(Other).
Means a high proportion of African-American are incorrectly rated at high risk which is not fai

Equal opportunity vs Maximum profit:
In case of Maximum profit algorithm according to the secondary optimization metric, COST is the maximally optimized in this method, but again if we observe other fairness metrics like TPR - for African-American, it is 0.85 and that for Other is 0.33. This means that in the case of African-Americans, the percentage of defendants who would recidivate are more correctly identified rather than in Other ethnicity. When TPR is high it means FNR is low which in turn means for those entries whose true label is false, the model still identifies them as true.
Likewise, If we observe for all groups, FNR value is least for African American ethnicity, which is unfair. Thus, cost wise maximum profit is the best method but if we look at other metrics such as TPR and FPR we can see substantial difference across the groups.

*Hence the best solution for fairness is, **Equal Opportunity** algorithm with **Cost as a metric** as mentioned before.*

References:
https://www.upenn.edu/learninganalytics/ryanbaker/LAK_PAPER97_CAMERA.pdf

https://arxiv.org/pdf/1803.04383.pdf
https://www.propublica.org/article/bias-in-criminal-risk-scores-is-mathematically-inevitable-researchers-say