Programming Assignment #3

Role: Employee of Publicly Traded Corporation

Group 48

Introduction to Machine learning

Saumya Dholakia((#50320175), Dhruv Patel (#50321707)

5/9/2020

Table 1: Table listing out the parameters selected as a baseline for testing the model.

Parameters	Values
Model selected	Naïve Bayes Classifier
Algorithm selected	Demographic Parity
Secondary optimization criteria	Accuracy
Overall system cost	-754,714,674$
Overall system accuracy	63%

1. What is the motivation for creating a new model to replace COMPAS? What problem are you trying to address?

   The COMPAS algorithm predicted Black Defendants who did not recidivate for a period of two years to be at a higher risk of recidivism as compared to their white counterparts in a ratio 2:1. Similarly white defendants who actually reoffended within a span of two years were marked at a lower risk inside the same margins. When other parameters such as gender, age, prior crimes etc. in addition to race were taken into account the black defendants were again deemed 45% more risky as compared to whites. A similar trend was observed considering the 'violent recidivism' parameter too. This clearly indicates a marked trace of predictive parity inside COMPAS and raises questions about its ability to over predict for blacks and under predict for whites consistently under different situations [1]. Hence the main problem that needs urgent attention is that of predictive parity and to ensure a logical relationship between the magnitudes of True positives and True negatives respectively. This we think is possible by establishing a universality in post processing the data, supported by a well-defined objective function and minimal biases.

2. Who are the stakeholders in this situation?

   A stakeholder is either an individual, group or organization who is impacted by the outcome of a project [2]. In our case there are two parties at stake considering the outcome of the project. The first being the company or the corporation as a whole, which plays a direct role in implementing the selected model, not only claimed to replace COMPAS (a tried and tested model) but also to justify all algorithmic/ethical considerations related to the model. The life of the project implicitly governs the livelihood of employees dependent on it and also the finances, reputation, etc. of the company as a whole. The second party being affected are the defendants themselves whose lives are at stake along with their families and closely knit friends. The US department of justice and the crime departments also are at stake with respect to the costs associated with false positives and false negatives respectively.

3. What biases might exist in this situation? Are there biases present in the data? Are there biases present in the algorithms?

   As discussed earlier, the COMPAS algorithm does show a marked tendency to be partial with respect to a given race (specifically blacks) considering various sensitivity attributes such as race,gender,age etc.[1]. But the classification studies associated with the determination of an accurate risk scale considering the sensitivity and the specificity features as well as other tools such as AUC, ROC etc. clearly criticizes ProPublica's findings.[3]Irrespective of the views stated, biases are currently an unavoidable part of the family of predictive algorithms either injected historically or through underrepresentation of data.[4] In this case, bias was not a part of the data but might be indirectly associated with the Broward county's history in crime or due to several statistical and technical errors such as misspecified regression models, wrongly defined

classification terms etc.[1]Bias also may be fed in an implicit fashion by creating false proxies. A recent study explicitly classifies biases into nine different categories based on training data, algorithms and its allied processing, the transfer context bias, the interpretation bias,proxies,automation, consumer biases and the feedback loop bias.[4]

4. What is the impact of your proposed solution?

   The proposed solution uses the Naïve Bayes classifier as a benchmark when compared with the linear SVM considering the parameters such as a reduction in the computational time by one minute, the characteristic of being conditionally independent [5] which removed any associated biases in the corresponding data points being fed into the algorithm and a probabilistic approach which tends to save a lot of time involved in hyper parameter tuning and grid search related to the SVM's. Furthermore, the performance of linear SVM's chiefly depends upon the linear hyperplane and the selection of a kernel function always tending to produce a global optimum [6] which might not always be necessary.

   Apart from this, the best parameter coming out of the analysis is the 'Demographic parity result' which places more or less equal weightage on the false positive and negative rates associated with four different races. There is a high cost of -910,659,204$ associated with the 'Equal opportunity results' with a corresponding loss in accuracy, but the costs associated with the 'Demographic parity result' lie in the lower range of the spectrum at about -762,416,820$ and an accuracy of 62%.

   The most prominent impact coming out of the above discussion is therefore related to the probability of positive prediction for all the races which is more or less equal. This avoids marginalization amongst races, avoids unnecessary biases which were observed in COMPAS and most importantly gives the defenders a fair chance to be proven innocent or procure bail resulting in a fair trial.

5. Why do you believe that your proposed solution is a better choice than the alternatives? Are there any metrics (TPR, FPR, PPV, etc?) where your model shows significant disparity across racial lines? How do you justify this?

   The proposed solution as mentioned earlier is the use of the 'Demographic parity result' which provides approximately equal estimates of quantities such as the FPR and FNR for all the races. Other metrics such as the 'Predictive parity' provide a highly fluctuating range of FPR and FNR each having error bounds extending from 10 to 15%. The 'Maximum profit' metric also follows a similar vein with varying thresholds being set for all races as a result of the secondary optimization. Going further the 'Equal opportunity' metric obviously have fixed values of FPR and FNR, however, the accuracies differ for four different races, thus introducing inconsistencies. Lastly, the 'Single threshold' values also show varying rates of false positives and negatives thus eliminating it from being the best choice amongst all. Therefore the TPR, TNR, FPR and FNR are the most common metrics that show significant disparity amongst racial lines.

   References:

   1. *Machine Bias,* https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
   2. *Stakeholder (corporate),* https://en.wikipedia.org/wiki/Stakeholder_(corporate)
   3. *Northpointe's Response*, https://www.documentcloud.org/documents/2998391-ProPublica-Commentary-Final-070616.html

4. *Algorithms, Platforms, and Ethnic Bias: An Integrative Essay*, Selena Silva and Martin Kenney, In Phylon: The Clark Atlanta University Review of Race and Culture (Summer/Winter 2018) Vol. 55, No. 1 & 2: 9-37
5. *How do the naive Bayes classifier and the Support Vector Machine compare in their ability to forecast the Stock Exchange of Thailand?*, Napas Udomsak, Student, Bangkok Patana School
6. Huang, K. Z., Yang, H., King, I., & Lyu, M. R. (2008). *Machine learning: modeling data locally and globally,* Springer Science & Business Media.