Divided States of America: Uncovering Police Violence Discrimination in the USA

David Berthiaume
Harvard University
dberthiaume@g.harvard.edu

Jingxing Fang
Harvard University
jfang@g.harvard.edu

Hari Raval
Harvard University
hraval@g.harvard.edu

Abstract

Excessive police violence is a rampant issue in the United States. Police departments have often been placed under scrutiny for the use of unreasonable and discriminatory coercive actions against civilians. Analysis of fatal police shootings data, Census Bureau data, and FBI law enforcement data uncovers the factors and manners of discrimination that contribute to the application of coercive force, often leading to a fatal shooting. In this work, we classify whether a shooting will occur or not at the zip code tabulation area by applying logistic regression, random forest, boosting, and neural network modeling approaches. Moreover, we formulate a technique to assign risk scores to areas indicating the likelihood of a police fatal shooting. Our modeling results illustrate that the neural network model leads to the highest accuracy of 72.28%. Overall, total population, crime, (lack of) education, and poverty are the most important contributing factors in these fatal shootings.

1 Introduction

The United States enables the local, state, and federal governments to control a limited amount of "state monopoly on violence". In theory, the democratic processes in place ensure that citizens authorize police officers to exercise this monopoly as needed, by applying a reasonable and nondiscriminatory amount of coercive force. Moreover, the law restricts the use of force such that it can only be used under certain circumstances, where such coercive actions are warranted.

However, in practice, such a balance fails to exist. With the number of fatal police shootings steadily increasing on a yearly basis, we find concrete proof in the excessive use of police violence in the United States. Namely, the fatal and tragic experiences of individuals including Rodney Taylor, Breonna Taylor, and most recently George Floyd, among many others, provide further evidence of the excessive use of police force which impacts the lives of individuals, families, friends, and communities at large. While social movements such as Black Lives Matter have risen to the task to combat these injustices and voice their opinions, it is important that we support these movements by examining the data surrounding such unjust exhibits of force.

To better understand police violence in the United States, we need to uncover the latent patterns in the accompanying data to answer questions such as the following: What sorts of factors contribute to whether a fatal police shooting happens or not? Is discrimination a contributing reason behind shootings? What is the risk factor for police shootings when living in a certain area? By processing, investigating, and analyzing fatal police shootings data, Census Bureau data, and FBI law enforcement data we can begin to answer these critical questions which are pivotal to our society's future.

2 Related Work

Prior work has focused on police shootings and their impact on specific groups of people. Researchers have made use of methods to quantify fatal police shootings based on data from the Washington Post as well as mainstream media outlets. With this data, they make use of DBSCAN clustering to uncover fatal shooting hot-spots and examine factors related to social economics, demographics, political tendency, education, gun ownership rate, and police training hours to discover reasons and motivations behind these shootings [12]. Of specific interest, prior work has created regression models to predict police shooting rates at the state level, however work has also been done with classification models to predict the race of fatal police shooting victims. Classification methods used have included gradient boosting, logistic regression, and naive bayes [12].

Other research has included the creation of theoretical models which have combined social learning and social psychological theories to explain why police shootings occur of unarmed suspects. Studies in this area have found that neither theory independently can effectively answer the question as to why such shootings occur. However, when considering both theories in an integrated manner, a model can be created which can be used to generalize such shootings [4].

Furthermore, other researchers have attempted to specifically examine the use of racial discrimination in the application of police force and fatal police shootings. More specifically, prior work has found that on non-lethal uses of force, African Americans and Hispanics are more than 50% likely to experience such coercive treatment [6]. Moreover, even when adding controls to account for the civilian behavior patterns, there are still disparities between the experiences of African Americans and Hispanics compared to other races. Ultimately, this works yields results which make a case for discrimination existing in a fraction of coercive-based police interactions [6].

3 Data

3.1 Data Overview

Our project works with data from three primary sources: Washington Post Fatal Police Shooting data, Census Bureau data, and FBI law enforcement data. We compile all of this data from publicly available sources, as outlined in Section 3.2.2 below. The Washington Post data consists of records relating to every fatal shooting in the United States by a police officer since January 1, 2015. It contains columns relating to the name, date, manner of death, signs of mental illness, threat level, age, gender, race, city, state, and location of the incident. Next, the Census Bureau data consists of, at the zip code tabulation area (ZCTA) level, population characteristics including marital status, sex, age, education level, occupation, household income, computer usage, migration status, and travel time to work. Finally, the data table from the FBI consists of offenses known to law enforcement by state and city. It contains information relating to population, violent crime count for 2020, murder and manslaughter counts, rape, robbery, aggravated assault, property crimes, burglary, larceny, motor vehicle theft, and arson.

3.2 Data Processing

3.2.1 Determining the Geographical Unit

A critical decision for our project is to determine the ideal geographical unit to operate at across the US. We aim to have a balance between the following three considerations:

- For a given geographical unit, it is important to have a relatively homogeneous population and crime rate. If a given geographical unit is too large, such as a county or a state, then this criterion will not be satisfied and we could have drastically different crime rates or population characteristics within a single geographical unit.
- In general, it is desirable to have a relatively balanced data set with a large enough sample of data points containing a fatal shooting. If we choose too small of an aerial unit such as a census block or block group, the vast majority of these areas will not have any fatal police shootings.
- Given the limited data availability, we must ensure that there are enough data points to work with. Specifically, the FBI crime data is only available at larger aggregate levels and for highly populated areas in general.

After careful consideration of these three points, we decide to use ZCTA as our geographical unit. There are around $32,000$ ZCTAs in the US.

3.2.2 Data Sources and Description

For our analysis, we use three sources of data, including 5-year American Community Survey data, FBI crime data, and fatal police shootings data.

The American Community Survey (census) data is available as a series of detailed summary tables and contains $35,942$ variables. From these available variables, we hand-select 179 features that we felt would be most relevant for our analysis. These 179 predictors come from 10 separate summary tables. Table 1 lists the Census Bureau summary tables and their broad categories which indicate where our predictors are selected from.

