

Universidade Federal de Pernambuco - Centro de Informática

Ciência da Computação

IF706 – Introdução a Ciência dos Dados

Prof.: Renato Vimieiro

Aluno: Milton Vasconcelos da Gama Neto

Título: Análise de violência envolvendo a utilização de armas de fogo nos Estados Unidos

Proposta:

Realizar uma análise da violência com armas de fogo nos Estados Unidos, com objetivo de obter informações sobre os fatores que podem interferir nesses incidentes. Para isso, será utilizada uma base de dados com aproximadamente 240 mil registros relacionados com esse tipo de violência, os registros vão de 2013 a 2018. Os dados são da organização Gun Violence Archive (GVA), uma organização sem fins lucrativos.

Cada registro, tem informação do lugar do ocorrido, número de feridos e mortos, tipo da arma, características do incidente, entre outros. Como nos EUA a legislação sobre o porte de armas varia de acordo com os estados, os dados serão cruzados com outra base, que tem as informações das permissões sobre esse porte. Outro fator, é que existem tipos de permissões diferentes, ou seja, não é apenas um “pode ou não pode” para cada estado, mas sim com diferentes restrições, que também são consideradas na base. A intenção de cruzar esses dados, é verificar a quantidade de registros de violência por liberação do porte de arma. Analisar se existe alguma correlação entre esses dados. E validar a hipótese que essa liberação ocasiona um aumento no número de incidentes.

Vale ressaltar, que essa liberação do porte de armas não é o único fator que pode interferir, e talvez ele sozinho possa ser pouco relevante. Outros dados devem ser levados em consideração. Então, também será analisado a população, questões econômicas e sociais por estado. Ou seja, outra hipótese, seria o grande número de incidentes onde existe a liberação de armas e uma desigualdade social/pobreza.

Após algumas análises e tratamento dos dados, será criado um classificador para identificar o nível da quantidade de mortes (baixo, médio ou alto) em um estado causado pela utilização de armas de fogo. Os atributos utilizados serão informações de pobreza e legislação sobre a permissão para utilizar armas de fogo.

Etapas:

- Primeiro será realizado um pré-processamento nas bases. Para tratamento de dados faltantes, remoção de atributos que não serão utilizados, ajustes nos tipos (para datas e categorias, por exemplo).
- Criação de novas features. Por exemplo, nos dados de Gun Violence, um dos campos é a característica do incidente, as informações estão armazenadas em uma String com separadores, mas não é algo muito claro. Então será realizado uma análise e separação dessas informações, passando para uma estrutura melhor. Também será realizada sumarização dos dados para cada estado, pois algumas bases tem registros separados por cidades.
- Cruzamento dos dados das diferentes bases.
- Gráficos com número de incidentes por estado e por número de mortos.
- Visualização das informações referentes a pobreza nos estados.
- Análise da relação entre incidentes envolvendo armas e as legislações de permissão. Visualização de acordo com o “tipo de permissão” e correlação entre esses dados.
- Análise da relação entre incidentes envolvendo armas no estado e seus respectivos índices de pobreza. Visualização de ambos os dados por estado e correlação entre eles.
- Criação de classificador para identificar o nível da quantidade de mortes causadas pelo uso de armas em um estado.
- Apresentar história dos dados, separando as principais análises numa ordem que faça sentido, tenha relevância para verificação das hipóteses e também importância nesse contexto.

Datasets:

- Gun Violence:
 - <https://www.kaggle.com/jameslko/gun-violence-data/>
- Firearms Provisions in US States
 - <https://www.kaggle.com/jboysen/state-firearms>
- US Census Demographic Data
 - <https://www.kaggle.com/muonneutrino/us-census-demographic-data>