Aaron Kohn

Bellevue University, DSC 630

Fall 2021

Reducing Police Violence

# Abstract

Many encounters between police and civilian results in the use of physical force by the police officer. This study uses data collected by the NYPD in 2011 to see if there is a way to predict whether an encounter is likely to involve the use of physical force by a police officer, before any contact is initiated. The analysis will look at the variables collected and made available by the NYPD that are believed to be available at the earliest stages of an encounter.

# Introduction

Although the New York Police Department’s (NYPD) has discontinued its controversial stop and frisk program, stops and frisks occur all the time. Many daily encounters between police officers and civilian end in a stop and frisk. Even when an encounter is not classified as a stop and a frisk does not occur there are hundreds of interactions between police officers and suspects every day.

Many encounters involve the use of force by a police officer. If an officer can be made aware of the level of risk of an encounter before entering a confrontation, steps can be implemented to proceed with the necessary tools to ensure a safe resolution for all involved. This includes providing the officer with the tools to identify situations that are likely to result in the officer using force. The officers can be trained in techniques in how to conduct themselves during these interactions to ensure that they are resolved in the safest outcome possible. This benefits the individual stopped, the officer personally, the entire police force, and society in general.

# Methods

The NYPD has collected and made available data on stops conducted since the year 2003. Each year contains hundreds of thousands of encounters. This study at the first stage focuses solely on the 2011 years data. This year was selected since it is believed to be the peak of New York Cities stop and frisk program.

## Target Values

Identifying cases where physical force was used follows the NYPD’s classifications of physical force and includes the use of handcuffs, drawing a weapon along with pushing a suspect against a wall or the floor among others. Models will be tested to predict the overall use of force and if possible what kind of force was used.

## Variable Selection

In order to ensure that variables that are unknown at the time of model deployment do not affect the model, all variables need to be inspected to ensure that the information will be available to the officer prior to initiating an encounter. This requires manual inspection of each variable and an understanding of what it represents these variables are then removed from the data set that will be used for the model.

Ethical consideration must be given whether to include all variables in the final model. Tentatively, the study will not exclude any variables due to ethical considerations. This decision is made based on the purpose of this study which is to train police to be self-aware of their actions. If bias is discovered in current practice this should be addressed and decisions on ethical concerns can be made at that point.

Still to be addressed, is feature selection and engineering using various models to determine relevant features of the data set to the target variables.

Data Preparation

Histograms and bar plots are harnessed to quickly visualize the distribution of the individual variables. These also help identify outliers and mislabeled data. The age column which can be used in place of the dob column contains many wrongfully labeled data. When possible, correct age can be calculated based off the dob and datestop columns. Care must be exercised when converting the columns to date time objects that dates are parsed correctly.

Outliers need to be studied to ensure that variables are correctly labeled. Based on the data size it appears that missing data and outliers can be dropped without effecting the overall model. It is important to analyze outliers to see if there is any noteworthy patterns in these cases themselves.

## Model Selection

Still to be addressed, is modeling techniques to be implemented for the final project. There are a few different models that would be interesting to test on the data to see how they perform. More study is needed to identify the optimal models to use. One technique that is of interest would be to see the effectiveness of a decision tree model to predict physical force.

# Results

Results are not available at this time.

# Discussion/conclusion

No conclusions have been reached yet. This project has once again reinforced the importance of data cleaning. Originally the data appeared to be relatively clean with little missing data. Only on further inspection have challenges and bad data been discovered.

# Acknowledgments

A full list of acknowledgments will be present in the final submission. At this time I would like to thank Jake Rickord who has provided invaluable feedback at every step.

# References

Partial list

Data retrieved from https://www1.nyc.gov/site/nypd/stats/reports-analysis/stopfrisk.page