Census Table Name	Census Table Description
Age and sex	total population, and population by age groups and sex
Race	counts of the population by race
Commuting	distances and travel mode for commuters
Children and household relationships	characteristics of the household and the children
Marital status and history	counts of population that are married, widowed, divorced
Educational attainment	counts of highest education obtained
Poverty	counts of the population at different levels of poverty
Income	income breakdowns by count of population
Computer usage	factors and types of computer usage in the population
Housing characteristics	household sizes and family relationships

Table 1: Selected Census Bureau summary tables

The FBI law enforcement data is available as a table that includes crime type counts per town/city for $7,688$ areas across the US. We note that data for only a select set of populated towns and cities is available. This covers a subset of the ZCTAs for which the census data is available.

Finally, the fatal police shooting data is available as an event-level table that contains the latitude and longitude of each individual shooting along with the associated columns described in Section 3.1 above.

Overall, the most challenging aspect of using all this data is merging and cleaning the files, joining all three sets together at a ZCTA level, and ultimately converting every numeric predictor to per capita rates as opposed to integer counts. We detail this process in Section 3.2.3 below.

3.2.3 Data merging and cleaning

To merge and clean the data, we first download a shape file that contains the geometries (as multi-polygons) for every ZCTA in the US and convert this to EPSG:4326, the world geodetic system. Next, we transform the fatal police shootings data into a point layer using the provided latitudes and longitudes and perform an intersection join with the ZCTA multi-polygons to determine the number of shootings in each ZCTA. We note that we include a handful of events that are classified as "not exactly geocoded" because these events would likely still fall within the correct ZCTA.

For the census data, we perform a join on the 10 tables shown in Table 1 using ZCTA as the primary key. These tables contain our selected 179 predictors. To ensure a successful join, several preproccessing and cleaning steps are required pertaining to the column names. We then merge this combined data with our ZCTA polygons and fatal shooting data.

Next, although the FBI data is provided at the town and city level, we were able to obtain a separate spatial database of all of the town and city names in the US along with the ZCTAs that lie within them. After cleaning the data, we preprocess the names of the towns by removing spaces and non-alphabetic characters. We then use fuzzy string matching using Levenshtein distance to join our ZCTA data with this FBI crime data where it is available and merge this with our spatial ZCTA multi-polygon data. Once this join is complete, we end up with $9,906$ ZCTAs for which we have census data and FBI crime data spread across the US including Alaska, but not Hawaii. Of the original $6,414$ fatal shootings with latitudes and longitudes, $4,365$ of them are within our final ZCTAs. These joins are hand-checked for consistency and accuracy. Figure 1 shows the distribution of the final ZCTAs along with violent crime rate and the fatal shootings that occurred within them (Alaska not shown).

Finally, we note that all of the crime and census data that we obtained is in population counts, corresponding to the number of people that fit a particular category. Hence, our last processing step is to convert all of these counts into populations rates - the number within each category divided by the total population of that ZCTA. For example, property crime count was converted to property crime rate by dividing property crime count by the total population of the ZCTA.

After this data preprocessing, we end up with two primary data sets.

1. *Combined Census and Fatal Shooting Data* - For our first data set, we only include census data merged with fatal shootings data. This allows us to have the vast majority of populated ZCTAs to work with in the US. We make use of this data set to generate a risk score for every single ZCTA in the US by performing a classification and then predicting probabilities with logistic regression.
2. *Combined Census, FBI Crime, and Fatal Shooting Data* - For our second data set, we include merged census data, FBI crime data, and fatal police shooting data. We make use of this data set to perform the majority of

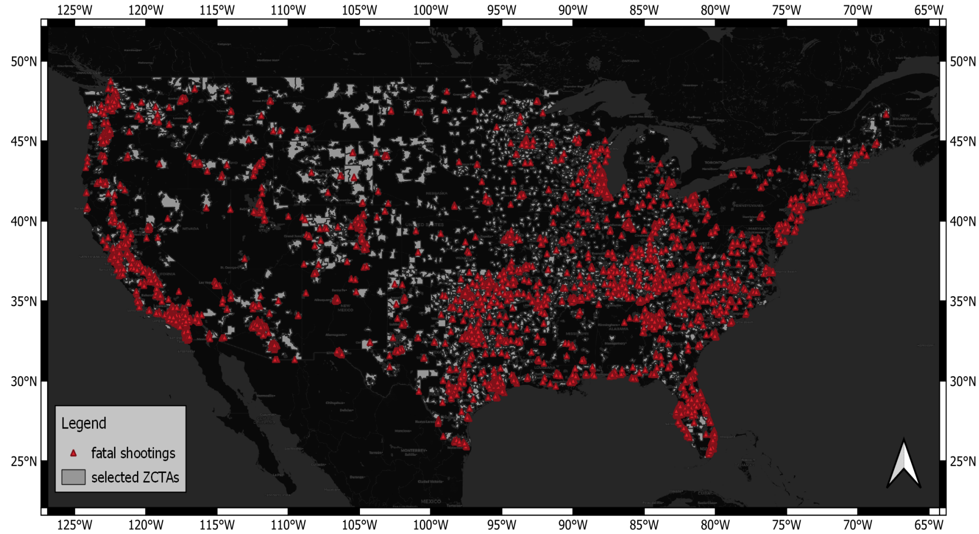

Figure 1: Map of CONUS showing the ZCTAs with FBI crime data in gray and fatal shootings in red

our deep analysis. Unfortunately, this data is not available for the entire US, but contains predictors that are very important in determining the presence of fatal shooting based on our feature importance analysis.

3.3 Exploratory Data Analysis

After merging our data sets as outlined in Section 3.2.3, each sample in the combined data set corresponds to a ZCTA. For each ZCTA, we have nine attributes describing nine types of crime rates, 179 attributes describing the demographic characteristics, and the count of fatal police shootings of this area (areas without fatal police shooting history have count 0). In our exploratory data analysis, we aim to understand the relationship between various crime and demographic statistics of an area and the number of fatal police shootings.

