

Análise da violência envolvendo armas de fogo nos Estados Unidos

Milton Vasconcelos da Gama Neto

CIn, Centro de Informática UFPE

Universidade Federal de Pernambuco

Recife, Brasil

Abstract

An analysis about violence involving fire gun in USA. From the public data provided by Gun Violence Archive organization, where were recorded all incidents with weapons. It was an analysis relating poverty and carrying of weapons legislation. Others two datasets were used, a USA census of 2015 and another with information about firearm laws in each state. In this paper, the relevance of poverty and the release of weapons was confirmed for the increase in the number of deaths involving weapons. With these attributes was built a machine learning model to predict the level of death caused by weapons in a state. The results show that these indicators are not the only ones that determine the level of death in a state accurately but they are relevant.

Resumo

Uma análise sobre a violência envolvendo armas de fogo nos Estados Unidos. A partir dos dados públicos fornecidos pela organização Gun Violence Archive, onde estão registrados todos incidentes com armas. Foi feita uma análise relacionando a pobreza e a legislação do porte de armas. Outras duas bases de dados foram utilizadas, uma do censo de 2015 dos EUA e outra com as informações sobre as leis do porte de arma em cada estado. Neste trabalho foi confirmada a relevância da pobreza e das liberação das armas para o aumento do número de mortes envolvendo armas. A partir desses atributos foi construído um modelo de aprendizagem de máquina que faz a predição do nível de mortes com armas em um estado. Os resultados apontam que esses não são os únicos indicadores para determinar o índice de morte num estado com precisão, mas que eles são relevantes.

1. Introdução

O tema da liberação e armas é algo bastante polêmico e sempre está em debate. Devido a sua importância, foi realizada uma análise do impacto dessa liberação com os incidentes com armas e as mortes decorrendo da utilização desta. Como nos Estados Unidos as leis para o porte de arma variam de acordo com o estado, foi realizada uma análise considerando os 50 estados norte americanos com suas respectivas leis.

Outro fator analisado foi o índice de pobreza em cada estado. Partindo da hipótese que lugares mais pobres ocorre mais crimes, violência e eventualmente armas de fogo estão envolvidas. Então, a liberação de

arma de fogo num estado mais pobre tem um peso diferente, porém esse não é o único fator que ocasionará incidentes e fatalidades.

1.1. Datasets

Foram utilizadas três base de dados e realizado um cruzamento dos dados com os estados, atributo comum em todas.

1.1.1. Gun Violence

Base de dados da organização Gun Violence Archive com aproximadamente 240 mil registros. Os dados vão de 2013 a 2018. Foi realizado um pré processamento para selecionar os atributos relevantes para essa pesquisa e remoção dos atributos com muitos dados faltantes ou irrelevantes para análise, como id do registro na base original e links. As informações tem uma granularidade alta quando se trata de localização, para esse estudo foram agrupadas as informações por estado.

1.1.2. Census 2015

Dados do censo norte americano do ano de 2015, com informações da população para cada cidade. Assim como na base de dados anterior, as informações foram agrupadas por estado. Foi utilizado apenas a quantidade total da população e o indicador de pobreza.

1.1.3. Firearm

Base de dados com as informações das leis sobre o porte de arma nos Estados Unidos de 1991 até 2017. As leis utilizadas foram as de 2015, por ser o mesmo ano do censo e não ter grandes mudanças entre 2013 e 2018. Foram codificadas 133 atributos que representam uma determinada restrição, esses atributos são distribuídos em 14 categorias.

2. Análise dos dados

Após realizar um pré-processamento e preparar as bases de dados para a utilização, deixando todas informações agrupadas por estado e normalizadas, foi realizada uma análise para validar as hipóteses levantadas. Outra consideração importante na preparação dos dados, foi a sumarização de algumas informações mas ponderando pela população, ou seja, o índice de pobreza num estado é a soma de todos os índices de pobreza das cidades pertencentes aquele estado utilizando diferentes pesos de acordo com a população da cidade.

2.1. Relação dos incidentes e mortes envolvendo armas e o nível de pobreza

Muitos fatores são considerados para o aumento do número de crimes num determinado local. Por exemplo, desigualdade social, índice de desemprego, segurança, educação, entre outros. Os dados utilizados são sobre violência com armas em geral, então, foi analisado se existe alguma relação entre esses números e a pobreza por estado.

