Introduction to Machine Learning (Spring 2020)
Programming Assignment 3 – Group 41

Report

Vaibhav Chhajed (vchhajed)

Yeshwanth Badineni (ybadinen)

Hemant Koti (hemantko)

Market Submission Model	
Model	SVM
Algorithm	Equal Opportunity
Secondary optimization criteria	Accuracy
Overall system cost	$-752,119,222
Overall accuracy	0.6367742833286947

1. Machine learning algorithms often don’t account for the data with inherent deficiencies, biases which tend to produce models that possibly amplify the bias and result in unfair decisions. Our motivation as volunteers of a humanitarian NGO is to eliminate those inherent biases by creating a model that scrutinizes and eliminates these deficiencies. There are two major concerns in COMPAS that we try to address (also stated by independent news organization *ProPublica*):
    a. The algorithm was likely to falsely flag black defendants as future criminals, wrongly labeling them at almost twice the rate as white defendants.
    b. White defendants were mislabeled as low risk more often than black defendants.

2. We identified potential stakeholders who are diligent in proactively addressing factors that contribute to bias. Our primary stakeholders would be
    a. The defendants who will be impacted by the decision of the authority
    b. The U.S. criminal justice system that takes necessary actions using our model
    c. The government to manage the financials and the public
    d. Businesses that promote ethical AI practices for the necessary funding

3. Racial bias and systematic bias are very likely to exist in this situation. The data set might also contain biases in demographic distribution (Broward County, Florida in this case) which could likely be induced in COMPAS. Also, as stated earlier, machine learning models tend to amplify the bias in data sets. This could likely be the case with COMPAS as the system fails to balance the racial bias in their predictions. Others may include systematic bias within the justice system that may distort the measurement of recidivism.

4. The proposed solution uses equal opportunity measure which strives to ensure that all the races have an equal chance to be labeled as a recidivist. This essentially tries to suppress any kind of bias in the data thereby providing fairness towards all the races.

Our primary responsibility as volunteers of a humanitarian NGO is to ensure that we model a system that is sensitive to every race. This trade-off could certainly reduce the overall accuracy of the model but results in a fair system.

5. The proposed solution is better than other alternatives as we obtain a True Positive Rate (TPR) with an utmost difference of 0.02 between all the races. Also, in our secondary optimization method, we chose accuracy as the criteria over cost which remains in line with our idea to create a fair and just system over any sort of business needs. The output data below lists the accuracy, f1 score, TPR, threshold values across all the races. As we can see the TPR/ FNR is constant with an utmost difference of 0.02 across all the races. To generate a model with equal opportunity for all the races we emphasize TPR as the primary metric against all other metrics.

Accuracy on training data: 0.6352886621840946

Cost on training data: $-596,868,332

F1 Score on training data: 0.6939688715953307

Metrics for training data

TPR for African-American: 0.8346613545816733

TPR for Caucasian: 0.789032749428789

TPR for Hispanic: 0.785

TPR for Other: 0.7570093457943925

TPR for all training data 0.803559360216265

The threshold for African-American: 0.1

The threshold for Caucasian: 0.1

The threshold for Hispanic: 0.08

The threshold for Other: 0.1

Accuracy on test data: 0.6168137482582443

Cost on test data: $-150,820,708

F1 Score on test data: 0.7007616974972796

Metrics for test data

TPR for African-American: 0.8962025316455696

TPR for Caucasian: 0.896

TPR for Hispanic: 0.875

TPR for Other: 0.9230769230769231

TPR for all test data 0.8961038961038961

The threshold for African-American: 0.08

The threshold for Caucasian: 0.06

The threshold for Hispanic: 0.06

The threshold for Other: 0.04

6. References

[Retrieved from]. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

[Retrieved from]. https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm

[Retrieved From]. https://github.com/propublica/compas-analysis/blob/master/Compas%20Analysis.ipynb