Aryan Deorah and Chris Goodhart

Mr. Writer

AOS Junior Research

**Background Research on Machine Learning and the Relationship Between Gun Laws and Gun Violence**

Gun violence has been a constant threat throughout American history and has become a threat to the well being of many Americans. Early gun laws date all the way back to the Constitution, with the second amendment being the right to bear arms. Ever since the founding of the country, the number of guns per population has increased, while gun violence rates have varied. Gun violence rates spiked in the early-1980’s and 1990’s during crime surges, then receded until the late 2000s, when gun violence rates began to increase again and are still increasing (Cohn et al., 2013). Though gun ownership per household has decreased over the last 20 years from 51% in 1994 to 36% in 2017, number of guns per population has increased from 90 guns per 100 people in 1996 to 113 guns per 100 people in 2013. The number of guns in America surpassed the number of people in 2009, when the population was 310 million (Ingraham, 2015; Ingraham, 2016). In essence, the trends show that less people own guns, but there are more guns and more victims of gun violence overall.

These trends are only seen in few other countries outside of the U.S.; as of 2007, half of the civilian owned guns in the world are in the US, over 6 times that of any other country in the world. As a result, Americans own the most guns per population, nearly double that of any other country in the world. America also has the greatest gun violence rates of all high-income nations, with American per-capita homicide rates nearly 25.2 times the average of high-income nations. Because of all of the gun violence, the guns have been more hotly debated over the last few years. Horrid mass shootings over the recent years, such as that of Sandy Hook Elementary School in 2012 that claimed 26 lives, many of which were children, have received increased levels of media coverage and have fueled further debates and policy changes. Though no major policy changes have occurred on the federal level in recent years, many states and municipalities have made policy changes regarding various gun laws. One example is a Florida gun law, passed following the Parkland school shooting that killed 17 people, that raises the minimum age to purchase guns and bans bump stocks. (Fox, 2018; Astor, 2018). Another example is the “red flag” law passed in Rhode Island that aims to take guns away from violence-prone people. Similar laws are being considered by dozens of states across the country (Seelye and Bidgood, 2018).

In 2015, gun violence claimed 36,252 lives in the United States, approximately 11.3 lives per 100,000 people; This number increased to 11.8 per 100,000 in 2016. This number varies greatly from state to state, with gun violence rates in Massachusetts at 3.4 per 100,000 and Alaska at 23.3 per 100,000 in 2016 (CDC, 2017; CDC, 2018). In addition to that, gun ownership and gun violence tend to correlate, as states with higher rates of gun violence tend to have higher gun ownership rates (Siegel, Ross, and King, 2013). These trends clearly show room for improvement. There must be differences between Massachusetts and Alaska, likely in gun legislature specifically, that lead to the vast gap in gun homicide rates.

There are various categories of gun laws in states and municipalities that may affect gun violence. These categories include right to carry laws, assault rifle and machine gun bans, mandatory background checks, gun permit requirements, and violence criminal gun bans. These laws could affect gun violence; for example, if violent criminal gun bans are in place, then potentially less criminals will get guns, and there will be less victims of gun violence.

Machine learning is a math modeling system in which programed models can teach themselves how to solve an endless amount of problems in today’s world. These models can learn how to make decisions as well as analyze and process data on a scale and proficiency that no other mathematical or programmed model can. Machine learning will greatly benefit our research because social science has many inputs and controls for the data, and there is no other form of statistics that can take these various inputs into account better than machine learning.

Tensorflow is an open source code framework that works using neural networks and can be applied to machine learning and non-machine learning projects alike. Essentially, neural networks bring data through multiple layers, each one with inputs and outputs. Each layer has several “neurons”, which represent computations, and certain inputs activate certain neurons, which provide outputs that move to serve as the inputs for the next layer. Neural networks are used for a variety of purposes, from diagnosing Alzheimer’s using medical imaging to sorting photographs using digitized files. Many major companies that use data science, primarily Google, have applied Tensorflow to their work (Rampasek and Goldenberg, 2016).

There are two major types of machine learning: supervised and unsupervised learning. The main difference is whether the machine is taught externally or internally; supervised learning systems are trained externally by a data scientist with expected outputs, while unsupervised learning systems are trained internally by the machine itself. Supervised systems are relatively simpler and more accurate, while unsupervised systems are very complicated and tend to yield inaccurate results when dealing with smaller sets of data. Unsupervised systems are considered to be closer to artificial intelligence because the machine can train itself without external interference (Schmidhuber, 2015; Castle, 2017).