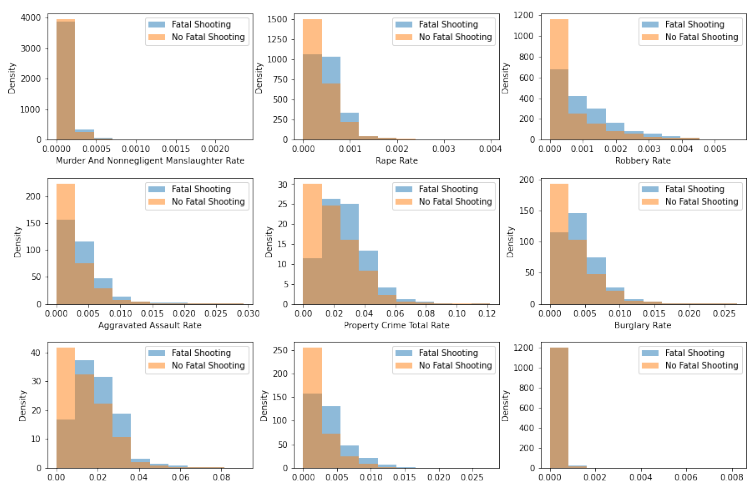

Figure 2: Distributions of nine crime rates among areas with and without fatal police shootings

In Figure 2, we plot the distribution of nine types of crime rates for areas with and without fatal police shooting histories. We observe that areas without fatal police shootings tend to have lower crime rates for all types of crime, and this makes intuitive sense. Most notably, the property crime rate, larceny-theft rate, robbery rate, and burglary rate all have the most stark differences in the distribution between no fatal samples and fatal samples across the US.

In Figure 3, we show the race distribution for 4 notable sample areas. The top-left subplot corresponds to an area with multiple fatal police shootings, while the other three subplots have no fatal police shootings. By examining various

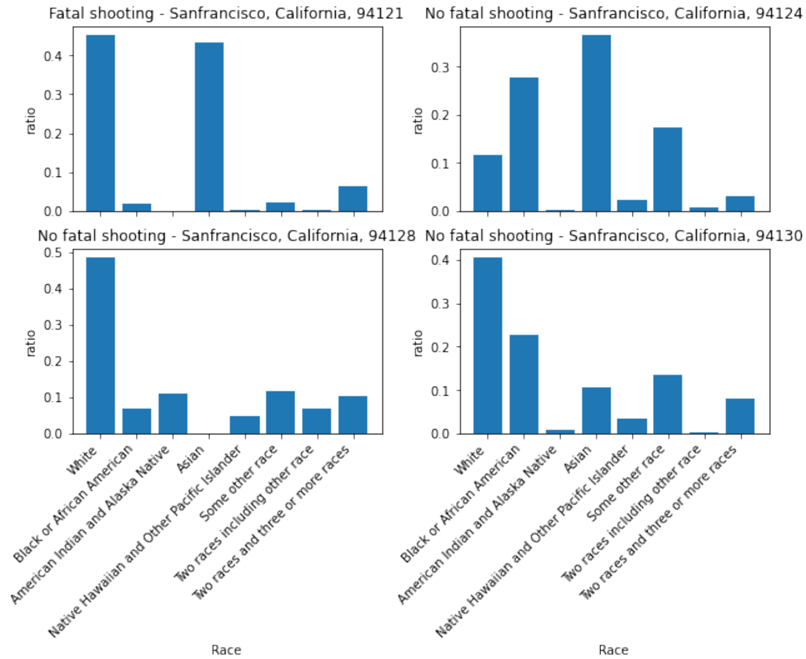

Figure 3: Population distribution by race for four different zip code areas. The top left sub plot corresponds to an area with fatal shootings and the other three subplots correspond to areas with no fatal shootings.

zip code areas in this manner, we began to uncover a key trend in our data: areas with more race diversity tend to have less/no fatal police shootings. This is captured in Figure 3 as we can see that the races are closer to uniformly distributed in the three subplots with no fatal police shootings, while the area with fatal police shootings is primarily made up of a White and Asian population. We believe that the lack of diversity leads to a ”majority” group dominance situation and a certain group is racially discriminated against.

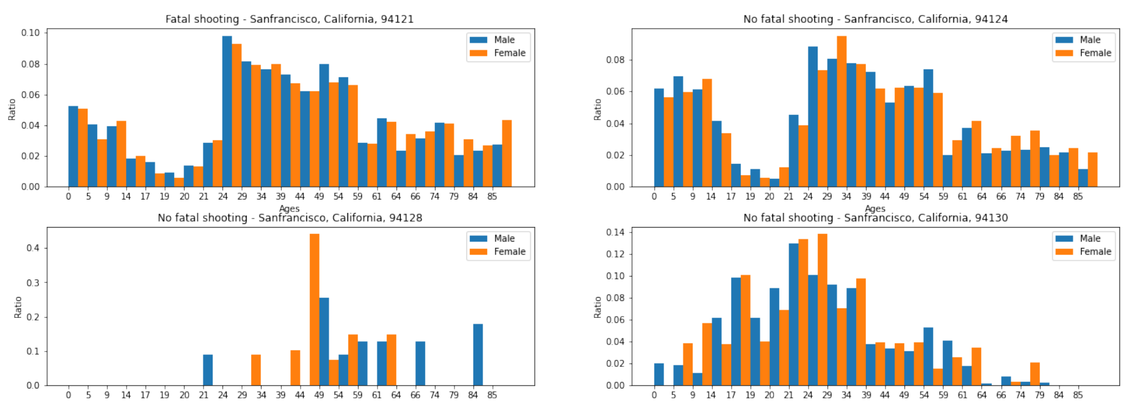

Figure 4: Population distribution by age and sex for four different zip code areas. The top left sub plot corresponds to an area with fatal shootings and the other three subplots correspond to areas with no fatal shootings

In Figure 4, we visualize one area with fatal police shooting (top left subplot) and three of its adjacent areas without a history of fatal police shootings. This enables us to control the variance of the population age distribution due to economic or geographic factors, and focus on how the age distribution correlates with fatal police shootings in an area. We observe that, compared to areas with fatal police shootings, areas without a history of fatal police shootings tend to have fewer 39-to-59-year-old people than the area with fatal police shooting.

