**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**: SVM

**Postprocessing algorithm:** Equal Opportunity

**Secondary optimization criteria:** Accuracy

**Total Cost:** $-738,582,624

**Overall Accuracy:** 0.6398552741441692

1)

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.

The data collected from the real world will have inherent biases and the typical predictive algorithms generally ignores some hidden biases which may give unfair results towards class of data. 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. In the given scenario we trying to address two of the following key issues:

* The white defendants were mislabeled as low risk more often than black defendants.
* The formula was particularly likely to falsely flag black defendants as future criminals, wrongly labeling them this way at almost twice the rate as white defendants.

2)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.

3)In the given situation, racial bias, demographic bias might exist in the data given as it has been collected from the real world. So, the biases 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.

4)Our model tries to reduce the racial bias present in COMPAS by providing equal opportunity to all races. We achieve TPR ratio of all races within a permissible variation of 0.01. By providing equal opportunity, we are trying to achieve the ultimate goal of fair treatment for individuals within the U.S. criminal justice system.

5)As volunteers of a humanitarian NGO, our primary responsibility lies in creating a model which 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:

1. It tends to keep the TPR/FNR values for all races as consistent as possible.
2. 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:

1. 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.
2. Accuracy is a more measurable parameter compared to cost as we cannot justify what is the correct cost of a correct or a wrong prediction.

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 which would be the same for all races considering equal opportunity enforcement and it can be considered that our model gives equal opportunity to all the correct people who has not committed the crime to prove themselves that the right justice has been served for the positive people. Similarly, our FNR values came out to be constant for all races saying that the wrong people are proven to wrong and committed a mistake and justice has been served for them too.

|  |
| --- |
| Accuracy on training data: 0.6383028054718294  Accuracy on test data: 0.6483975847654435  Cost on training data: $-595,051,358  Cost on test data: $-143,238,136  -----------------------------------------------------------------  Attempting to enforce equal opportunity...TEST\_DATA  PPV for African-American: 0.6997635933806147  PPV for Caucasian: 0.60625814863103  PPV for Hispanic: 0.39344262295081966  PPV for Other: 0.5277777777777778  FPR for African-American: 0.44876325088339225  FPR for Caucasian: 0.45345345345345345  FPR for Hispanic: 0.4868421052631579  FPR for Other: 0.34  FNR for African-American: 0.25063291139240507  FNR for Caucasian: 0.256  FNR for Hispanic: 0.25  FNR for Other: 0.2692307692307692  TPR for African-American: 0.7493670886075949  TPR for Caucasian: 0.744  TPR for Hispanic: 0.75  TPR for Other: 0.7307692307692308  TNR for African-American: 0.5512367491166077  TNR for Caucasian: 0.5465465465465466  TNR for Hispanic: 0.513157894736842  TNR for Other: 0.6599999999999999  -----------------------------------------------------------------  Attempting to enforce equal opportunity...TRAIN DATA  PPV for African-American: 0.6482213438735178  PPV for Caucasian: 0.6149125498619208  PPV for Hispanic: 0.5547445255474452  PPV for Other: 0.5503355704697986  FPR for African-American: 0.5226510067114094  FPR for Caucasian: 0.48549323017408125  FPR for Hispanic: 0.47843137254901963  FPR for Other: 0.432258064516129  FNR for African-American: 0.23771580345285526  FNR for Caucasian: 0.23686214775323686  FNR for Hispanic: 0.24  FNR for Other: 0.2336448598130841  TPR for African-American: 0.7622841965471447  TPR for Caucasian: 0.7631378522467631  TPR for Hispanic: 0.76  TPR for Other: 0.7663551401869159  TNR for African-American: 0.4773489932885906  TNR for Caucasian: 0.5145067698259187  TNR for Hispanic: 0.5215686274509803  TNR for Other: 0.567741935483871  (For Both TRAIN\_DATA & TEST\_DATA)  Threshold for Other: 0.0967741935483871  Threshold for African-American: 0.12013348164627363  Threshold for Caucasian: 0.10789766407119021  Threshold for Hispanic: 0.09010011123470522  ----------------------------------------------------------------- |