Texas A&M University-Commerce

Master of Computer and Information Science

CSCI 526.01W Database Systems

Gun Violence Hotspots Predictive Model

Instructor: Dr. Lee

Submitted by: Megha Shrestha

CWID: 50320145

12/03/2023

Total Pages: 20

ABSTRACT

This term paper addresses a sensitive issue of gun violence in the United States by proposing a robust database system for a sophisticated machine learning model that predicts the hotspot for possible incidents. The information system aims to be a centralized data source for various kinds of data such as details of incidents, offenders, victims, demographic data, and geospatial data. The paper discusses the utilization of a robust dataset to train a machine learning model - CNN, to recognize complex patterns within the data to identify the hotspots. This paper also highlights the consideration of ethics and privacy protocols that ensure the handling of sensitive information. The paper outlines the database design, implementation of a system, its challengers, ethical guidelines, evaluation metrics, and ways for its improvement. There is an emphasis on the crucial role of the database in strategic decision-making, public safety discussion, and future enhancement. This paper discusses a holistic approach to reducing and attempting to eliminate gun violence

Introduction

In recent times, there has been a spike in gun violence in the United States. Every other day there is news related to gun violence. Most of the time the coverage is limited to local news stations based on where the incident happened. Sometimes depending on the impact of incidents can also gather national coverage and a lot of media attention. The total number of deaths from gun-related injuries in the U.S. in 2021 was 48,830 (Gramlich, 2023). Most of the incidents that involve guns result in deaths and severe injuries. While major initiation must include changes in the legislation to make stricter laws for guns, a machine learning program can be created to construct a prediction mode. The predictive model research aims to use sophisticated machine learning techniques to proactively identify possible hotspots for gun violence by utilizing data gathered from multiple sources.

Various prior studies identify that numerous characteristics, including age, sex, race, economic and social disadvantage, place of residence, criminal history, social network characteristics, and prior exposure to violence, are linked to participation in criminal gun violence as either an attacker or the victim (Braga et al, 2010; Fowler et al 2015; Wheeler, Worden and Silver, 2019). Hence, using mentioned characteristics from prior research, the predictive model can discover hotspots based on demographic, socioeconomic, and geographical statistics. The machine learning algorithm uses the curated database to identify patterns and correlations that can assist law enforcement in implementing preventative measures and public safety plans. Likewise, the model can also predict chances of gun violence if the location/event comes under a certain radar. With this prediction, law enforcement can add more security based on the output given by the model which could be improved with time. Furthermore, as of right now, the police system does not use any predictive model or validated instruments that can predict future gun violence (Wheeler, Wooden, and Siver, 2019)

Background

In the US, incidents involving weapons have become dominantly common and excessive in recent years. Historically, the purpose behind those incidents has regularly been characterized as responsive conduct as opposed to reactive conduct. Crime prediction made by Victor Rotaru and the team was one of the models that developed a forecasting method that uses past events' spatiotemporal connections to predict crime events (Rotaru et al., 2022). With strong results spanning communities of varying sizes, from varying areas, and with

varying levels and histories of violence, this study generates predictions that are more precise than earlier models (Papachristos, 2022).

Law enforcement or authorities security commonly responds to occurrences once they have taken place. The initiative can allow the police department to make changes and adaptations based on the provided result. It recognizes the features by identifying patterns in historical data on gun violence incidents.

The main purpose of creating a predictive version for gun violence hotspots is to pick out patterns as these incidents don't randomly occur but rather are influenced by a selection of factors, which include socioeconomic, political, and demographics and other to name a few. The dataset produced for this project provides a basis for an in-intensity look. We can also position laws into place, regularize methods, and bolster safety by way of designating excessive-hazard areas so one can help deter these acts.

Current Issues and Suggest Topics

Multiple researchers and companies are working on developing systems or software that could detect Gun violence. Scylla is a company that is working on an AI system for real-time Gun detection (Scylla Technologies Inc., n.d.). There have been previous studies that have predicted firearm violence in urban communities based on population (Polcari et al, 2023). While the initiative taken by Scylla is great and detects real-time firearm detection, there are still significant challenges and gaps that need to be covered in the context of this topic. The system detects the firearm which might or might not stop an incident from happening. On the other hand, as the mentioned research predicts the firearm is only based on population, there are far more factors that contribute to incidents of these kinds. The existing systems often rely on limited data as resourceful data requires integrated datasets from multiple sources. It is very challenging to collect different data such as crime records of the offenders, socioeconomic details of the community, geographic characteristics, and details of the victims.