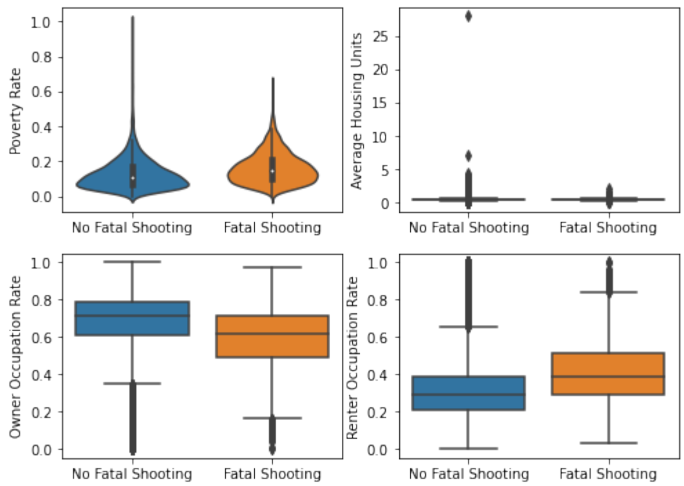

Figure 5: Distributions of poverty rate, average housing units, owner occupation rate, and renter occupation rate across areas with no fatal shootings versus at least one fatal shootings

In Figure 5, we plot the distributions of the poverty rate, average housing units and ratios of housing units occupied by owners and renters for areas with and without fatal police shootings. Most notably, the top two sub plots demonstrate that areas with a higher poverty rate or lower average housing units per resident have a higher probability of facing fatal police shootings. This uncovers a key trend in our data set: discrimination on economic status. Moreover, the bottom two subplots unveil that areas with a higher number of renters also tend to have higher chances of fatal police shootings. This was a surprising result, but is suggestive of the fact that a larger number of renters corresponds to a larger number of unstable or "floating" people, which usually correlates with worse public security.

Overall, through our in-depth exploratory data analysis (additional figures included in the Jupyter notebook), we found multiple noteworthy correlations between the demographic and crime rate attributes and fatal police shootings in the combined data set. These correlations and patterns in the data motivated and inspired our project.

4 Modeling Approach

4.1 Combined Census and Fatal Shooting Data Approach

Our first analysis is accomplished with the census and fatal shooting merged data set, where the data is split into 80% for training and 20% for testing. While this data set has fewer predictors, namely it is missing the crime data that is so relevant to fatal shootings, a key benefit of this more limited data is that we have coverage for most of the US (all populated ZCTAs where census data is available). This data does not have crime rates in it, and as we will see below, crime is a critical component, perhaps not too surprisingly, in predicting whether a fatal police shooting will occur or not. In general, crime is positively associated with police contacts and each police contact has a small chance of becoming fatal in the worst case scenario.

Our goal is to create a map that shows a "risk score" from 0 to 1 of a police fatal shooting occurring within each ZCTA across the entire United States. This is accomplished by performing a cross validated logistic regression where we hypertune an L_2 regularization parameter in the range from 1×10^{-6} to 1. Given that we have nearly 180 predictors in this data set and many of these predictors are highly correlated, we feel it is important to perform some regularization.

Even though our response variable is unbalanced, with just over 13% of the ZCTAs having a fatal shooting in them, for this analysis we do not do any up-sampling such as SMOTE. We decide on this approach since we are generating probabilities, and we also wish to use the original training data for interpretability of the coefficients. We scale the data to ensure that the L_2 regularization does not penalize certain predictors more severely than others just based on scale.

4.2 Combined Census, FBI Crime, and Fatal Shooting Data Approach

While demographics, income, travel time to work, and education, are all important features of a population, we also want to look at data relating directly to crime. Luckily, the FBI provides crime data for a large portion of the US. Most of our work uses this data which covers about one third of the US.

Our second analysis is accomplished with the merged census, crime, and fatal shooting data. In this merged data set, we have nine types of crime rate data, which intuitively seem more relevant to the response. For the analysis of this data set, we aim at figuring out how the features in crime and census data would contribute to a ZCTA area having fatal police shooting events. We select 40 of the most significant features according to their permutation importance scores out of the total 193 features to make our models more focused on relevant predictors.

We use logistic regression, random forest, AdaBoost and a neural network to model the data and compare their results. Our merged data set is quite imbalanced with only 27% positive rate, so to prevent our models from naively biasing towards the majority class (not having fatal police shootings), we up-sample the data through Synthetic Minority Oversampling TEchnique (SMOTE) during preprocessing [2]. The data set is also scaled using a standard scalar before modeling to avoid biased penalization in the logistic regression.

4.3 Methods Used

4.3.1 Feature Importance

We fit a random forest model with 100 estimators at max depth = 15 and calculate the permutation importance for all the features [1]. The permutation importance algorithm requires a trained model $\hat{f}$,the feature matrix X, the response vector y and the measurement of error $L(y, \hat{f})$. To calculate the importance score of the jth feature, we create X_{perm} which is identical to X except that the values of the jth predictor are permuted. Finally, the importance score can be calculated by $FI_j = L(y, \hat{f}(X_{perm})) - L(y, \hat{f}(X))$. Higher FI_j implies that the jth feature is more important.

