Aaron Kohn

Bellevue Univrsity DSC630 Predictive Analytics

Use of Force in NYPD Encounters

September 12, 2021

# Introduction

## Background

Although the New York Police Department’s (NYPD) controversial stop and frisk has been found unconstitutional, stops and frisks occur all the time. Many daily encounters between police officers 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. 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.

## Problem Statement

Many encounters involve the use of force by a police officer. Avoiding the use of excessive force during an encounter can benefit the safety of all involved. 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 a profile of individuals likely to act in ways that would typically require the use of force as well as situations that are likely to result in the officer using force.

## Scope

The object of this project is to accurately predict encounters that would end in physical force being used. An accurate model should be able to predict what type of force was used.

This project is being performed solely by Aaron Kohn, and the responsibility to deliver all work in a timely manner will be his.

The data used is assumed to be accurate.

## Document Overview

This proposal outlines the technical approach, data sources, development requirements, model selection and deployment, testing and evaluation, and the project plan and risks.

# Requirements

## Technical Approach

The project will be conducted following the CRISP-DM methodology for data analysis. Using the methodology as a guide will ensure that the necessary steps are completed accurately. The primary goal of this project is to be able to predict at an acceptable level of accuracy the level of force used by officers in an encounter

## Data

The data for this project has been retrieved from [Publications, Reports - NYPD (nyc.gov)](https://www1.nyc.gov/site/nypd/stats/reports-analysis/stopfrisk.page), and uses the year 2011 dataset. The dataset contains more than 685,000 encounters from the year 2011 with 112 variables for each encounter. There are nine target variables indicating if a specific type of physical force was used. Many of the variables are not of interest since they may not be known at the time decisions would need to be made. There may be more than one time in an encounter when decisions are made it would be necessary to identify the point most beneficial for implementing the model. Identifying the information available at every stage is required to ensure the model usability.

## Requirement Development

The requirements to complete the project include

Data that is accurate and complete. The accuracy and completeness of the data is assumed based on the methods used to collect the data. An understanding of what the variables mean in order to determine the chronological order of when the data is available.

Programs to be used include Python (Jupyter Notebook) which will be the primary tool for data wrangling. R which will be used for exploratory data analysis. Other steps will be performed by the program that will be determined to be optimal at the time.

## Model Deployment

Once a model is selected and determined to have reached an acceptable level of accuracy, the model can be deployed. When deployed on actual encounters, results would be measured by the improvement based on actions taken. These actions would need to be controlled for in further use of the model to continually improve the accuracy of the model and be able to update the model.

## Testing and Evaluation

The model will be evaluated using a confusion matrix as well as accuracy, precision, recall, and F1 scores. A baseline would need to be established to measure improvement of model predictions.

Testing of models can be performed using either a subset of the data or using instances from a later year’s data that is available.

# Expected Results

The expected result of the project is to be able to find variables that contribute to the likelihood of an encounter resulting in the use of force. These variables should be able to be used to grade the probability of an encounter reaching a point requiring the use of force.

# Project Execution

## Project Plan

The project will follow the following timeline

1. Week 1. Milestone 1 (project plan) Done
2. Week 2. Milestone 2 (project proposal) Done
3. Week 3. Data wrangling and data understanding
4. Week 4. Model selection
5. Week 5. Milestone 3 (preliminary analysis) Due
6. Week 6. Model tuning
7. Week 7. Final model tuning and visualization selection for presentation
8. Week 8. Project status assessment
9. Week 9. Milestone 4 (project presentation and status) Due
10. Week 10 – 11. Final revisions to model and presentation
11. Week 12. Milestone 5 (Project completion) Due

## Project risks

Whenever one is dealing with human interaction it is important to be wary of racial bias that may result from predictive analytics. The program in its original for was found to be unconstitutional due to racial profiling. Care must be given that the project is not abused in an illegal manner.  
The accuracy of the project depends on being able to rely on the data collected. If there is any way for data to falsely entered in order to protect the parties involved the study will be compromised. The study assumes the information is verified and accurate.

Care must be exercised to only use variables to create the model that are available at the time the model is expected to be implemented. All variables will need to be analyzed to ensure that there is no crossover into data that is going to be unknown.