Predicting NYPD Misconduct Case Outcomes and Penalties

Helen Wang

Data Science II, Spring 2025

WHY

1
million people/year
report negative experiences with law enforcement

1/3
complaints
are actually filed against officers

1/24
misconduct cases
result in actual sanctions for the officers involved

GOALS

01

Develop supervised models to ***predict case and penalty outcomes*** for police misconduct cases

02

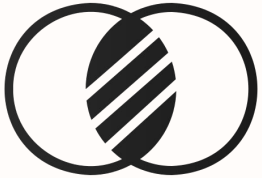
Assess the ***impact that media coverage / visibility*** has on case outcome and resulting penalty.

03

Identify ***other factors most predictive*** of police misconduct case and penalty outcomes.

PREPROCESSING

1



Merged Datasets from the the Civilian Complaint Review Board (CCRB)

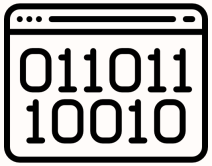
2

Gather media coverage data from Mapping Police Violence dataset and by scraping NYT API

3

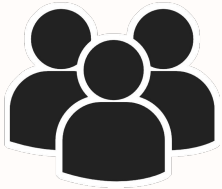
Imputing Missing Data via MICE

4



Dummy Coding and Z-score Normalization

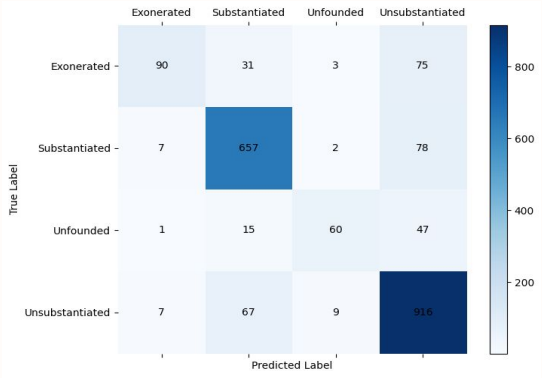
5



Synthetic minority oversampling technique (SMOTE)

COMPLAINT DISPOSITION MODELS

MULTICLASS



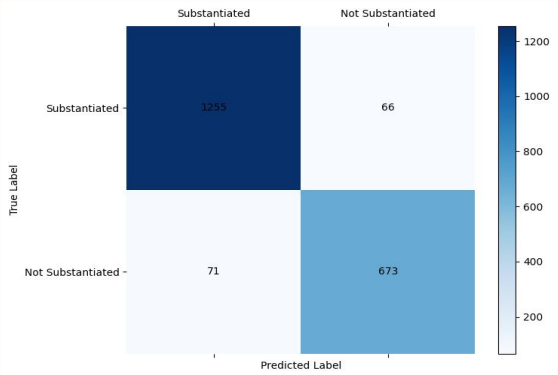
Random Forest

F1: 0.825

Most Important Features:

- Total Substantiated Complaints
- Location and Time Incident Occurred
- Officer Days on Force at Incident
- Video Evidence
- Arrest at Encounter

BINARY



Gradient Boost

F1: 0.933

Most Important Features:

- Video Evidence
- Total Substantiated Complaints
- Recent Previous Case that resulted in Penalty
- Location and Time Incident Occurred

PENALTY OUTCOME MODELS

TYPE

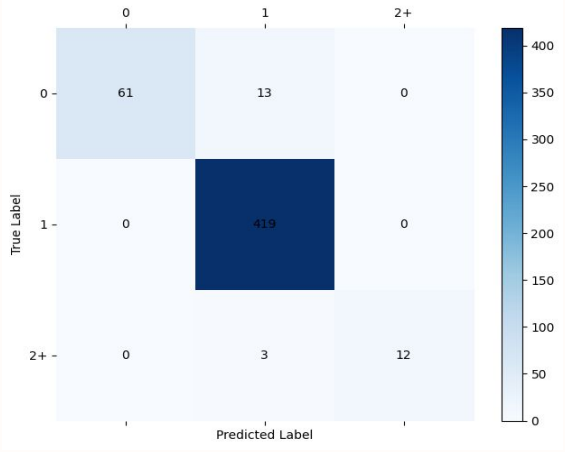
Random Forest

F1: 0.897

Most Important Features:

- Location and Time Incident Occurred
- Officer Days on Force at Incident
- Video Evidence
- Total Complaints

COUNT



Random Forest

F1: 0.967

Most Important Features:

- Location and Time Incident Occurred
- Officer Days on Force
- Total Substantiated Complaints
- Officer Race