4.3.2 SMOTE

SMOTE creates and adds synthesized minority class samples to an imbalanced data set for the purpose of oversampling. SMOTE works as follows:

1. Choose a random sample $\mathbf{x}_i$ from the minority class and find its k nearest neighbors in the minority class. Randomly select several samples from the k nearest neighbors according to the re-sampling rate.
2. For each $\mathbf{x}_j$ in the selected neighbors, generate a synthetic sample for the minority class according to
$$\mathbf{x}_{new} = \mathbf{x}_i + uniform(0, 1) \times \mathbf{x}_j$$
3. Repeat Steps 1 and 2 until the re-sampled data set is balanced.

The classic up-sampling method increases the proportion of the minority class by naively duplicating minority class samples, which will easily lead to overfitting. SMOTE can overcome this issue through synthesizing new samples.

4.3.3 Logistic Regression

Logistic regression is a machine learning algorithm that models the probability of a binary outcome from an input variable [3]. It assumes a linear relationship between the logit, i.e., the log-odds, of the outcome and the input variables:

$$\beta_0 + \beta_1 X = \ln\left(\frac{P(Y=1)}{1-P(Y=1)}\right),$$

Above, we note that X is the input and $P(Y = 1)$ is the probability that X belongs to category 1.

From a maximum likelihood estimation (MLE) perspective, we want to choose parameters β_0 and β_1 such that the likelihood of observing our data set is maximized, i.e., minimizing the negative log-likelihood (NLL loss) of the data:

$$L = -\sum_i \left[y_i \ln \frac{1}{1+e^{-(\beta_0+\beta_1 X_i)}} + (1-y_i)\ln\left(1 - \frac{1}{1+e^{-(\beta_0+\beta_1 X_i)}}\right)\right].$$

Besides the naive logistic regression model above, we also tune a logistic regression model with L_2 regularization through cross validation. Regularization adds an additional term $\lambda\beta_1^T\beta_1$ to the loss function during the training to discourage learning a more complex or more flexible model, which prevents overfitting.

4.3.4 Random Forest

A random forest is an ensemble method based on decision trees [8]. A decision tree classifier recursively partitions the data set into two halves in the feature space according to a single feature selected by a certain criteria that maximizes the purity gain of the split. A decision tree can easily overfit the data when the max depth of the tree is too large, as the decision boundary becomes overly complex after too many splits. A random forest reduces the high variance of the single decision tree through fitting single decision trees on bootstrapped data sets and aggregating the results of all decision trees. Specifically, when building these decision trees, the data set is partitioned by a feature from a random subset of m predictors chosen from the full set of predictors, which helps decorrelate the decision trees.

In our random forest model, the size of the random subset of features at each split is the square root of the entire feature set, and we choose Gini Index as the split criteria. There are 30 single decision trees in our random forest and we use cross validation to select the optimal max depth of single decision trees.

4.3.5 AdaBoost

When the max depth of a decision tree is small, the decision tree will have high bias but low variance. Boosting grows a sequence of simple trees, each of which makes use of the error information of the previously grown trees to reduce the bias of the entire model. In AdaBoost, we initialize a model with a simple decision tree and even weights for all samples [5]. Next, at each step, we update the weights of samples through up-weighting misclassified samples and down-weighting correctly classified samples from the previous tree. We then fit a new simple decision tree to the data set and add this new estimator to the model with an estimator weight. For our work, we fit an AdaBoost classifier with 5 estimators and a learning rate of 0.05. We tune the max depth of single decision trees through 5-fold cross validation.

4.4 Model Evaluation and Interpretation

The purpose of this study is to explore the relationship between crime data, census statistics, and the occurrence of fatal police shootings. We want to reach high accuracy with our model, so that the relationship between features and the response learned and represented by the model will also be accurate and informative. Therefore, we use accuracy as the metrics of all our models, which is the ratio of correctly classified samples among all samples.

We leverage SHapley Additive exPlanations (SHAP) to interpret how each feature value in each sample contributes to the result in our model. The main idea of calculating the SHAP value of a feature is as follows (take the jth feature as an example):

1. Get all subsets of features which do not contain feature j and compute the predictions
2. Add feature j to all subsets in 1. and compute the predictions
3. Compute the contribution of having feature j based on results from 1. and 2. and aggregate all contributions to calculate the marginal contribution of feature j:

$$SHAP_{feature}(x) = \sum_{feature \in set} \left[|set| \times \binom{F}{|set|} \right]^{-1} \left[Predict_{set}(x) - Predict_{set \setminus feature}(x) \right]$$

(where set is the set of all features)

SHAP has several advantages over other techniques to calculate feature importance. It not only shows feature importances, but also shows if the feature has a positive or negative influence on results. In addition, SHAP values can be calculated for the prediction result for a single sample and used to analyze how features contribute to that single prediction result. A fast algorithm called TreeSHAP was proposed to calculate SHAP values specifically for tree-based models, such as random forest and boosted trees[9].

5 Spotlight Technique: Neural Network

5.1 Motivation

One motivation for tackling the challenging task of processing data for the entire US is to ensure we have enough samples to comfortably explore a neural network model. With around 200 predictors and $10,000$ data points, we believe that a neural network is a suitable choice to leverage an advanced technique not covered in AC 209A. Moreover, we note that predicting whether a fatal shooting occurs or not is a challenging problem, due to the relative rarity of the event. Hence, a neural network is well-suited to this rare task because by the Universal Approximation Theorem, a neural network is able to create very complex decision boundaries [11].

5.2 Data Preparation

For the neural network model, instead of using SMOTE to up-sample our data, we make use of a simple sampling with replacement scheme to create a balanced data set. This approach is used because adding synthetic data is not necessary for a neural network in this case. As before, we scale the data using a standard scalar.

Moreover, instead of choosing the top 40 predictors, as we do for our other models, we use the top 80 predictors. We can reduce the chance of overfitting by decreasing the size of our neural network and by performing early stopping, which reduces the number of epochs that we train for. As we'll see below, we also apply some modifications to our architecture to further help prevent overfitting.

5.3 Architecture Selection

The type of neural network that we explore in our work is a multi-layer perceptron, one of the simplest types of neural networks. Given the dense connections between the layers of this model, it is interesting to see how the interaction terms between the predictors, implicitly generated by the neural network, help the model predictions [7].

