Terry Traffic Stops Arrest Model:

Top Features to Predict Arrests

Crissy Bruce ●01.21.2021

Overview

Stakeholder
Bruce Law Firm

Data
Provided by Flatiron School

Problem
The law firm has several clients that are accusing police of targeting them due to race, gender, age and more. So, I have been tasked to determine if there is a link between race or other factors to Terry Stop arrests.

Data Issue: Arrest Flag Data Imbalance

After performing SMOTE, the imbalance has been resolved.

Before SMOTE:
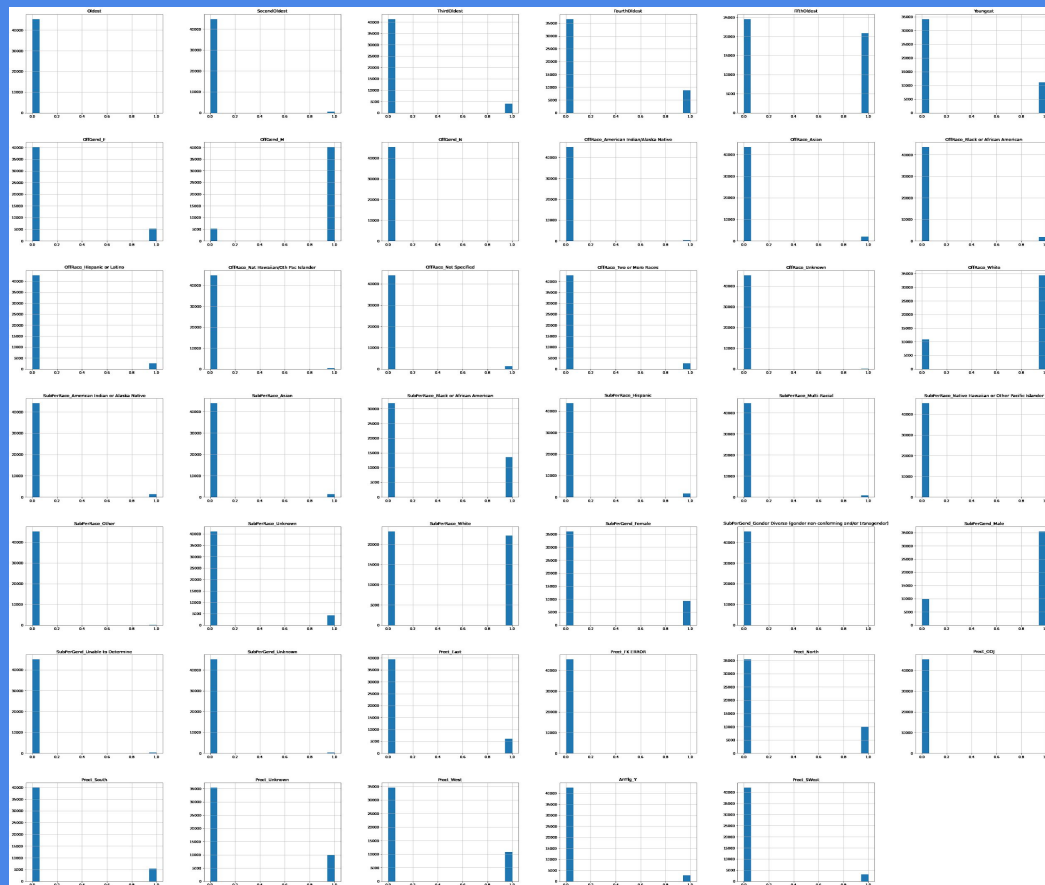
```
Original class distribution:

0    42640
1     2736
```

After SMOTE:

```
Synthetic sample class distribution:

1    29845
0    29845
```

Histograms show that there are certain features that result in higher # of arrests:

- certain precincts
- younger officers
- subject gender.

Final Model

The model that produced the best results was Bagged Trees.

	Decision Tree (Baseline) Model	Bagged Trees Model
Training Accuracy Score	0.77	0.67
Testing Accuracy Score	0.67	0.71
Precision, arrests	0.10	0.12
Precision, no arrests	0.96	0.96
Recall, arrests	0.57	0.58
Recall, no arrests	0.68	0.72
F1	0.76	0.78

Conclusion

The precision score of arrests was quite low in most models, so overall, I think more work is needed.

Although the Bagged Trees ended up being the best, that may change providing the following future work:

- Digging more into the data to find out if Sector & Precinct are related and how to combine them if so.
- Checking the % of demographics to the population in the area to determine if the numbers are in line with the demographics of those being arrested.
- Determine why there are so many Precnt_Unknowns in the data.
- Try other models.
- Explore the idea that maybe training is a factor in that certain precincts have more arrests and younger officers make more arrests.

Thank you!

Project Repo:
https://github.com/crissymae/TerryTrafficStops