To overcome this challenge, the database created for this project incorporates all the necessary data that might be crucial toward creating a robust system for predicting the hotspots of Gun violence. The aggregated data from a source all around will ensure a more accurate prediction model.

Furthermore, let's discuss the challenges and issues that this project might face and suggest some solutions to them. One of the major issues that might be faced is ethical and privacy concerns of the collected data. As we collect data from various sources, the

consideration of the establishment of ethical guidelines and privacy protocols for sensitive data of this kind becomes very crucial. Therefore, this part is also taken care of in this project ensuring all that the predictive model gets is within the legal and ethical boundaries from reliable sources. Likewise, a big ethical challenge can be seen lurking as to who and how the model will be used. The prediction from the model can also lead to certain prejudices or biases. This can ultimately result in neglect in certain cases. Since the model can rely only on the provided data, it is not mandatory that the model can tell exactly where and what is going to happen but can give a prediction on how likely certain geographical locations are likely to see gun-related crimes.

Furthermore, the transition from theory to practice is a challenge on its own. As this is a very sensitive topic, it is very tough to predict another incident or the hotspot. The system might provide false positives which might create unnecessary chaos in the community as well as the safety department but being preventive is always better than facing any consciousness. The strategies for overcoming these challenges are emphasizing continuous integration of the predictive model into the existing safety framework and keeping the database up to date. This will help model for continuous improvement and adaptability. This part also explores how database design facilitates the update, and the addition of the new data sources and enhances the accuracy of the machine learning model.

In conclusion, addressing the above current issue and exploring the suggested topics ensure the development of the project in a holistic approach.

Methods and Evaluations

For the development of a Gun Violence Hotspot Predictive Model, we need to take various things and requirements into consideration. In this section, we dig deeper into the steps and methods of developing a robust system with the highest effectiveness and accuracy.

1. Data Collection and Database Design:

   This process involves the collection of various data from different sources such as government organizations, police departments, or safety departments. The data involves various crime records, demographic statistics, geospatial data, incident/crime reports, and data on offenders and victims.

   After this process, designing a database with accurate tables, attributes, and data types is very crucial. The following is a database tables table and data dictionary for the project.

DATABASE NAME: GunVoilenceData

Table Name:

- Incidents
- Offender
- Victims
- Demographics
- GeospatialData
- PredictiveModelData

Table Name	**Attribute Name**	**Data Type**	**Key Type**	**Description**
Incidents	incident_id	NUMBER	PK	Unique identifier for each incident
	incident_date	DATE		Date when the incident occurred
	zip_code	NUMBER		Zip code where the incident took place
	number_injured	NUMBER		Number of individuals injured in the incident
	number_killed	NUMBER		Number of individuals killed in the incident
	weapon_type	VARCHAR2(255)		Type of weapon involved in the incident
Offenders	offender_id	NUMBER	PK	Unique identifier for each offender
	incident_id	NUMBER	FK(Incidents)	Reference to the incident in which the offender was involved

	name	VARCHAR2(100)		Name of the offender
	age	NUMBER		Age of the offender
	gender	VARCHAR2(10)		Gender of the offender
	criminal_history	VARCHAR2(255)		Details of the offender's criminal history
Victims	victim_id	NUMBER	PK	Unique identifier for each victim
	incident_id	NUMBER	FK(Incidents)	Reference to the incident in which the victim was involved
	name	VARCHAR2(100)		Name of the victim
	age	NUMBER		Age of the victim
	gender	VARCHAR2(10)		Gender of the victim
	race	VARCHAR2(100)		Race of the victim
Demographics	zip_code	NUMBER	PK	Unique identifier for each geographic area
	incident_id	NUMBER	FK(Incidents)	Reference to the incident in which the demographic data is associated

	state	VARCHAR2(255)		State where the demographic data is collected
	population_density	NUMBER		Population density of the area
	ethnic_diversity	VARCHAR2(100)		Ethnic diversity of the area
	median_age	NUMBER		Median age of the population
	unemployement_rate	FLOAT(5)		Unemployment rate in the area
	poverty_rate	FLOAT(5)		Poverty rate in the area
GeospatialData	geo_id	NUMBER	PK	Unique identifier for each geospatial data record
	zip_code	NUMBER	FK(Demographics)	Reference to the geographic area to which the geospatial data belongs
	presence_of_school	NUMBER(1)		Indicator of the presence of a school in the area
	presence_of_bar	NUMBER(1)		Indicator of the presence of a bar in the area
	is_an_event	NUMBER(1)		Indicator of whether an event is happening in the area