Specifically, our architecture consists of 6 dense layers with 98, 64, 32, 16, 8, and 1 nodes per layer, respectively starting from the inputs to the final output layer. We use ReLU activation functions in each dense layer except the final layer.

To help prevent overfitting, we add dropout layers with a rate of 0.3 in between the fourth and fifth dense layers, and the fifth and sixth dense layers. Without these dropout layers, we note that overfitting occurs more rapidly. Dropout helps to reduce this effect by setting certain weights randomly to zero in our neural network. This helps to prevent the neural network from learning the noise in our data and hence, generalize better. Finally, the output dense layer has a sigmoid activation function which allows us to predict probabilities and generate binary predictions. Figure 6 shows our neural network architecture, a slight modification of a simple multi-layer perceptron.

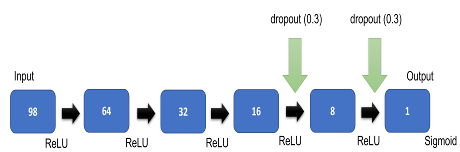

Figure 6: Modified multi-layer perceptron architecture used for neural network

5.4 Training

We use binary cross entropy as our loss function and the Adam optimizer. We use a learning rate of 0.000025 which was selected by trial and error. We train the neural network for 100 epochs. It is important to note that we use a validation set to monitor any potential overfitting and do not use the test data until the very end. Our data here is split into training, validation, and test sets.

6 Results

6.1 Combined Census and Fatal Shooting Data

For our census and fatal shooting data only, a cross validated logistic regression model with L_1 penalty (with no SMOTE up-sampling) leads to a test accuracy of 88.13%, as shown in Figure 8. The best value for C, our penalizing term, is 0.0464. With this L_1 penalization, we observe that most coefficients are set to 0. Figure 9 shows the largest *positive* coefficients in this logistic regression model. These results indicate the predictors that most likely tend to increase the chance of a fatal shooting occurring.

We see immediately that total population dominates as a predictor here and this is also indicative of what we will see with the FBI crime data in Section 6.2. With a greater population in a ZCTA, given all other factors remaining constant, there are more police contacts and therefore a greater chance of a fatal shooting occurring. Our second coefficient is a bit surprising, as it suggests that, in general, when households have two our more computing devices, the ZCTA tends to have more fatal shootings by police. Next, the third coefficient is particularly telling. We find that a higher rate of African Americans in a ZCTA tends to *increase* the chance of a fatal shooting occurring. Unfortunately, we have seen that police discrimination is real and that African Americans can be targets in police violence. In this regard, finding this element as a top coefficient is not surprising. The fourth and fifth coefficients are intuitive since higher rates of poverty and divorce tend to correspond to increases in crime and as we will see below, this corresponds to higher rates of fatal shootings by police.

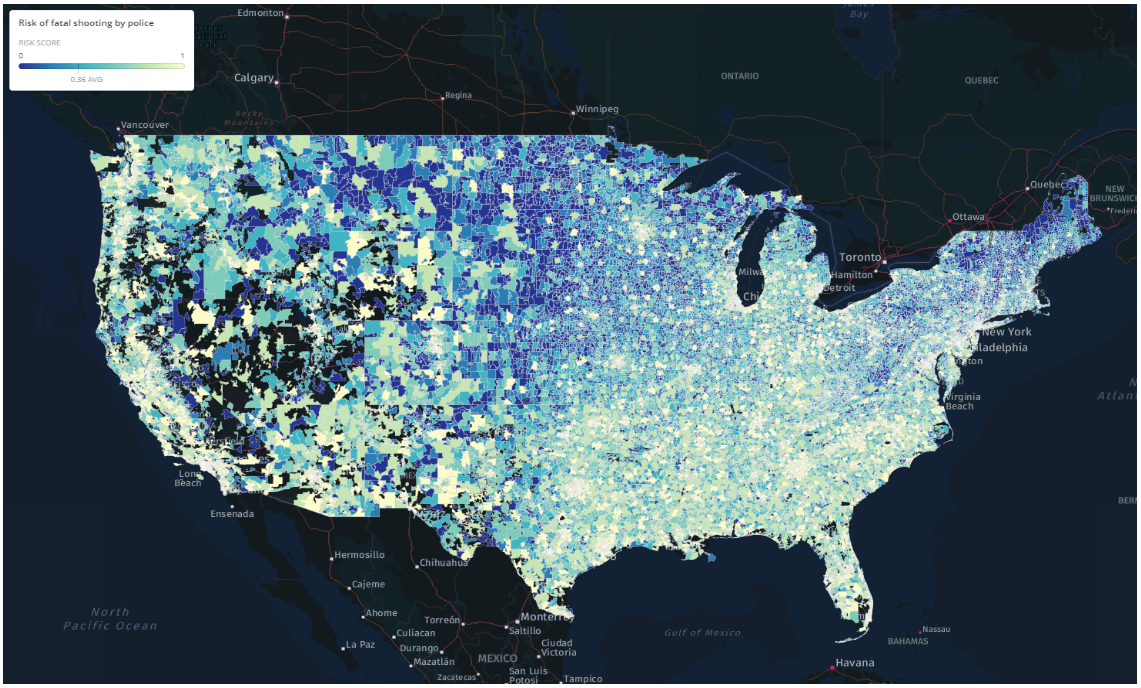

Figure 7: Census only risk score for every available, populated ZCTA in CONUS

Using the census and fatal shootings model, we predict the risk score of a fatal shooting for every populated ZCTA in the US that has census data available. We then aggregate these results together and visualize them in a map for the Continental United States as shown in Figure 7. We also host an interactive and animated web map of these results such that users can hover over an a specific area and view a risk score between 0 and 1: `https://dbsmooth.carto.com/builder/b483b4a5-5672-44dc-b084-203b054ac71d/embed`.