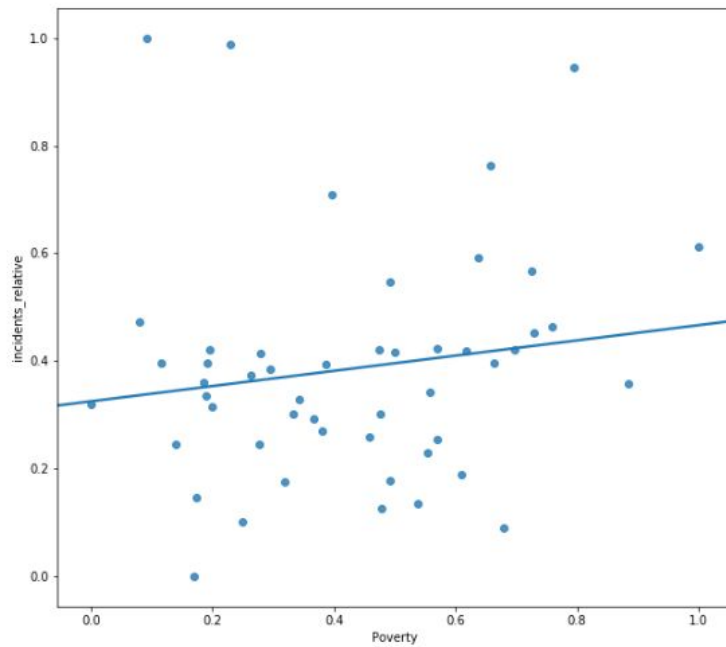


Fig. 1. Distribuição do número de incidentes por pobreza.

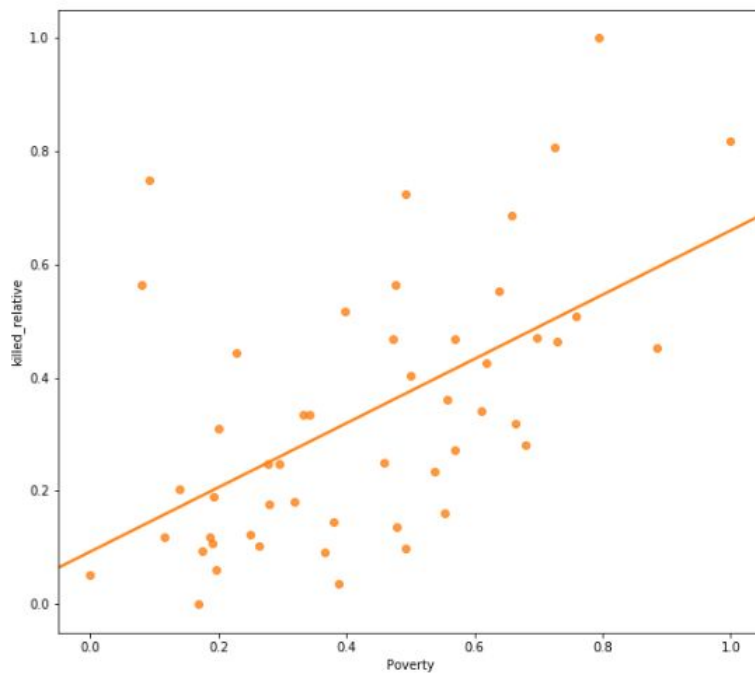


Fig.2. Distribuição da quantidade de mortos por pobreza

As variáveis utilizadas foram: Poverty, incidents_relative, killed_relative, que representam a pobreza, quantidade de incidentes relativo à população e número de mortos relativo à população, respectivamente.

A quantidade de mortos tem um aumento maior de acordo com o aumento da pobreza (Fig. 1) comparado com a quantidade de incidentes (Fig. 2).

Com isso, está confirmada a primeira hipótese da relação que existe com a pobreza. Outra comprovação disso foi feita através da correlação de Pearson (Fig. 3).

	incidents_relative	killed_relative	Poverty
incidents_relative	1.000000	0.725088	0.152392
killed_relative	0.725088	1.000000	0.561863
Poverty	0.152392	0.561863	1.000000

Fig.3. Correlação de Pearson

2.2. Relação entre a permissão de do porte de armas com o número de mortes

A segunda hipótese era que estados mais flexíveis com o porte de arma teriam índices de mortes maiores. Existem diversos tipos de permissão, aqui consideramos o somatório de todos. Onde um número alto indica uma liberação maior.

