Group 17

Nimal Kumarr Arunkumar (801317922)

Samhitha Mudiam (801329504)

Hrushi Pappuri (801060321)

Tanuj Darla (801316530)

Shravan Naidu Gollu (801318790)

Big Data Analytics for Competitive Advantage

ITCS 6100

Deliverable 1

Team Members

Nimal Kumarr Arunkumar (801317922)

Samhitha Mudiam (801329504)

Hrushi Pappuri (801060321)

Tanuj Darla (801316530)

Shravan Naidu Gollu (801318790)

Repository and file storage

GitHub repository :

https://github.com/darlatanuj10/Big_Data_project

Google Drive link :

https://drive.google.com/drive/folders/1oTZcGHwaM4rpUTlRjLQMYEjOGVE5BV_v?usp=share_link

Communication plan to include project artifact repository

This communication plan outlines how we will communicate about our project. This communication plan details how and where we will communicate about our project artifacts during our group project.

To distribute project artifacts, we will use the following communication channels:

- WhatsApp: For real-time messaging and file sharing.
- Google Drive is a service for storing and sharing project documents.
- For virtual meetings and screen-sharing sessions, use Zoom.

Repository of Artifacts: Google Drive will serve as our central store for all project artifacts, including:

- Other Documentation
- Code Files
- Datasets

At the following periods, we will communicate about project artifacts:

- Team meetings once a week: We will review project progress, any difficulties or risks, and changes to project artifacts.
- Discussion Hours: We will setup other doubt and meeting sessions when an additional help for a team member is needed.

Domain

This domain focuses on analyzing and building predictive models using a gun violence dataset containing information on incidents in the United States from January 2013 to March 2020. Preprocessing steps include data cleaning, feature engineering, data transformation, and feature selection. Various machine learning models can be applied, such as decision trees and random forests, gradient boosting machines, and neural networks, classification methods to achieve good performance metrics. Metrics such as precision, recall, F1-score, and the Area Under the Receiver Operating Characteristic (ROC) curve should be used to measure model performance as they provide more informative results than accuracy alone for imbalanced datasets. By using Python 3.x and necessary libraries like pandas, numpy, matplotlib, seaborn, and scikit-learn, analysts can gain insights into the patterns and factors contributing to gun violence incidents and build models to predict future incidents.

Gun Violence Data

Data :
https://www.kaggle.com/datasets/jameslko/gun-violence-data
https://www.kaggle.com/datasets/konivat/us-gun-violence-archive-2014
https://www.gunviolencearchive.org

Business Problem or Opportunity

One of the primary business problems that can be addressed with this dataset is to identify the factors associated with gun violence incidents. This includes understanding the demographics of the victims and perpetrators, the types of weapons used, the location and timing of incidents, and other key variables that may be relevant. By analyzing these factors, law enforcement agencies, policymakers, and community organizations can develop targeted interventions to prevent gun violence and improve public safety.

Another business opportunity that can be identified with this dataset is to explore the impact of gun control policies on gun violence. This includes analyzing the effectiveness of existing policies, identifying gaps in policy implementation, and exploring potential policy solutions to reduce gun violence. This information can inform policy development and implementation at the local, state, and national levels.

Moreover, the dataset can be used to identify high-risk areas and populations for gun violence and develop interventions that address the root causes of gun violence, such as poverty, unemployment, and mental health issues. By targeting these factors, stakeholders can reduce the incidence of gun violence and improve the health and safety of communities.

In summary, the gun violence dataset presents a significant business opportunity to better understand the complex issue of gun violence, develop evidence-based interventions, and inform policy decisions. By analyzing this data, stakeholders can work together to reduce the harm caused by gun violence and promote public safety.

Research Objectives

This project's aim is to examine a comprehensive dataset of gun violence occurrences that took place in the US between January 2013 and March 2020. Finding major patterns and trends that can be used to pinpoint the key causes of these accidents is the goal. A further objective of the project is to develop a machine learning model that can precisely forecast the probability of gun violence occurrences in various situations and environments. Utilizing the proper evaluation metrics, the generated model's performance will be assessed. Finally, the initiative aims to offer data-driven suggestions to legislators, law enforcement groups, and community organizations to implement focused interventions and policies to lower gun violence events.

Questions

1) What are the most typical patterns and trends in gun violence instances across the United States from 2013 to 2020?

2) Which aspects (such as place, time, socioeconomic variables, etc.) have the biggest bearing on the frequency of gun violence incidents?

3) Can the discovered factors be used to predict gun violence events using machine learning models like logistic regression, decision trees, random forests, gradient boosting machines, and neural networks?

4) Can gun violence occurrences be accurately predicted using the given parameters using machine learning models such as logistic regression, decision trees, random forests, gradient boosting machines, and neural networks?

5) How does the predictive performance of the various machine learning models compare when evaluated using metrics like precision, recall, F1-score, and the Area Under the Receiver Operating Characteristic (ROC) curve?

6) What practical lessons can be drawn from the data to help policymakers, law enforcement groups, and community organizations create focused strategies for stopping and intervening in gun violence incidents?

7) How did the economic crisis change in the gun violence statistics in the United States ?

Reference

- https://www.aafp.org/about/policies/all/gun-violence.html
- https://ucr.fbi.gov/crime-in-the-u.s/2019/crime-in-the-u.s.-2019
- Jaffe, S. (2018). Gun violence research in the USA: the CDC's impasse. The Lancet (British Edition), 391(10139), 2487–2488. https://doi.org/10.1016/S0140-6736(18)31426-0
- https://www.jec.senate.gov/public/_cache/files/69fcc319-b3c9-46ff-b5a6-8666576075fe/the-economic-toll-of-gun-violence-final.pdf
- https://www.w3schools.com/python/matplotlib_pyplot.asp