PredictiveModelData	id	NUMBER	PK	Unique identifier for each predictive model data record
	zip_code	NUMBER	FK(Demographics)	Reference to the geographic area to which the predictive model data belongs
	feature1	NUMBER		Result of the predicted model-Accuracy Percentage
	feature2	NUMBER		Another possible result of model- F1 score
	feature3	NUMBER		Another possible result of model- F1 score- Precision score

Table 1. Data Dictionary

The following are SQL statements for creating the above tables with necessary attributes and data types:

```
sql> CREATE TABLE Incidents (incident_id NUMBER PRIMARY KEY, incident_date DATE, zip_code NUMBER, number_injured NUMBER, number_killed NUMBER, weapon_type VARCHAR2(255));
sql> CREATE TABLE Offenders (offender_id NUMBER PRIMARY KEY, incident_id NUMBER, name VARCHAR2(100), age NUMBER, gender VARCHAR2(10), criminal_history VARCHAR2(255), CONSTRAINT fk_offenders FOREIGN KEY (incident_id) REFERENCES Incidents(incident_id));
sql> CREATE TABLE Victims (victim_id NUMBER PRIMARY KEY, incident_id NUMBER, name VARCHAR2(100), age NUMBER, gender VARCHAR2(10), race VARCHAR2(100), CONSTRAINT fk_victims FOREIGN KEY (incident_id) REFERENCES Incidents(incident_id));
sql> CREATE TABLE Demographics (zip_code NUMBER PRIMARY KEY, incident_id NUMBER, state VARCHAR2(255),  population_density NUMBER, ethnic_diversity VARCHAR2(100), median_age NUMBER, unemployement_rate FLOAT(5), poverty_rate FLOAT(5), CONSTRAINT fk_demographics FOREIGN KEY (incident_id) REFERENCES Incidents(incident_id));
sql> CREATE TABLE GeospatialData (geo_id NUMBER PRIMARY KEY, zip_code NUMBER, presence_of_school NUMBER(1) DEFAULT 0 NOT NULL, presence_of_bar NUMBER(1) DEFAULT 0 NOT NULL , is_an_event NUMBER(1) DEFAULT 0 NOT NULL, CONSTRAINT fk_geospatialdata FOREIGN KEY (zip_code) REFERENCES Demographics(zip_code));
sql> CREATE TABLE PredictiveModelData (id NUMBER PRIMARY KEY, zip_code NUMBER, feature1 NUMBER, feature2 NUMBER, feature3 NUMBER, CONSTRAINT fk_predictivemodel FOREIGN KEY (zip_code) REFERENCES Demographics (zip_code));
```

```
SQL> CREATE TABLE Incidents (incident_id NUMBER PRIMARY KEY, incident_date DATE, zip_code NUMBER, number_injured NUMBER, number_killed NUMBER, weapon_type VARCHAR2(
255));

Table created.

SQL> CREATE TABLE Offenders (offender_id NUMBER PRIMARY KEY, incident_id NUMBER, name VARCHAR2(100), age NUMBER, gender VARCHAR2(10), criminal_history VARCHAR2(255)
, CONSTRAINT fk_offenders FOREIGN KEY (incident_id) REFERENCES Incidents(incident_id));

Table created.

SQL> CREATE TABLE Victims (victim_id NUMBER PRIMARY KEY, incident_id NUMBER, name VARCHAR2(100), age NUMBER, gender VARCHAR2(10), race VARCHAR2(100), CONSTRAINT fk_
victims FOREIGN KEY (incident_id) REFERENCES Incidents(incident_id));

Table created.
SQL> CREATE TABLE Demographics (zip_code NUMBER PRIMARY KEY, incident_id NUMBER, state VARCHAR2(255),  population_density NUMBER, ethnic_diversity VARCHAR2(100), me
dian_age NUMBER, unemployement_rate FLOAT(5), poverty_rate FLOAT(5), CONSTRAINT fk_demographics FOREIGN KEY (incident_id) REFERENCES Incidents(incident_id));

Table created.

SQL> CREATE TABLE GeospatialData (geo_id NUMBER PRIMARY KEY, zip_code NUMBER, presence_of_school NUMBER(1) DEFAULT 0 NOT NULL, presence_of_bar NUMBER(1) DEFAULT 0 N
OT NULL , is_an_event NUMBER(1) DEFAULT 0 NOT NULL, CONSTRAINT fk_geospatialdata FOREIGN KEY (zip_code) REFERENCES Demographics(zip_code));

Table created.

SQL>
SQL> CREATE TABLE PredictiveModelData (id NUMBER PRIMARY KEY, zip_code NUMBER, feature1 NUMBER, feature2 NUMBER, feature3 NUMBER, CONSTRAINT fk_predictivemodel FORE
IGN KEY (zip_code) REFERENCES Demographics (zip_code));

Table created.
```