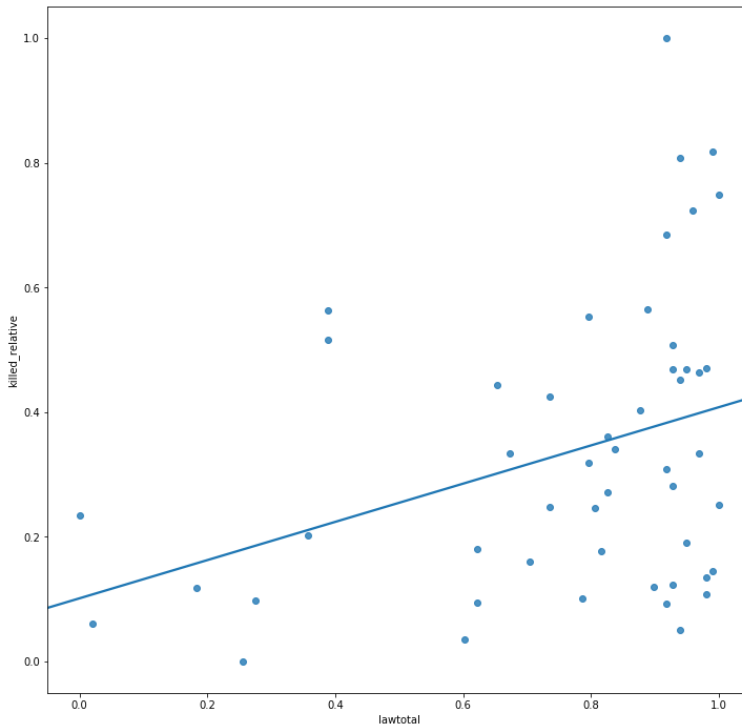


Fig. 4. Distribuição do número de mortos pela quantidade de permissões

O gráfico mostra uma concentração maior de estados com uma quantidade maior de permissões, porém, confirma o crescimento do número de mortes com o aumento das permissões. Vale salientar que apesar da curva mostrar o crescimento, muitos exemplos não se comportam seguindo fielmente a reta, porém

é notório o crescimento. A correlação de Pearson para essas duas variáveis é 0.346.

A comparação com o total de permissões é genérico, algumas leis analisadas separadamente tem uma correlação maior com o número de mortes, por exemplo, as variáveis *nosyg* (0.442) e *mcdvdating* (0.376). As variáveis indicam que *o uso de força letal não pode ser um primeiro recurso em público* e que *a lei proíbe a posse de armas por pessoas condenadas por um delito de violência doméstica*, respectivamente.

3. Predição

Após confirmar que existe uma relação entre as mortes causada pelo uso de armas de fogo em um estado, seu índice de pobreza e sua legislação a respeito do porte de arma, foi criado um classificador para determinar qual seria o índice de um estado apenas com esses indicadores citados acima.

3.1. Alvo

A quantidade de mortes num estado é um valor numérico entre um e zero, pois foi normalizado. Foi criada classes com intervalos desse valor para ser utilizada como alvo do modelo. As classes são:

- L (Low) - Para o intervalo $0 \leq x < 0.33$
- M (Medium) - Para o intervalo $0.33 \leq x < 0.66$
- H (High) - Para o intervalo $0.66 < x \leq 1$

3.2. Cross Validation e Oversampling

Como a quantidade de exemplos é a quantidade de estados norte americanos, temos uma limitação de dados. Também temos um problema de desbalanceamento, pois as classes com número alto de mortes (H) tem poucos exemplos comparados às demais.

Para contornar estes problemas, foi utilizada uma validação cruzada (Cross Validation) para aumentar a quantidade de exemplos utilizados no teste e ter um valor mais preciso da avaliação do modelo.

E foi utilizada a técnica SMOTE para gerar mais exemplos (Oversampling) das classes minoritárias.

3.3. Atributos

Os atributos utilizados no modelo foram o índice de pobreza do estado, as 25 leis com maior correlação com a quantidade de mortes e o total de permissão das 14 categorias de permissões. A ideia de agrupar por categoria e utilizar a soma das permissões é para reduzir a dimensionalidade do problema e realizar a criação de novas features.

3.4. Random Forest

O modelo utilizado foi Random Forest, com a configuração padrão da biblioteca Scikit-Learn e obteve uma acurácia de 55%.

4. Resultados e Discussões

Apesar da acurácia não ter sido alta, ainda assim ela atinge um valor maior o aleatório (33.3%, por serem três classes). O que comprova o fato das leis e índice de pobreza interferirem no percentual de mortes com a utilização de armas de fogo, como os gráficos e correlação mostraram anteriormente. Porém, esse não são os únicos fatores para determinar esse número e a quantidade pequena de dados dificulta que o modelo crie uma representação que generalize bem o problema.

5. Conclusão e trabalhos futuros

Foi possível validar as hipóteses da interferência da pobreza e da liberação do porte de armas no índice de mortes em um estado. Esses não são os únicos indicadores que devem ser utilizados para determinar um valor exato, mas mostram uma relevância. Esse estudo enfatiza o cuidado que deve ser tomado quando se pensa na liberação de armas, vários fatores devem ser levados em consideração e tomados como referência.

Para trabalhos futuros, uma análise mais profunda das leis individualmente e considerar outros aspectos que podem ser considerados para o crescimento (ou diminuição) do número de mortes causados por armas de fogo. E também aprimorar o modelo utilizado, executando experimentos de aprendizagem de máquina para encontrar o melhor modelo e os parâmetros para esse problema.

References

Provost, F., & Fawcett, T. (2013). *Data Science for Business: What you need to know about data mining and data-analytic thinking*. " O'Reilly Media, Inc."