HW1-XAI

Paweł Narkiewicz

October 2024

1 Task 2

Here, I trained logistic regression model on adult dataset from benchmark. We can notice that statistical (demographic) parity is much higher compared to equal opportunity and predictive parity value. It means that Logistic Regression predicts that man are 6 times more likely to achieve income over 50k dollars rather than woman.

Metric	**Value**
Statistical Parity	6.1483
Equal Opportunity	1.1091
Predictive Parity	1.0310
Accuracy	0.8474

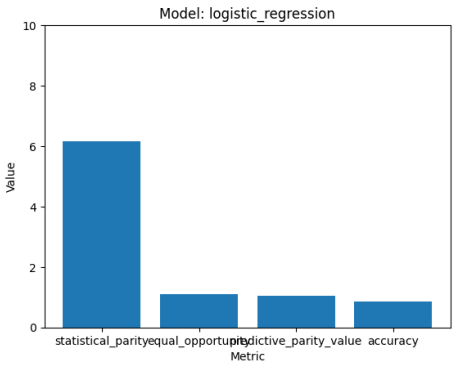

Figure 1: The columns in order are: Statistical parity; Equal opportunity; Predictive parity value; Accuracy

The results of Decision Tree are very similar compared to Logistic Regression results. However, the small differences are: higher statistical parity, equal opportunity, accuracy and smaller predictive parity value.

Metric	Value
Statistical Parity	6.5989
Equal Opportunity	1.1230
Predictive Parity	0.9726
Accuracy	0.8563

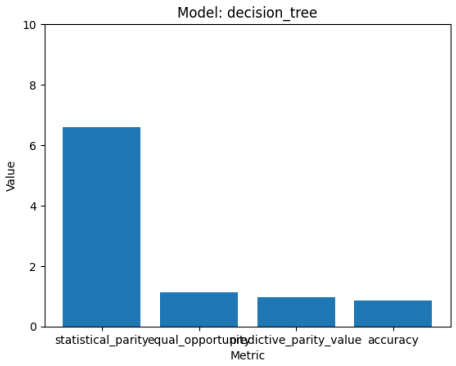

Figure 2: The columns in order are: Statistical parity; Equal opportunity; Predictive parity value; Accuracy

The class imbalance technique that I used is supplying the dataset with random samples of the underrepresented class with respect to the income (target column). It resulted in higher statistical parity, equal opportunity. It also looks like the predictive parity value is almost 1, but model accuracy is lower.

Metric	Value
Statistical Parity	7.2668
Equal Opportunity	1.2460
Predictive Parity	0.9975
Accuracy	0.8211

There seems no correlation between Statistical Parity, Equal Opportunity, Predictive Parity vs Accuracy. However, it looks like there might be positive correlation between Statistical Parity and Equal Opporutunity. Moreover, the class balancing made both statistical parity and accuracy worse which can be linked to the fact that man are overrepresented in those having an income above 50k dollars.

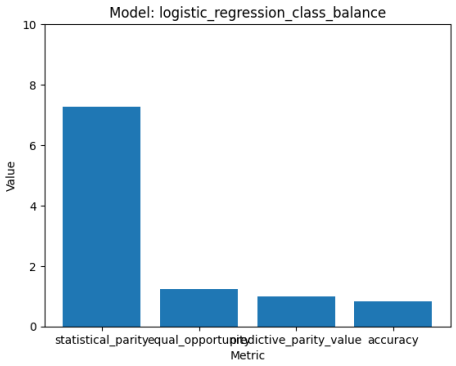

Figure 3: The columns in order are: Statistical parity; Equal opportunity; Predictive parity value; Accuracy