Did feature selection and reduced the attributes to 19 from 53

Created dummy values for 19 attribute values

Train test split -> 33% for test data

Applied random forests -> accuracy = 97%

Scaled the data using standard scalar and applied logistic regression -> accuracy = 96%

Applied KNN with k=5 and got accuracy = 83%

Applied XG Boost and got accuracy = 96%

**Task 1: [False +ve’ s]**

To check for false positive rates, i.e., predicted individual as recidivated but did not actually recidivate

for African American and Caucasians

African Americans = 43 people were misclassified as recidivated by the random forests model, but they didn’t actually recidivate.

19 Caucasians were misclassified as recidivated by the model, but they didn’t actually recidivate.

**Task 2: [true +ve’ s]**

Probability that an individual recidivates given that they both are **predicted positive** by your model and categorized as **African American** by the race variable and **compare** that to the probability that an individual recidivates, and both **predicted positive** by your model and categorized as **Caucasian** by the race variable.

290 / 290 + 677 = 290/967 = 30% – Caucasian

578 / 578 + 389 = 578/967 = 59.7%– African American

Where 967 individuals are predicted +ve and actually recidivated (**true positives**)

**Using Random Forests**

**Accuracy = 90%**

Without Race attribute:

False positives = 118

Number of African Americans incorrectly classified as recidivated= 64

HIGH = 27 MED = 16 LOW = 21

Caucasians = 28

HIGH = 4 MED = 6 LOW = 18

Total number of True positives = 930

Number of African Americans correctly classified as recidivated = 571

HIGH = 209 MED = 193 LOW = 169

Number of Caucasians correctly classified as recidivated = 269

HIGH = 55 MED = 85 LOW = 129

With RACE attribute:

Total number of False Positives = 97

Number of African Americans incorrectly classified as recidivated = 55

HIGH = 25 MED = 17 LOW = 13

Number of Caucasians incorrectly classified as recidivated = 24

HIGH = 4 MED = 6 LOW = 14

Total number of True positives = 929

Number of African Americans correctly classified as recidivated = 567

HIGH = 208 MED = 190 LOW = 169

Number of Caucasians correctly classified as recidivated = 271

HIGH = 56 MED = 84 LOW = 131

**Fair - Demographic Parity Classifier**

Accuracy = 91%

Total number of false positives = 62

Number of African Americans incorrectly classified as recidivated = 34

HIGH = 9 MED = 15 LOW = 10

Number of Caucasians incorrectly classified as recidivated = 18

HIGH = 2 MED = 5 LOW = 11

Total number of True positives = 905

Number of African Americans correctly classified as recidivated = 553

HIGH = 196 MED = 190 LOW = 167

Number of Caucasians correctly classified as recidivated = 265

HIGH = 53 MED = 83 LOW = 129

Objective:

Is to predict whether a defendant in a criminal case will recidivate (be arrested and charged for another crime) within the next two years.

Tried different models and used the model with high base accuracy.

The model is more likely to misclassify African-Americans as high risk over Caucasians.

Table 1: Accuracy of predicting the validation set using different models

|  |  |
| --- | --- |
| Classifier | Accuracy (%) |
| Random Forest | 90 |
| Logistic Regression | 90 |
| KNN | 74.50 |
| XG Boost | 89.92 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **USING RANDOM FOREST CLASSIFIER** | | | | | **USING DEMOGRAPHIC PARITY CLASSIFIER** | |
| **RACE** | **With Race Feature** | | **Without Race Feature** | |
| False Positives | True Positives | False Positives | True Positives | False Positives | True Positives |
| African-Americans | 55 | 567 | 64 | 571 | 34 | 553 |
| Caucasians | 24 | 271 | 28 | 269 | 18 | 265 |
| Total | 97 | 929 | 118 | 930 | 62 | 905 |
| ***In Percentage (%)*** | | | | | | |
| African-Americans | 56.7 | 61.03 | 54.24 | 61.4 | 54.84 | 61.1 |
| Caucasians | 24.74 | 29.17 | 23.73 | 28.92 | 29.03 | 29.28 |

**Analysis of Classifications**

By Random Forest Model [Without Race Feature] By Demographic Parity Classifier

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| \- | **False Positives** | | **True Positives** | |
| **Race** | High Risk | Low Risk | High Risk | Low Risk |
| African-Americans | 27 | 21 | 209 | 169 |
| Caucasians | 4 | 18 | 55 | 129 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| \- | **False Positives** | | **True Positives** | |
| **Race** | High Risk | Low Risk | High Risk | Low Risk |
| African-Americans | 9 | 10 | 196 | 167 |
| Caucasians | 2 | 11 | 53 | 129 |

Slide 1:

Hello everyone, this is Ramaswamy.

So, I’m gonna guide you through the next steps of how we developed our model and discuss about our findings from the experiments.

So, as usual we followed all the standard methods of developing a classifier that includes analyzing the type of features present in the dataset, checking for null values. As mentioned, we chose the two-years variant of the available COMPAS datasets, in which there was too many unnecessary features, and everything was almost present as categorical variables. So, we had to undergo preprocessing steps where we removed columns like names, first name, last name and the columns with arrest-date, screening-date, case number and many others.

So, after cleaning the data, we transformed the selected features into binary forms using the label encoder and onehot encoder and performed train test split to reserve 33% of the dataset as a validation set on which we can predict using our model trained on the rest.

We then applied a few machine learning classifiers such as logistic regression, random forest, etc.. and got the accuracies for the prediction as mentioned in the slide.

So after comparing the results on applying different models, we chose to go with the Random forest classifier as our model for performing experiments to check for bias in terms of opportunity cost and calibration.

Slide 2:

After fitting the training dataset on the Random Forest classifier, we can clearly see that the false positive rates of African-American defendants is around 57% which is significantly higher than the false positive rates of the Caucasian defendants. This means that the model is more likely to misclassify African-americans as high risk that is twice as much as it misclassifies over Caucasians.

Also when we check the true positive rate differences between the two races, it is still in favor of the Caucasians by labeling most of them as low risk and not classifying them as recidivated even though there might be many Caucasians who actually recidivated in the two years.

Even after removing the race feature and conducting the whole experiment, there was no significant change in the rates and the model always seems to be biased against the African-Americans.

So, This bias observed might be due to other features in the dataset such as decile score, gender, etc., which may also play an important role in predicting whether the criminals recidivated or not in the next two years.