Classifier	**Test Set Accuracy**
Logistic Regression	88.13%

Figure 8: Classifier accuracy results for predicting fatal shootings based on every populated ZCTA available from the census data. This model has hyper parameters tuned via cross validation and accuracy reported on a 20% test data set

Predictor	**Coefficient**
Total population	0.9005
2+ computing devices	0.2205
Black or African American	0.1236
Poverty rate	0.1231
Divorced rate	0.1193
Travel time to work 10-14 min	0.0953
Has some college education rate	0.0693

Figure 9: Top 7 *positive* coefficients in our logistic regression model for census data and fatal shooting data across the US

6.2 Combined Census, FBI Crime, and Fatal Shooting Data

When we combine the census data, FBI crime data, and fatal shooting data, the first approach we take is to perform feature selection on our 188 predictors. We perform a permutation-based faced feature importance using a random forest and we show the top 10 predictors in Table 2.

Similar to our results in Section 6.1, the population predictor dominates compared to the rest. With higher populations, there are more police contacts in general and hence a higher risk for a fatal shooting by a police officer. The second most important feature is robbery rate. This is the number of robberies that occurred in 2020 divided by the total population of the ZCTA. As expected, robbery rate, is significant in predicting fatal police shootings. As robbery rate increases, the number of fatal shootings tend to increase. Robberies are high-tense situations, especially if the perpetrator is caught in the act. In this regard, it is reasonable to see perceive these situations can increase the risk of a police fatal shooting. The third most important feature is city population. This is similar to the total population of the ZCTA (our most important predictor), but is instead the population of the town or city that a ZCTA is in. Given that cities often have multiple ZCTAs, this is a one to many mapping. The next four important predictors are all crime rates: motor vehicle theft rate, arson rate, murder rate, and and violent crime rate. In addition to being important features, they all tend to increase the chance of a fatal shooting by police. Finally, the next two predictors in our important features list are surprising: Race: American Indian and Alaska Native and Race: two or more races. We see from our prior analysis that race plays an important role in predicting fatal shootings, but once we include crime data, race is no longer in the top 3. The proportion of a population that is American Indian or Alaska Native tends to *decrease* the chance of a fatal shooting occurring.

Feature Name	Feature Importance	$\pm$ 1 Standard Deviation
Total population	0.0337	0.0035
Robbery rate	0.0052	0.0009
City population	0.0042	0.0005
Motor vehicle theft rate	0.0031	0.0007
Arson rate	0.0027	0.0006
Murder and non-negligent manslaughter rate	0.0023	0.0008
Violent crime rate	0.0021	0.0006
Race: American Indian and Alaska Native	0.0017	0.0003
Race: two or more races	0.0017	0.0005
Larceny-theft rate	0.0016	0.0007

Table 2: Top 10 features according to permutation-based feature importance. First column is the name of the feature, second column is the assigned feature importance, and last column is $\pm$ 1 standard deviation of the importance value

With regards to model performance, we illustrate our test set accuracy results in Figure 10. Namely, we find that the poorest performing models are logistic regression and boosting, with accuracies of 70.95% and 70.29%, respectively. Specifically, we find that the boosting model overfits to the training set, despite tuning its major hyper parameters. We note that, in terms of accuracy, the best performing models where random forest and neural network. The random forest yields an accuracy of 72.23% while the neural network performs slightly better with an accuracy of 72.28%. Despite the similar accuracy results for both these models, we can “break the tie” by looking further into their performance metrics. Figure 11 illustrates an ROC curve for all four model results. Of particular interest, we note that the neural network curve tends the most toward the top-left, indicative of superior performance across the majority of the thresholds. With this, we conclude that the neural network model yields the best overall performance.

Classifier	Test Set Accuracy
Logistic Regression	70.95%
Random Forest	72.23%
AdaBoost	70.29%
Neural Network	72.28%

Figure 10: Classifier accuracy for predicting fatal shootings based on every populated ZCTA available from FBI law enforcement data. Models have major hyper parameters tuned via cross validation and accuracy reported on a 20% test data set

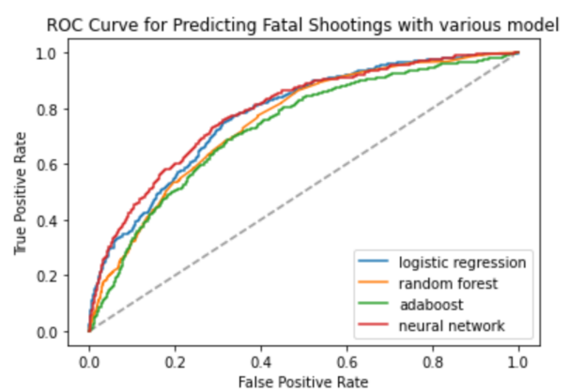

Figure 11: ROC curves for predicting fatal shootings based on every populated ZCTA available from FBI law enforcement data

7 Discussion

7.1 SHAP Analysis

For this project, the interpretation of the predictors and their impact on the risk of fatal shootings are of paramount importance. We not only want a reasonably accurate model to predict fatal shootings, but we also want to understand the relationship between our predictors and our response variable.

To give another view into these relationships, we decide to use SHapley Additive exPlanations (SHAP). Similar to neural networks, this is another technique that was not covered in AC 209a. We calculate the SHAP values for class 1 (having fatal police shootings) for our random forest model. Figure 12 shows the summary plot of the most important predictors from our random forest model (from the original most important 40 features selected) with our permutation-based feature selection. We see that total population, education, certain crime rates, race, and poverty are the top predictors. In the SHAP summary plot, the abscissa corresponds to the SHAP values of the predictors and each row corresponds to one predictor. Each dot in a row represents a single prediction made by the model and the color of the dot represents the value of this feature - red corresponds to a higher value and blue corresponds to a lower value. A positive SHAP value of a feature in a prediction means that this feature makes positive contribution to this sample being predicted as class 1, and the contribution is larger for higher SHAP values. The SHAP summary plot can give us a general sense of how each feature relates to fatal police shootings.

