CSE 574 INTRODUCTION TO MACHINE LEARNING

Group Number : Programming Assignment 30

Group Members: Venugopal Chinthappareddy Govindarajan(50336845)
Shathesh Kumar(50336820)
Hari Guhan Thillairajan(50338189)

COMPAS++

Model choice: Kernel SVM
Post Processing technique: Equal opportunity
Secondary optimization metric: Accuracy
Accuracy on test data: 71.01%
Cost on Test data: $-125,545,504
Accuracy on train data: 85.7%
Cost on training data: $-339,444,050

1. The main motivation for creating a new model (**COMPAS++**) to replace **COMPAS** is to avoid biases against certain groups or races. On analysis of COMPAS tool, Propublica found that black defendants were far more likely than white defendants to be incorrectly judged to be at a higher risk of recidivism, while white defendants were more likely than black defendants to be incorrectly flagged as low risk. Even, Black defendants were also twice as likely as white defendants to be misclassified as being a higher risk of violent recidivism by COMPAS (**propublica analysis**). NorthPointe's **defense** is that Propublica did not take into account the different base rates of recidivism for Blacks and Whites which resulted in false assertion. So, this motivates us to create a new model that satisfies the **fairness definition** and also maximizes the expected **public safety(reference[Section 1])** and at the same time considering all the metrics of the individuals like sex, age, charges on him and priors count.

2. As our company takes the task of designing the new model **(COMPAS++)** to replace COMPAS. The key stakeholders are **Shareholders**, **Investors, Directors**, **Employees**, **Government** and its agencies. As the model scores are used to predict whether the individual is likely to be recidivated or not by the judges so the other key stakeholders are **Judges**, **Defendants**, **Attorneys, Police** and other **law enforcement agencies.** For every false negative prediction by the model, the potential impact is a crime with the general public or tax-payers being the victims. This includes the **general public or tax-payers** also as the key stakeholders.

3. This data is released by **ProPublica**.

	WHITE	AFRICAN AMERICAN
Labeled Higher Risk, But Didn't Re-Offend	23.5%	44.9%
Labeled Lower Risk, Yet Did Re-Offend	47.7%	28.0%

Yes, there exist biases in the Northpointe's assessment tool as per Propublica which stated that COMPAS predicts recidivism 61 percent. But African Americans are almost twice as likely as Caucasians to be labeled a higher risk but not actually re-offend. It makes the opposite mistake with whites. They are much more likely than blacks to be labeled lower risk but go on to commit crimes. This motivates to model a new system.

Yes, there is a bias present in the algorithm, an African American(refer to id 480 in csv) who has no prior count and the current case is third-degree grand theft and was given 9 as the decile score by COMPAS. Whereas a Caucasian(refer to id 613 in csv) who has 4 prior counts and current case is similar like an African American but was given a decile score as 1 by COMPAS. Interpreting COMPAS decile scores as binary predictions, the True positive rates of Afican Americans is significantly higher than Caucasians. The trend seems to be the same in terms of ages, the TPR for age category below 25 is much higher than TPR for age category over 45. However, in terms of gender, there is not much difference(**Refer Gragh3**). From our perspective of fairness, the model is pretty **biased** along the lines of **age and race** and fair with respect to gender.

(**Refer Graph2**) This graph is analysed from the given COMPAS Broward data. This comparison is between the African-American and Caucasain groups, we found that the decile

score provided by COMPAS was slightly biased towards Caucasians as, for people with [Section 2.7 - COMPASS Guide] less than 5, but they did go on to recidivate, this number was high in the Caucasian who did go on to recidivate. So, COMPAS gave Caucasians who were actually recidivating with a lower decile score.

4. We believe that all groups should be treated equally, as in this case it refers to being labelled as recidivist. There should be absence of discrimination based on attributes such as racial or ethnic origin. So, a model should predict the correct rate of those who recidivate among all races. Our proposed solution takes **Kernel SVM** as the model and **Equal Opportunity** as the post processing technique and **Accuracy** as the secondary optimization. We were able to achieve best overall accuracy on train data of neary **85.7%** and the corresponding profit to be **$-339,444,050** maintaining the fairness among all demographic groups with certain epsilon tolerance. We applied the best threshold that we got from the training data predictions and applied the same threshold to the test data and we were able to achieve the best accuracy of **71%** and the corresponding profit to be **$-125,545,504**. Post applying these thresholds to test data epsilon has deviated to 0.07 which is not much significant for test data. Our risk assessment tool must predict the defendants in trial accurately, as many bureaucrats use the score provided to make decisions - so there should not be any chances of misclassification, especially when the defendant's future is at stake. So, in order to maintain the credibility of the company, and continue business with the government, we figured **Accuracy is more important than the cost**.

5. Fairness in the context of this problem statement, is correctly predicting the high risk cases of each demographic group with parity. That is, even if we are given a skewed dataset in which 80% of cases are labeled positive from one group and only 40% of cases labeled positive from another group, we believe it is only fair to correctly predict the same proportion of positive cases from either demographic group. To enforce such a kind of fairness, Equal opportunity works the best for us, which is the reason we are going with it *(Refer Graph1).* To enforce our perspective of fairness, base rates of recidivism among different groups is the pivotal factor.

   Both demographic parity and positive predictive parity both attempt to impose fairness by manipulating the false positive rates, which opens up the possibility of classifying low risk defendants as high risk which is really unfair to them. Maximum profit is a bit better as it attempts to maximize the accuracy of prediction for each group. However, this fails when our model is very good for one group and poor for another group.

