

Model Choice: Support Vector Machine (SVM)

Algorithm Choice: Equal Opportunity

Secondary Optimization Criteria: Accuracy

Overall Cost of the System's Choices to the Society: \$-757,737,810

Overall Accuracy: 0.6269598292977085

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

COMPAS results in unfair disadvantages for certain ethnicities. For e.g. according to the [analysis conducted by ProPublica on COMPAS](#), even when they ran a statistical test that isolated the effect of race from criminal history and recidivism, as well as from defendant's age and gender: *"Black defendants were still 77 percent more likely to be pegged as at higher risk of committing a future violent crime and 45 percent more likely to be predicted to commit a future crime of any kind."*

Even after removing the race as the sensitive attribute, we still arrive at a hurdle where the sensitivity is hidden in other attributes which skews the result of the model. We are trying to address this disparity for the different race groups.

2. Who are the stakeholders in this situation?

Stakeholders in this ethical setting is anybody who is affected by the organization's actions, objectives and policies or has an interest in this situation. The different stakeholders are:

Defendants: The defendants will be affected by the system's decision, to classify them as high or low risk.

Public: The public has an interest in this situation as it wants fairness across all racial groups. The public can also be affected by the system's decision as falsely labelling an individual as low risk who goes on to commit future crimes will impact them.

Shareholders: The shareholders will be affected by this situation because if our model doesn't perform well our company might not receive future contracts which would cause the shareholders to lose money.

Employees: The employees of the company are affected by this situation because if our model doesn't perform well our company might not receive future contracts resulting in some of them being laid off.

People in the Judicial System: The people of the Judicial System are affected by this because they can come under public scrutiny for being unfair to a particular demographic.

Taxpayers: The taxpayers are affected by the system as the cost to feed prisoners and employee more prison staff will increase if innocent people are falsely classified as high risk and denied bail.

Creditors: Creditors will be affected by this situation because if the model doesn't perform well our company might not receive future contracts and will be unable to pay back its creditors.

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

Age, Race, Sex Hidden bias: Decile Score, Prior count, is recidivist, charge of degree are the biases that might exist in this situation.

Yes, the data on which the model is trained is obtained from a previous judicial system where there are cases of improper judgement which could be a result of conflicted societal definitions for different groups of people. For example: race, sex, age. Also, human error contributes to skewed data values.

There may be biases in the algorithm, decisions made by one designer, or team of designers, may be obscured among the many pieces of code, created for a single program. These biases may create new patterns of behavior in relationship to specific technology as the code interacts with other elements of society.

4. What is the impact of your proposed solution?

Our proposed solution makes sure that the True Positive rates for all the groups are the same.

We also see that our solution gives an acceptable maximum difference between the false positive rates and the true positive rates across racial lines. After taking different methods into consideration, our proposed solution does not discriminate along hidden bias / borderline cases. Our system also reduces the cost to society

Considering the legal system in a public sphere and how disparate impact laws amount to affirmative action by systems, equal opportunity is a standard that given other pursuits of justice doesn't discriminate in any other way to cover up one discrimination but is substantially stated to converge groups of ethnicity as equals and dissolve the difference to a plausible value.

Impact on Stakeholders: Our system produces good results for the fairness parameters thus treating the defendants across racial lines equally. The public will be safer when the system correctly labels high risk individuals as they won't be able to commit future crimes. It reduces the cost the cost to society, thus impacting the shareholders, taxpayers and creditors.

5. Why do you believe that your proposed solution 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?

Other models have a higher disparity in the TPR, TNR, FPR and FNR values across racial lines.

The following table show the difference between the highest and the smallest values of TPR, TNR, FPR and FNR for the different races.

	TPR	TNR	FPR	FNR	Total Cost (- \$)	Total Accuracy
Demographic Parity	0.0394234	0.0383567	0.0383563	0.0394234	757,300,176	0.62798033212
Equal Opportunity	0.0079184	0.0160341	0.0160341	0.0079184	757,737,810	0.62695982929
Maximum Accuracy	0.5260903	0.5267135	0.5267135	0.5260903	739,317,978	0.63929863623
Predictive Parity	0.5301089	0.5776250	0.5776250	0.5301089	749,328,894	0.63196957046
Single Threshold	0.26812955	0.32752377	0.32752377	0.2681295	743,987,808	0.63605158177

As we can see from the above table Equal Opportunity seems to be the fairest when considering the TPR, TNR, FPR and FNR values. Our model gives TPR, TNR, FPR, and FNR values that are around 1% of each other across racial lines. Its readily apparent that Maximum Accuracy, Predictive Parity and Single Threshold give results that have a significant disparity w.r.t these fairness parameters. Demographic Parity might seem like a good approach initially but we are just ensuring that the percentages of people classified as recidivist remains the same across racial lines. For e.g. if African-Americans have 50% of people being classified as recidivist then Demographic Parity just ensures that the same 50% of people from other racial groups get classified as recidivists irrespective of whether are actually high or low risk.

Market Evaluation Competition:

Model Choice: Support Vector Machine (Support Vector Regression SVR)

Algorithm Choice: Equal Opportunity

Secondary Optimization Criteria: Accuracy

Overall Cost of the System's Choices to the Society: \$-111,415,594

Overall Accuracy: 0.761263353460288

Runtime: 141.53085851669312 seconds

References:

1. [*Machine Learning Fairness Primer: Dakota Handzlik*](#)
2. [*Equality of Opportunity \(Stanford Encyclopedia of Philosophy\)*](#)
3. [*Algorithmic Bias \(Wikipedia\)*](#)
4. [*Approaching Fairness in Machine Learning*](#)