



# Terry Stop Arrest Predictor

BY Qi Cai





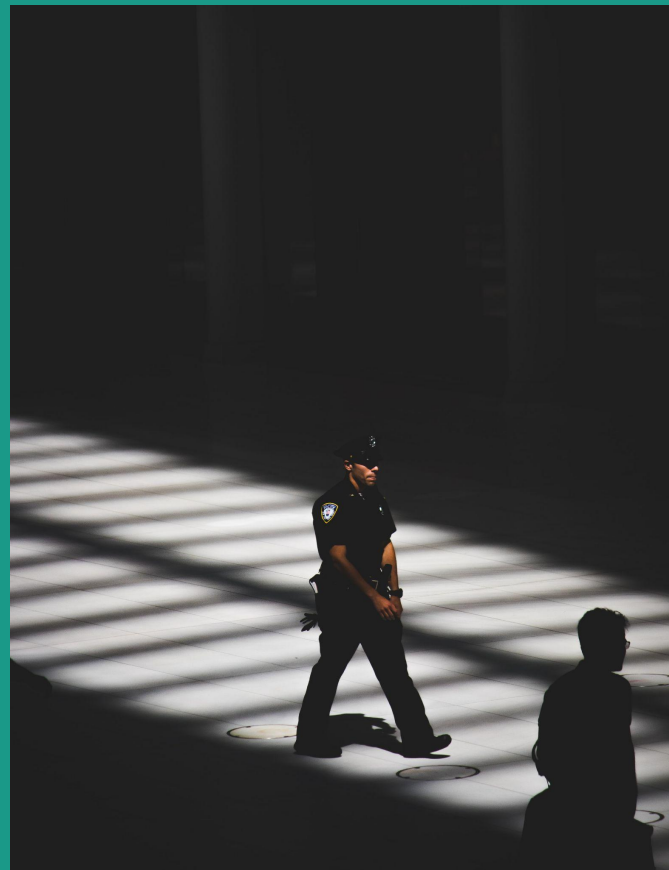
## Overview / Goal

Build a binary classifier to predict whether an arrest was made after a Terry Stop

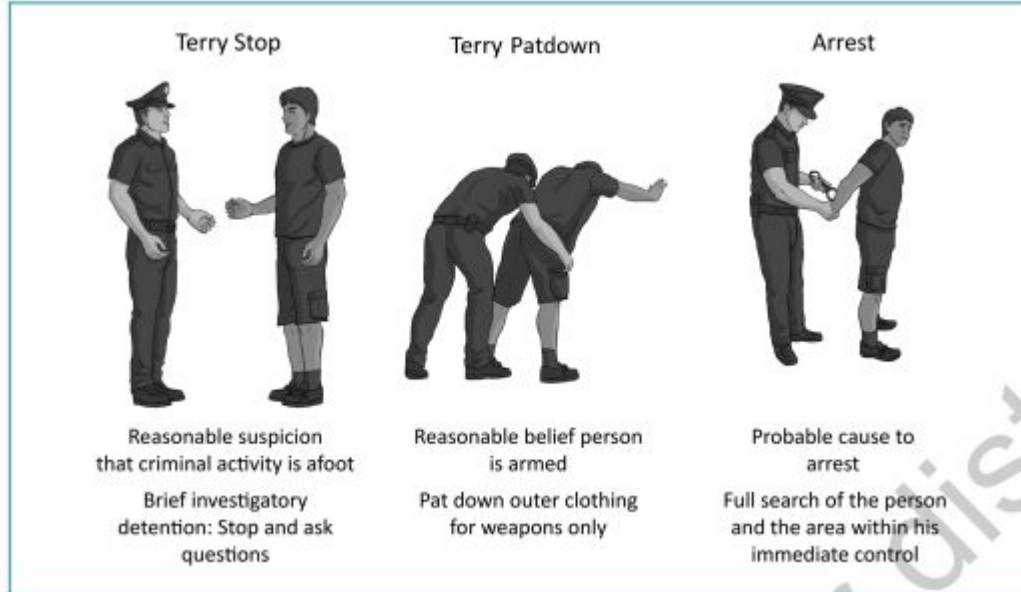


# Stakeholder:

Fellow data scientists who are interested in working in Law enforcement agencies



# What is a Terry Stop?





# Data Understanding

1

Dataset containing 2.96 M rows for every terry stop made by Seattle police officers from 2017 to current.

2

Data Exploration.

Defining Target, features.

Data Cleaning.

3

Trying various Classification Modeling

4

Pick a metric: F1 score

Evaluating Model performance



# Methods and Evaluation Metric:

1. OneHotEncoder
2. SMOTE
3. F1-Score

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} \quad (1)$$

$$precision = \frac{TP}{TP + FP} \quad (2)$$

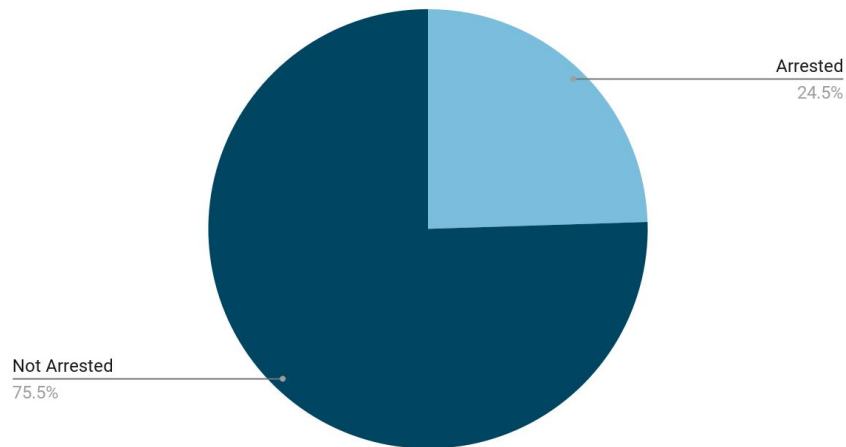
$$recall = \frac{TP}{TP + FN} \quad (3)$$

