How Does the Public Interact with Police in Washington, D.C.?

Statistical Machine Learning–Stat 627

Jithendra Sadu, Samuel Mweni, Wesley Wilson

2023-12-06

## Executive Summary

Now more than ever, state and local governments are actively working to curb police bias and brutality while promoting community policing. The effective use of available data is essential in achieving transparency, accountability, and well-informed decision-making within law enforcement agencies. In this context, we endeavor to comprehend individuals’ interactions with the police in Washington, D.C. Through the construction of machine learning models, we aim to predict key aspects such as police stop duration, the likelihood of a search occurring, and the overall outcome of police interactions. Our classification model demonstrated an approximately 60% accuracy in predicting stop outcomes. Our Lasso model effectively estimate average stop duration and identified key predictors of police stop durations, including ethnicity and specific stop reasons, signifying their significant impact and highlighting potential areas for investigating fairness and bias in policing practices. These findings are especially relevant to the local D.C. council members, police officers, and the general public. [[1]](#footnote-20)

## Research Question

How do individuals interact with the police in Washington, D.C., and what are their implications for community-policing?

## About the Data

The data originated from the District of Columbia’s Open Data portal and includes stops by the Metropolitan Police Department (MPD) from January 1, 2023, to June 30, 2023 D. C. Open Data (2023). The data covers various types of stops, including vehicle, pedestrian, bicycle, and harbor stops, and encompasses outcomes such as tickets, investigatory stops, protective pat-downs, searches, and arrests. Much of the data provided was in response to the legislative requirements from the Neighborhood Engagement Achieves Results Amendment Act of 2016 (NEAR Act), which was designed to improve public safety in D.C “D.c. Law 21-125. Neighborhood Engagement Achieves Results Amendment Act of 2016. | d.c. Law Library” (2016). The dataset includes details from both tickets and MPD’s Record Management System (RMS), and each row represents an individual stop. Information in tickets includes the subject’s race, gender, reason for the stop, and duration. The Open Portal homepage also mentions nuances in data representation, such as differentiating between ticket and non-ticket stops. Additionally, the data glossary outlines data auditing processes, defines variables, and emphasizes the ongoing nature of data quality checks and the possibility of figures changing due to delayed reporting D.C. Open Data, ArcGis (2023). The original data set includes 72 columns and 34,292 rows.

### Data Challenges

The police data encompasses all police activities in D.C., including non-traditional policing, such as harbor police, who provide law enforcement support for boating traffic and D.C. imports and exports. However, as our focus is on understanding interactions between regular community members and the police, observations related to these non-traditional activities are excluded from our analysis. The dataset also contains numerous dummy variables that may only hold relevance for specific instances of police stops. For instance, these variables capture details such as whether the person subjected to the stop consented to a search, if the search was executed in response to a warrant, and whether any property was seized. To streamline our analysis and maintain simplicity, we primarily concentrate on the fundamental variable indicating whether an individual was subjected to a search. This decision allows us to prioritize clarity and coherence in our analysis.

Data entry errors emerged as a concern during our analysis, particularly evident when examining stop durations. Some stop durations exhibited negative values, while others lasted unrealistically long, even up to 10,000 minutes, implying that an officer spent an entire week on a single stop. While it is plausible for an officer to dedicate several hours to a single stop, especially in response to a severe infraction early in their shift, it is important to note that D.C. police policies restrict officers from working more than 18 hours in a single day. Given the unlikely occurrence of stop durations exceeding 8 hours naturally, we classify such instances as data entry errors or outliers and exclude them from our analysis. This ensures the integrity and accuracy of our examination of police interactions within the specified parameters.

## Methodology

### What will the Outcome of my Stop Be?

The first way we wanted to understand policing in D.C. is by predicting stop outcomes. When a police stop happens, will it result in a no action, a warning, a ticket, or an arrest? Thus, our overall approach was to our independent variables–such as stop duration, subject demographics, time of day, etc.–to build several classification models and select the model that most accurately predicted stop outcomes. We first created the stop outcome variable by dividing the stop-type variable in the dataset into mutually exclusive categories; this was challenging because a subject might receive a warning, a ticket, and an arrest for different infractions within a single stop. Thus, we defined the overall stop outcome as the “worst” of all possible outcomes. For instance, if a subject was arrested and also received a ticket, we classified that instance as an arrest.

