Programming Assignment-3 Report

Team Members:

1. Sandesh Kumar Srivastava
2. Charan Reddy Bodennagari
3. Venkata Narayana Rohit Kintali

Team Number: 35

Course Number: CSE 574

Model: Support Vector Machine
Postprocessing algorithm: Equal Opportunity
Secondary optimization criteria: Accuracy
Accuracy on training data:0.6394620913517274
Accuracy on test data:0.6516488620529494
Cost on training data: $-594,192,398
Cost on test data: $-142,424,158

We as humans are biased naturally and it inherently reflects in the way we observe data. Since machine learning is based on this perceived data, it is quite possible that some biases creep in our machine learning model. If we use this biased model for prediction in a pivotal system such as the US criminal justice system, the effects could be catastrophic. Moreover, the results of prediction often effect the environment which is the society in this case and it results in an endless loop where the bias keeps on increasing. Hence, the need of devising a model which is fair to all sections of society. As volunteers of a humanitarian NGO (non-governmental organization), our motivation is to identify such biases and create a fair model by minimizing the impact of biases.

Any disparities in the model would affect the stakeholders involved. The stakeholders in this situation would be of different kinds namely the criminals whose lives will depend on the decision made by the algorithm, the justice department who is responsible for maintaining the crime rate low, the government who has to maintain the prisons and protecting the public and the general public who might be impacted by the actions of the criminal.

There might be different sources from which biases might occur, in this situation, racial bias might exist in the data given as it has been collected from the real world. So, the biases in the real world will be carried into the data. Even in the algorithm biases exist considering biases exists in data. So, taking COMPAS algorithm into consideration, this algorithm reflects racial biases and leads to some issues which are unfair and we would be proposing a model which balances the biases occurred in the model.

As volunteers of a humanitarian NGO, our primary responsibility lies in creating a model which is fair to all sections of society. By selecting equal opportunity as measure of fairness for our model and keeping accuracy as the secondary optimization we do not allow cost to be the guiding factor for US criminal justice system.

For our model we have selected “equal opportunity” as the measure of fairness because:

- It tends to keep the TPR/FNR values for all races as consistent as possible.
- Since TPR reflects the fraction of people who are correctly predicted to recidivate, if we get this ratio uniform across all races we can conclude that our model does not discriminate individuals when making a correct prediction.

We have selected “accuracy” as the secondary optimization because:

- Accuracy measures the correct predictions we make with our model and so by maximizing accuracy we try to make predictions as close to actual values as possible.
- Accuracy is a more measurable parameter compared to cost as we cannot justify what is the actual cost of a correct or a wrong prediction
- Pure economic figures may not accurately account for suffering by the defendant or their families.

The proposed model shows disparity in PPV, FPR/TNR across all the racial lines whereas the TPR/ FNR remains approximately constant across all the racial groups. In order to create a system where all the individuals will get equal opportunity for a fair trial irrespective of biases we need to give more importance to TPR compared to other metrics.

Therefore, we demonstrate that our model will be preferable with “Equal Opportunity” as the apt fairness measure.

```
Accuracy for African-American: 0.6666666666666666
Accuracy for Caucasian: 0.6436870642912471
Accuracy for Hispanic: 0.6203703703703703
Accuracy for Other: 0.6973684210526315

FNR for African-American: 0.25569620253164554
FNR for Caucasian: 0.2736
FNR for Hispanic: 0.25
FNR for Other: 0.2692307692307692

TPR for African-American: 0.7443037974683544
TPR for Caucasian: 0.7263999999999999
TPR for Hispanic: 0.75
TPR for Other: 0.7307692307692308
```

References:

[1] NINAREH MEHRABI, FRED MORSTATTER, NRIPSUTA SAXENA, KRISTINA LERMAN, and ARAM GALSTYAN. A Survey on Bias and Fairness in Machine Learning, https://arxiv.org/pdf/1908.09635.pdf

[2] Fairness in Machine Learning, http://sitn.hms.harvard.edu/uncategorized/2020/fairness-machine-learning/