There are four major types of machine learning models: regression, dimensionality reduction, classification, and clustering. Regressions are meant for predicting continuous values and quantities and will be used the most in the research because gun violence death rates are values. Regressions use equations, such as linear regressions, and mathematical relationships to predict quantities. Dimensionality reduction is used as a prediction model based on reducing the number of dimensions involved in the prediction. Dimensionality reduction has many practical applications that involve high dimensional data with hundreds of dimensions, such as image classification. In this case, each potential color of each pixel would be its own dimension. However, since we are not intending to have that many dimensions, dimensionality reduction will not have an application to our project. Classification is used to predict non-continuous values. The most common form of classification is binary classification, that classifies predictions into 0s and 1s. Since it seems that classification will not give us the quantities we are looking for, it will not have an application to our project as of right now. Clustering uses various equations to “cluster” data together and divide data points into different groups based on their values. Since clustering does not seem to output anything useful to the project, it will not be used in our research.

Research regarding machine learning, gun violence, and social sciences over the last 5 years has gone in various different directions. One example of a gun violence study is the Siegel et. al (2013) paper “The Relationship Between Gun Ownership and Firearm Homicide Rates in the United States, 1981–2010.” In the study, Siegel et. al collected data from state to state, as well as a number of other geographic, demographic, and economic factors present in each state, such as income, race, and unemployment rates. These factors were used to adjust the data. They then used a generalized estimating equation (GEE) approach, which created a negative binomial expression for the model of the data. They had discovered that states with significantly greater gun ownership rates also had higher rates of gun violence. In addition to that, they found various factors, such as crime and incarceration rates, that had the greatest effect and tended to increase gun violence (Siegel et al., 2013).

An example of a gun violence study that showed the relationship between gun laws in one area and gun violence in another is titled “Cross-Border Spillover: U.S. Gun Laws and Violence in Mexico.” Dube et. al (2013) focused on the 2004 expiration of the U.S. Federal Assault Weapons Ban and how it increased gun violence in certain Mexican communities near the U.S.-Mexico border. According to the study, gun violence rates increased dramatically, by more than 60% in Mexican communities near the Texas, New Mexico, and Arizona border ports as a result of the 2004 expiration. Unlike the other border states, California had restrictions in place regarding assault weapons. This caused gun violence rates to increase 60% more in Mexican communities in the vicinity of non-Californian border ports than communities in the vicinity of California’s border ports. This study was an example of how gun laws from one location can directly affect gun violence in another location (Dube, Dube, and Garcia-Ponce, 2013).

An example of deep learning being applied to behavioral trends is “Estimating Treatment Effect Heterogeneity in Randomized Program Evaluation.” The study focuses on the effect that different social programs and medical treatments have on social behavior. The study used various estimation algorithms to determine how much a given social program or medical program, such as transitional housing and job-training programs, had on a given population. For example, various positive and negative treatment effects of a job training program were quantified to show how much that social program helped or hurt a given demographic of people, such as race, age, and gender. This concept can be applied to determining the positive or negative effects of certain gun laws on different types of gun violence (Imai and Ratkovic, 2013).

There are various problems facing research regarding gun violence, social sciences, and machine learning. As for gun violence, the CDC was banned by Congress from researching causes and solutions with gun violence in 1996. This has multiple implications, including the fact that solely private institutions can conduct gun violence research, which reduces the accuracy of available gun violence studies. In addition to this, controversies surrounding gun violence research can make it hard to receive funding and convince others of the legitimacy of the findings. These controversies also lead to increased bias in the research. Furthermore, underreporting has been shown to impact the validity of the results of past studies. Social sciences are hard to model and are highly unpredictable, meaning that even the best of models are only guesses. This unpredictability makes it difficult for the model to understand the patterns in social statistics. As for machine learning, machine learning systems are complicated, and unsupervised systems require enormous amounts of data. Without these large quantities of data, unsupervised models will provide inaccurate results. Overall, there are multiple concerns to consider regarding the data and the models in these fields of research.

Our research differs from past studies for a number of reasons. Firstly, it is focused on the correlation between multiple different gun laws and types of gun violence rates within a municipality. Most past research focuses in on a variety of factors that can affect gun violence, while our research focuses only on the factors that lawmakers can change. In addition to this, most gun violence research is done on the national level, while this is on the state level, so there will be less potential errors and biases because a national study fails to account for multitudinous state factors that differ between states such as state legislature. Next, our research is controlled only by economic factors and external gun law factors, whereas most other research includes other factors such as race, unemployment rate, violent crime rate, and age. This difference is significant because a decent number of relevant factors will be taken in, but not too many that there is data bias by the people conducting the research. These relevant factors include median income and general crime rates. Along with that, there are few known applications of machine learning to gun violence. This is important because this study could provide a more complete picture of the relation between gun violence and gun laws. Finally, the training data used for the model will be provided using a stringency algorithm, which outputs the strength of the aforementioned categories or subcategories of gun laws for a given state. This is important because many studies in the social science categories have to manipulate certain data to get expected outputs or have to use unsupervised systems, so this project will be unique in how it gets its expected outputs.