We then prepared our independent variables and conducted an exploratory analysis of the cleaned data, which revealed that the independent variables were not normally distributed or had equal variances between the different stop outcomes. Thus, we preliminarily believed that the non-parametric KNN method would perform best, followed by Quadtric Discriminant Analysis (QDA) and Linear Discriminant Analysis (LDA). Given the size of our dataset, we conducted a 70/30 training-testing data split for cross-validation.A significant challenge arose from the independent variables forming a perfect linear relationship with the outcome variable. For instance, two of our independent variables were the counts of tickets or warnings issued. If a subject received two tickets, it implied they would not be categorized as having received a warning. This issue recurred with various other independent variables, including the traffic division’s involvement, the stop’s reason, or whether the subject’s property was searched, an action typically associated exclusively with an arrest.

To regularize the model, we employ lasso regression for variable selection. Interestingly, the lasso regression drops certain levels of the model matrix variables but not any one variable exclusively. Therefore, we retain all variables. We then prepared our data for KNN. Since only a handful of our independent variables were quantitative, we first consolidated several independent variables into dummy variables where possible. For example, we transformed gender into male (yes|no) and ethnicity into white or non-white. For the first knn model, we tried to retain the remaining categorical variables by transforming our data into a model matrix, which enumerated the factor variables into dummy variables. We then fit two KNN models using the regular, non-matrix stop data: a non-tuned KNN model with k = 1, and a tuned KNN model. Overall, the most flexible KNN performed the best, with an overall accuracy of 59%.

To regularize our model, we employed lasso regression for variable selection. Notably, lasso regression dropped certain levels of the model matrix variables without excluding any single variable exclusively. Consequently, we retained all variables. Subsequently, we prepared the data for k-Nearest Neighbors (KNN). Given that only a few independent variables were quantitative, we consolidated several variables into dummy variables where applicable. For instance, we transformed gender into a binary variable (male: yes/no) and ethnicity into white or non-white. In the first KNN model, we attempted to retain the remaining categorical variables by transforming the data into a model matrix, which enumerated factor variables into dummy variables. We then fitted two KNN models using the regular, non-matrix stop data: a non-tuned KNN model with k = 1 and a tuned KNN model. The most flexible KNN model yielded the best performance, achieving an overall accuracy of 59%. If we include the number of tickets or warnings variable, our model accuracy increases nearly 20%; however, these variables create perfect linear relationships with the outcome variables are excluded for practical purposes.

Next, we fit cross-validation LDA and QDA models on the stop data. The LDA and QDA models resulted in 49% and 48% accuracy rates, respectively. Considering the presence of unequal variances in the data, it was anticipated that the QDA model would outperform the LDA model. However, both models faced challenges with certain categorical variables. The QDA model encountered difficulties accepting many categorical variables, suggesting that the observed performance discrepancy is more likely due to the restricted use of independent variables rather than an inherent superiority of either method. The accuracy rates of all classification models is below:

|  |
| --- |
| Stop Outcome Classification Model Accuracy |

Overall, the most flexible, non-parametric KNN model performed the best. Still, with an accuracy rate of only 59%, the KNN model leaves something to be desired. This could be due to our feature selection or the available predictors not providing a clear trend in predicting stop outcomes. If the latter is the case, there may be unobserved factors influencing stop outcomes, like the demeanor of the subject or the race of the officer. Given more time, segmenting our outcome variables into two classes and using logistic regression, trees, or random forests might have produced a more accurate model.

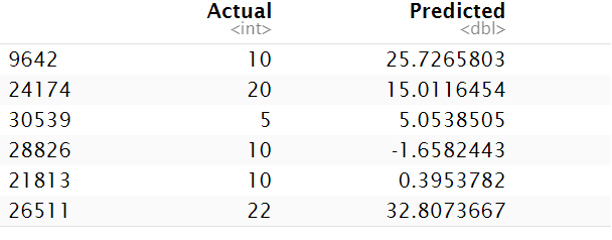
### How Long Will the Stop Last?