From the first line of Figure 12, we can see that areas with a higher total population have higher SHAP values, which is reasonable because fatal police shootings are a rare human event and so a larger population naturally means a higher probability of the occurrence of fatal police shootings.

Figure 12: SHAP summary plot for most important predictors

Different types of crime rates generally show quite clear SHAP value distributions according to Figure 12. For rape rate, motor vehicle theft rate, violent crime rate, property crime rate, aggravated assault rate, larcency-theft rate, and burglary rate, it can be seen that higher values of these crime rates correspond to higher SHAP values. For arson rate, robbery rate, and non-negligent manslaughter rate, although a low arson, robbery or murder rate does not necessarily correspond to a low or negative SHAP value, a high arson, robbery or murder rate always corresponds to a high and positive SHAP value. This observation means that the six types of crime rates mentioned above are positively correlated to the occurrence of fatal police shootings, which implies that in the areas with poorer public security, police tend to take more powerful actions like shooting or tasering when they are on duty.

Education level also has some notable patterns in the SHAP values. Seven features relevant to education levels are used in our random forest model. It can be seen that, for the feature of the ratio of the population having 10th grade educational level, a higher value of this feature will lead to higher SHAP value in predictions. For features describing the ratio of the population having 1st, 2nd, 4th, 5th, 6th, and 9th grade educational level and less than 1 year of college education level, although sometimes low feature values (blue dots) relate to high SHAP values, high feature values generally correspond to positive and high SHAP values, implying that they are positively correlated to being predicted as class 1. Education levels below college, and especially below the 10th grade, are considered as low education level. Therefore, from the SHAP values it can be seen that a high proportion of population with low education level makes positive contributions to fatal police shootings in an area. Poverty rate is also an important feature related to police violence. It can be clearly observed that a higher poverty rate makes a positive contribution to fatal police shootings.

We also care about how the race distribution of an area contributes to the police violence. The most significant race feature is the rate of American Indian and Alaska Native population, which is positively correlated to fatal police shootings according to Figure 12. On the other hand, the rate of “two races” and “other races” populations are not showing clear patterns in their contributions to fatal police shootings. Finally, the rate of the white population barely contributes to fatal police shootings, which is reasonable because Caucasians form the majority of the United States population and so every area has a high proportion of whites.

7.2 Concluding Remarks

We have consistently seen that total population, crime, education, and poverty all relate to fatal police shootings. The relationship between these predictors and police fatal shootings is complex. There are further data sets such as police misconduct and corruption which would help with the predictions. One limitation of this project is that we were not able to obtain such data at the geographical scale that we needed. As such, we are only looking at the potential victims rather than at the perpetrators (the police). This creates an inherent imbalance in our question. Another limitation of our specific modeling approach is the inherent noise in the data sets we are working with. Given the rarity of fatal police shootings, predicting the occurrence of such events becomes a challenging endeavour.

One of the strengths of this project is that we are able to process vast amounts of data and work with data across the entire US. This provides us with a large set of samples to work with and allows us to explore using neural networks as a model. This new technique performs the best overall with an accuracy of 72.28% and helps illustrate the strength of deep learning compared to our other modeling approaches.

Overall, throughout our analysis, we observe that crime and (lack of) education appear to be prominent in increasing the chance of a fatal police shooting. Furthermore, we find that both of these predictors are related to each other. By improving education, and access to education, it is likely that crime would decrease, and so would fatal shootings by police. On the other side of the coin, police training and education would likely reduce the number of situations where police feel like they need to use deadly force.

8 Future Work

This work can be expanded in a few key ways. First, it would be interesting to explore even larger data sets which would contain more instances of fatal police shootings in the United States. Gaining access to such data would yield a much more balanced data set. Furthermore, finding data pertaining to police officers involved in coercive situations would provide another dimension of analysis. By being able to extract features related to officers’ race, gender, age, and location, it would be possible to learn which types of officers are more likely to be involved in shooting incidents. Furthermore, another direction of future work would involve considering different, more complex models for our classification task. For instance, leveraging a support vector machine could potentially yield fruitful results, due to its robustness and ability to work well in a high-dimension feature space [10]. Finally, we can consider more complex classification tasks to test the strengths and weaknesses of our modeling approach. Instead of predicting whether a fatal shooting occurs or not in any given data sample, we could attempt to predict the race of the individual involved in the incident. Performing such a multi-class classification problem would enable us to further understand the patterns or characteristics of the victims involved in cases of overly coercive police interactions.

References

[1] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[2] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.

[3] D. R. Cox. The regression analysis of binary sequences. *Journal of the Royal Statistical Society. Series B (Methodological)*, 20(2):215–242, 1958.

[4] Maskaly J. Donner. A theoretical integration of social learning theory with terror management theory: Towards an explanation of police shootings of unarmed suspects. *American Journal of Criminal Justice*, 40:205–224, 2015.

[5] Yoav Freund and Robert E Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1):119–139, 1997.

[6] Jr Fryer, Roland G. An empirical analysis of racial differences in police use of force. Working Paper 22399, National Bureau of Economic Research, July 2016.

[7] M.W Gardner and S.R Dorling. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric Environment*, 32(14):2627–2636, 1998.

[8] Tin Kam Ho. Random decision forests. In *Proceedings of 3rd International Conference on Document Analysis and Recognition*, volume 1, pages 278–282 vol.1, 1995.

[9] Scott M Lundberg, Gabriel G Erion, and Su-In Lee. Consistent individualized feature attribution for tree ensembles. *arXiv preprint arXiv:1802.03888*, 2018.

[10] Kevin P. Murphy. *Machine Learning: A Probabilistic Perspective*. The MIT Press, 2012.

[11] Franco Scarselli and Ah Chung Tsoi. Universal approximation using feedforward neural networks: A survey of some existing methods, and some new results. *Neural Networks*, 11(1):15–37, 1998.

[12] Yuan Wang and Yangxin Fan. Us fatal police shooting analysis and prediction, 2021.