The first major step to this study will be collecting data. The laws of states, initially five states, will be categorized. As time goes on and the study progresses, laws from more states will be collected and categorized for use in the project. The rates of gun violence will also be collected from the examined states. The data will be gathered from a variety of journals of epidemiology, public health journals, and medical journals. Next, the data will be modified into a format that is easy to use for the model. The data will be visualized through graphs and charts so as to identify preliminary patterns in the data and locate features with large amounts of missing values. If there are missing values in a feature, they can be imputed. However, if too many values are missing, the column will be dropped. The data will control for factors such as GDP, unemployment, and external gun factors.

The unique aspect of this project will be training machine learning models on the law analysis. Several different regression models will be trained to take law analysis as inputs and output predicted rates of gun violence. Then, a function for measuring the performance of the models will be applied to isolate the best performing model. This model will be run to determine which changes in gun laws have the best effect on gun violence.

As a second part to this research, a stringency algorithm will be created to take in law characteristics as inputs and give a numerical value of the strictness of the law as an output. This algorithm will be based off of other stringency algorithms used in past research. The algorithm will be applied to the laws that were tested. The models will be retrained on the stringency scores rather than law analysis. The performances of the regression models will be evaluated and compared to the performances of the previously created models.

**References**

Astor, M. (2018, March 08). Florida Gun Bill: What's in It, and What Isn't. Retrieved June 10,

2018, from https://www.nytimes.com/2018/03/08/us/florida-gun-bill.html

Castle, N. (2017, July 13). Supervised vs. Unsupervised Machine Learning. Retrieved May 15,

2018, From

https://www.datascience.com/blog/supervised-and-unsupervised-machine-learning-algorithms

Cohn, D., Taylor, P., Lopez, M. H., Gallagher, C. A., Parker, K., & Maass, K. T. (2013, May 07).

Gun Homicide Rate Down 49% Since 1993 Peak; Public Unaware. Retrieved May 15,

2018, from http://www.pewsocialtrends.org/2013/05/07/gun-homicide-rate-down-49-since-1993-peak-public-unaware/

Dube, A., Dube, O., & García-Ponce, O. (2013). Cross-Border Spillover: U.S. Gun Laws and

Violence in Mexico. *American Political Science Review,* *107*(03), 397-417.

doi:10.1017/s0003055413000178

Fox, K. (2018, March 09). America's gun culture vs. the world in 5 charts. Retrieved May 15,

2018, from https://www.cnn.com/2017/10/03/americas/us-gun-statistics/index.html

Imai, K., & Ratkovic, M. (2013). Estimating treatment effect heterogeneity in randomized

program evaluation. *The Annals of Applied Statistics,* *7*(1), 443-470.

doi:10.1214/12-aoas593

Ingraham, C. (2015, October 05). There are now more guns than people in the United States.

Retrieved May 15, 2018, from

https://www.washingtonpost.com/news/wonk/wp/2015/10/05/guns-in-the-united-states-one-for-every-man-woman-and-child-and-then-some/?noredirect=on&utm\_term=.b63d442e95d8

Ingraham, C. (2016, June 29). American gun ownership drops to lowest in nearly 40 years.

Retrieved May 15, 2018, from

https://www.washingtonpost.com/news/wonk/wp/2016/06/29/american-gun-ownership-is-now-at-a-30-year-low/?utm\_term=.b9d3e013bbdc

National Center for Health Statistics. (2017, May 03). Retrieved May 15, 2018, from

https://www.cdc.gov/nchs/fastats/injury.htm

National Center for Health Statistics. (2018, January 10). Retrieved May 15, 2018, from

https://www.cdc.gov/nchs/pressroom/sosmap/firearm\_mortality/firearm.htm

Rampasek, L., & Goldenberg, A. (2016). TensorFlow: Biology’s Gateway to Deep Learning?

*Cell Systems,* *2*(1), 12-14. doi:10.1016/j.cels.2016.01.009

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks,* *61*,

85-117. doi:10.1016/j.neunet.2014.09.003

Seelye, K. Q., & Bidgood, J. (2018, February 27). What Are States Doing About Gun Violence

After the Florida Shooting? Retrieved June 10, 2018, from https://www.nytimes.com/2018/02/26/us/gun-control-laws.html

Siegel, M., Ross, C. S., & King, C. (2013). The Relationship Between Gun Ownership and

Firearm Homicide Rates in the United States, 1981–2010. American Journal of Public Health, 103(11), 2098-2105. doi:10.2105/ajph.2013.301409