Fig 1. Database Table Creation

Below is a data dictionary table that describes the columns of each table and respective data type:

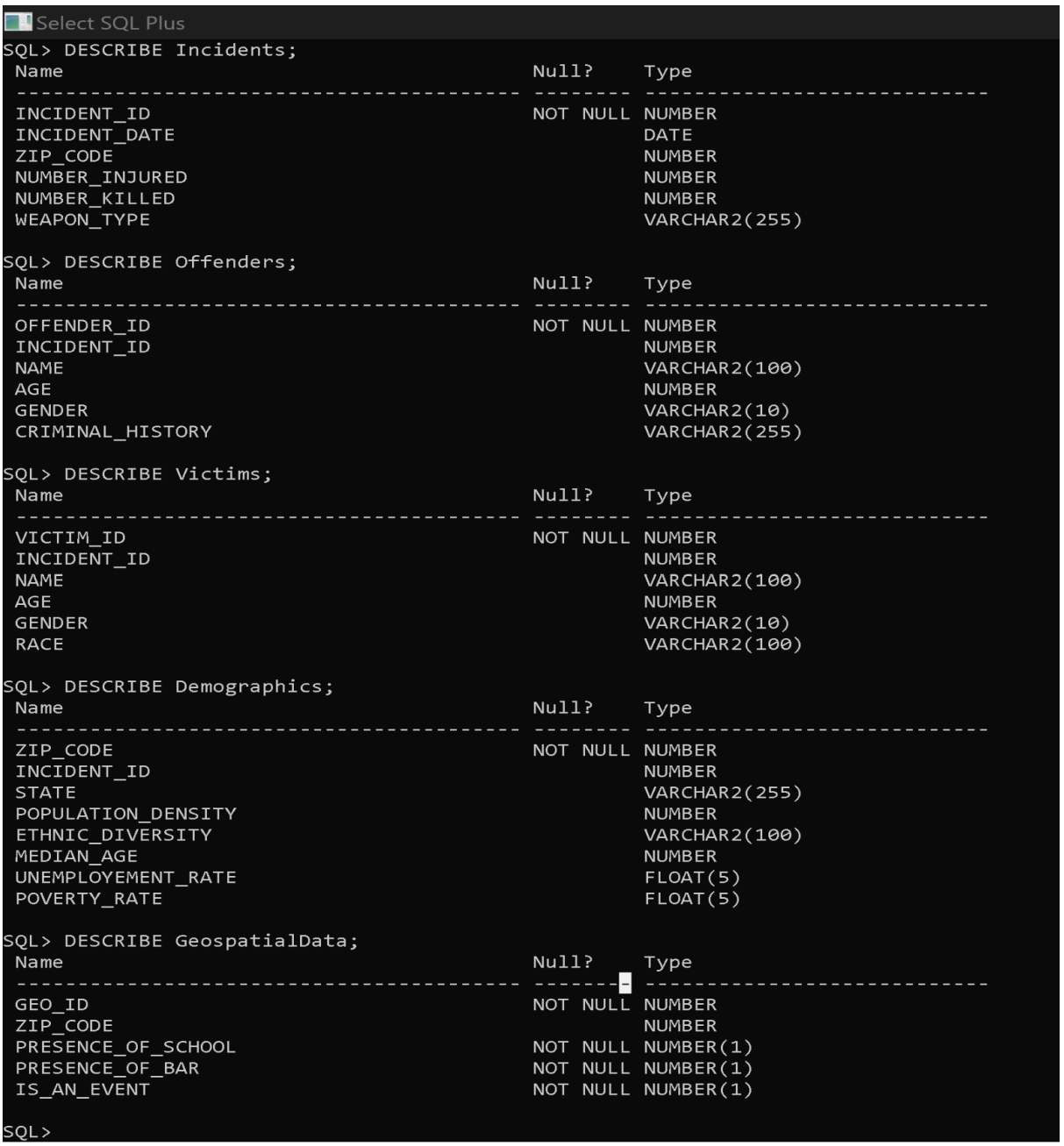

Fig. 2 Database tables and columns description.