Methods: Ridge and Lasso Regression

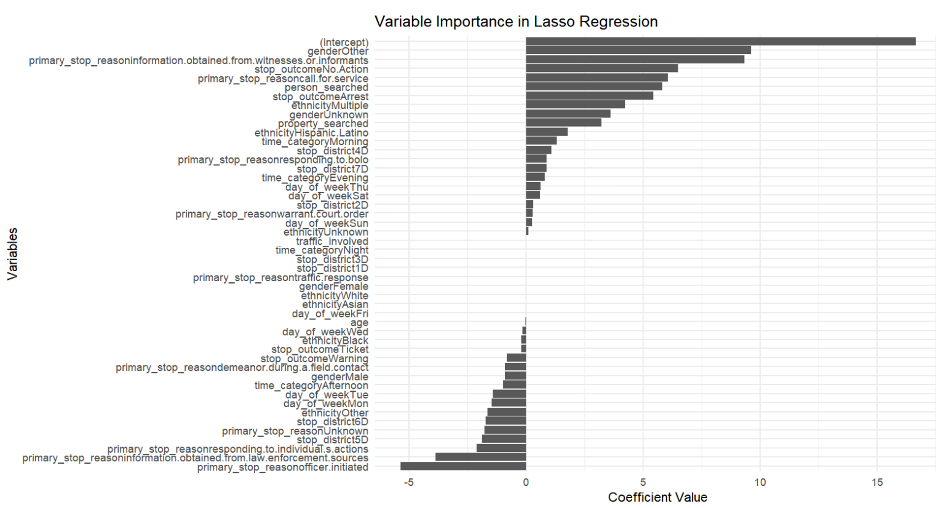
Mae For Ridge Regression : 1.3801

Mae For Lasso Regression: 0.3012663

We ended up selecting Lasso because it had the lowest Mae; This means that among the two model Ridge and Lasso, Lasso is the best Model in predicting Stop duration (see output below).



Variable Importance Plot



## Results

In predicting stop outcomes, the lasso regression retained all variables, and the most-flexible K-Nearest Neighbors (KNN) model performed best with a 59% accuracy. Cross-validation LDA and QDA models achieved 49% and 48% accuracy, facing challenges with certain categorical variables. The overall conclusion underscores the limitations of the 59% accuracy in the KNN model, pointing to potential issues with feature selection or the predictive capacity of available predictors in determining police stop outcomes.

### Stop Duration: Interpretation of Variable Importance from Lasso Regression

**Key Predictors Identified**

* The Lasso model identified **ethnicityOther**, **primary\_stop\_reasoninformation.obtained.from.witnesses.or.informants**, **primary\_stop\_reasonresponding.to.bolo**, **person\_searched**, and **primary\_stop\_reasoncall.for.service** as significant predictors of the duration of police stops.
* The significance of these variables suggests that they have a substantial impact on how long a police stop might last. For instance, stops involving individuals of **ethnicityOther** or stops initiated due to information obtained from witnesses or informants tend to have different durations compared to other stops.

**Interpreting the Predictors**

* **Ethnicity Factor**: The model suggests that the ethnicity of the person stopped (particularly **ethnicityOther**) plays a role in determining the stop duration. This might indicate different interaction dynamics or procedures based on ethnicity, which could be a point of interest for further investigation to ensure fair and unbiased policing practices.
* **Stop Reason**: The reasons for the stop, such as responding to a BOLO (Be On the LookOut) alert or calls for service, significantly influence the duration. This could be due to the complexity or urgency of the situations associated with these reasons.

### Interpretation of Average Durations

**Closeness of Actual and Predicted Durations**:

* The average actual duration of a stop is approximately 16.07 minutes, while the average predicted duration from the model is around 15.95 minutes.
* The closeness of these two figures suggests that the model is doing a good job in estimating the average duration of police stops. It indicates that the model captures the underlying patterns in the data well, at least at an aggregate level.

## Recommendations

* The model provides valuable insights into the factors influencing police stop durations, which can inform policy discussions and training programs aimed at improving efficiency and fairness in law enforcement.
* The close match between the predicted and actual average durations shows the model’s effectiveness, though it’s essential to remember that average predictions do not account for individual variability and specific circumstances of each stop.
* Understanding the key predictors can help in tailoring interventions or policies to address the specific needs and circumstances related to different stop reasons and demographic factors.

## References

D. C. Open Data. 2023. “Stop Data – Home Page.” <https://opendata.dc.gov/datasets/DCGIS::stop-data/about>.

“D.c. Law 21-125. Neighborhood Engagement Achieves Results Amendment Act of 2016. | d.c. Law Library.” 2016. <https://code.dccouncil.us/us/dc/council/laws/21-125>.

D.C. Open Data, ArcGis. 2023. “Stop Data Glossary.” <https://www.arcgis.com/sharing/rest/content/items/cc1b18f4c6954ae891d2e262b991b06a/info/metadata/metadata.xml?format=default&output=html>.

1. Machine learning models developed for learning purposes in policing should not be employed for operational use due to the risk of perpetuating biases and contributing to unjust practices. Careful scrutiny and ongoing ethical evaluations are imperative to prevent the unwarranted application of these models and ensure that their use aligns with principles of fairness, accountability, and transparency. [↑](#footnote-ref-20)