# Arrest Rate

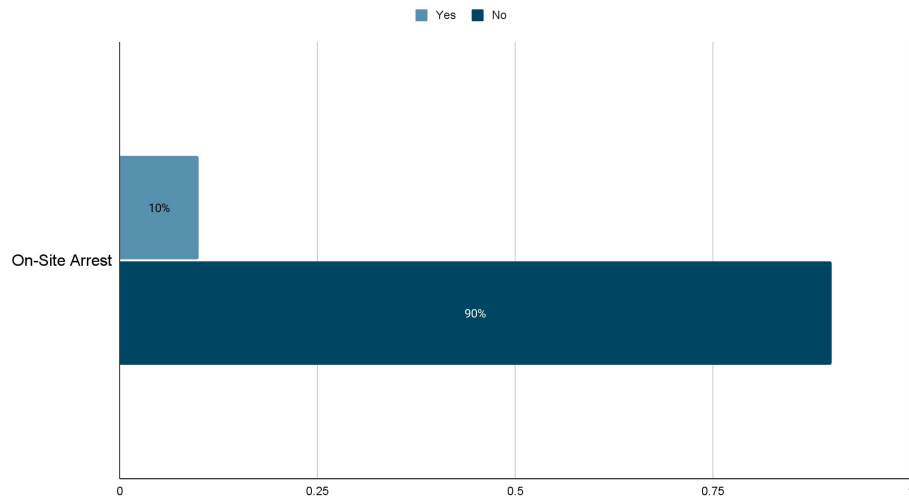


~25% of terry stops end in an arrest. However, only ~10% of arrest happen during the terry stop, while the rest are resolutions decided later on.

Stop Resolution



On-Site Physical Arrest Rate



Arrest Flag & Stop Resolution  
Pearson Coefficient: 0.58



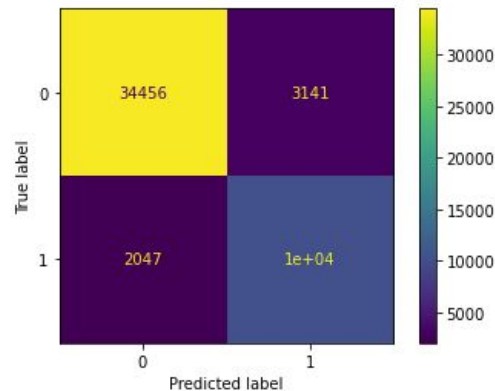
# Logistic Regression

Regression is a technique used to predict a response variable from one or more predictor variables.

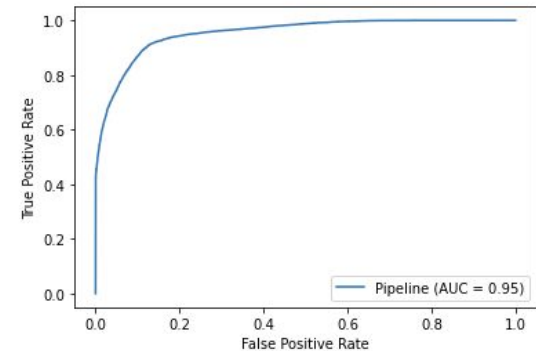
Accuracy = 0.90

F1 score = 0.80

Logistic Regression Confusion Matrix



ROC Graph







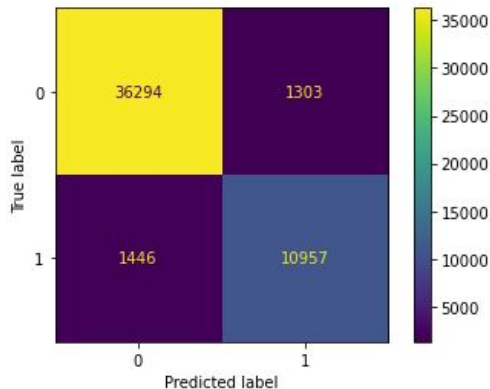
# Decision Tree

The model's decision-making process is represented as a tree-like structure, where each node represents a decision based on a feature and each branch represents the possible outcomes.

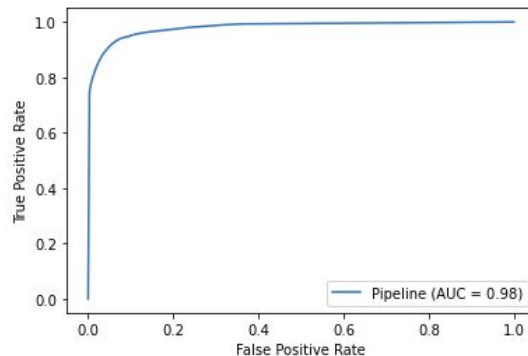
Accuracy = 0.94

F1 score = 0.89

## Decision Tree Confusion Matrix



## ROC Graph



# Conclusion / Recommendation

Decision Tree Classifier is the best model so far, classifying arrests with a F1 score of 0.89.

Feature importances:

Top 1: arrest flag

Top 2: frisk flag



# Recommendations & Next Steps

Using GridSearchCV for Tuning my model to get better F1 score

Looking into feature importance, correlations

Try other models

A short horizontal bar with a teal-to-orange gradient.

# Thank you.

Any  
questions/concerns/comments?

Qi Cai

 [qiechocai](https://www.linkedin.com/in/qiechocai)

[qicai1995@gmail.com](mailto:qicai1995@gmail.com)