   Our model is much better than the above models as in Equal opportunity, we attempt to equalize the true positive rates (refer to below graph), and since our secondary optimization maximizes the accuracy so false positive rates are minimized which mitigates the risk of low risk defendants being labeled high risk. Our model shows a significant disparity along the lines of positive predictive value and probability of positive prediction. However, as I said above these metrics manipulate false positives, which we dont believe is fair on the low risk defendants.

   Although our model reduces uncertainty about an offender's probable future conduct because of the high accuracy on test data, every risk assessment model is subject to errors and should be considered as reference rather than basing decisions on the score [Ref].
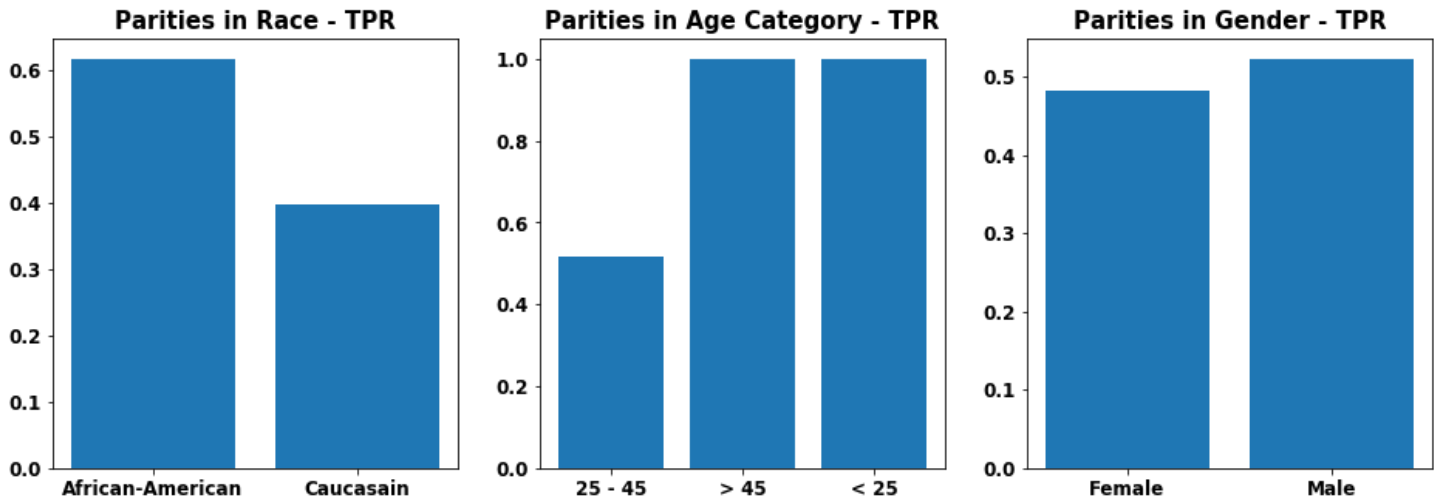
Graph1

TPR for all races - preserved

Graph2

Decile Score below 5 but went on to recidivate
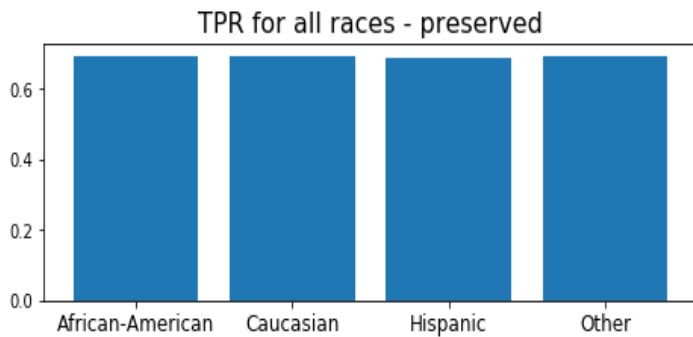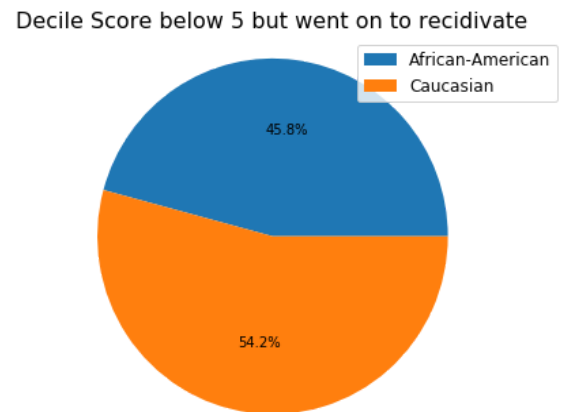
Graph 3 - COMPAS Parities across various metric

References :

1. Propublica Analysis on Compass Recidivism Algorithm.
   https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm
2. Northpointe's defense on Propubilica's accusation of COMPAS is racially biased against black.
   https://www.documentcloud.org/documents/2998391-ProPublica-Commentary-Final-070616.html
3. Paper on algorithmic decision making to maximize public safety while satisfying formal fairness constraint.
   https://arxiv.org/pdf/1701.08230.pdf
4. Northpointe practitioner guide to understand how decile score works.
   http://www.northpointeinc.com/files/technical_documents/FieldGuide2_081412.pdf
5. How do uncertainty and risk tolerance factor model's decision.
   https://fas.org/sgp/crs/misc/R44087.pdf