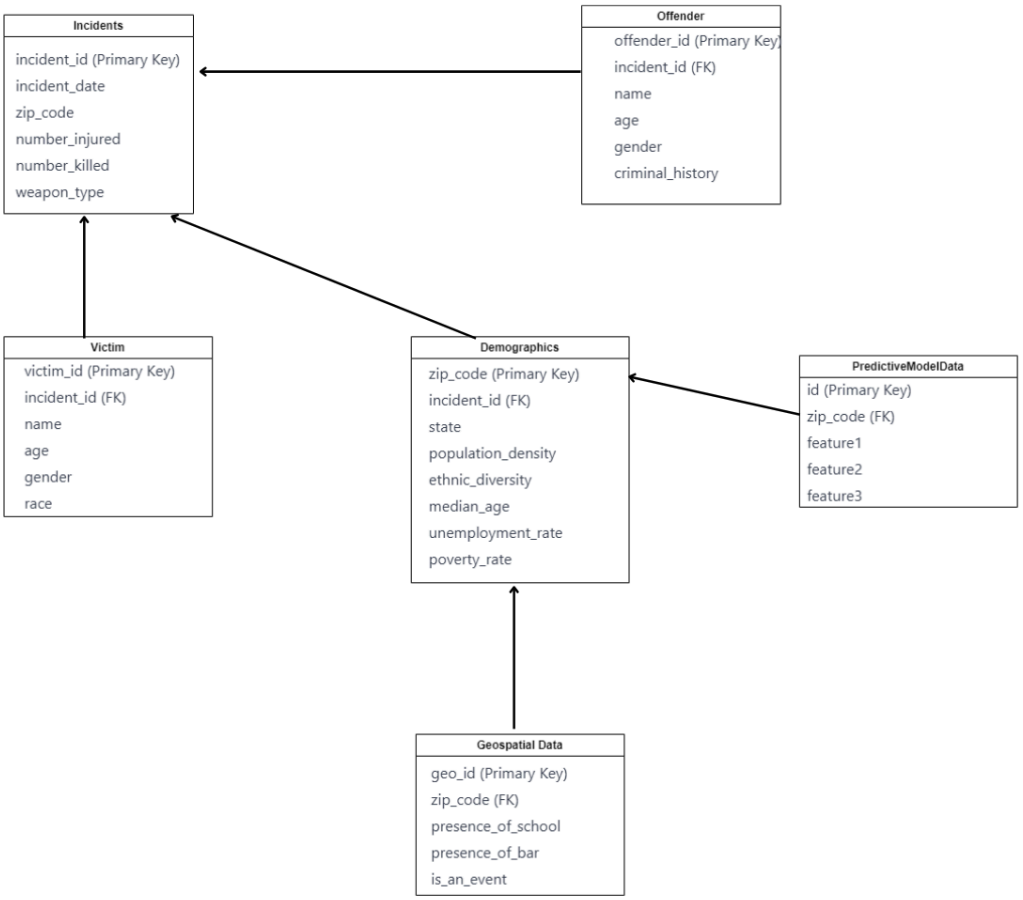

Fig. 3 Entity Relationship diagram

Table Name	Related Table	Relationship Type
Incidents	Offenders Victims Demographics	One-to-Many One-to-Many Many-to-One
Offenders	Incidents	Many-to-One
Victims	Incidents	Many-to-One
Demographics	Incidents GeospatialData PredictiveModelData	One-to-Many One-to-Many One-to-One

Table. 2 Relationship Table

The following are SQL statements for populating the Database with sample data:
Insertion in Incidents Table:

```
INSERT ALL
  INTO Incidents(incident_id, incident_date, zip_code, number_injured, number_killed, weapon_type) VALUES (1, TO_DATE('2023-01-01', 'YYYY-MM-DD'), 10001, 3, 1, 'Firearm')
  INTO Incidents(incident_id, incident_date, zip_code, number_injured, number_killed, weapon_type) VALUES (2, TO_DATE('2020-01-02', 'YYYY-MM-DD'), 20002, 2, 0, 'Knife')
  INTO Incidents(incident_id, incident_date, zip_code, number_injured, number_killed, weapon_type) VALUES (3, TO_DATE('2019-01-03', 'YYYY-MM-DD'), 30003, 1, 1, 'Firearm')
SELECT * FROM dual;
```

Likewise populated other tables with sample data using same INSERT SQL statement.

Normalization of the database to 3NF:

```
sql> CREATE TABLE States (state_id NUMBER PRIMARY KEY, state VARCHAR2(255) NOT NULL);
sql> ALTER TABLE Demographics ADD state_id NUMBER;
sql> ALTER TABLE Demographics DROP COLUMN state;
```

2. **Development of Machine Learning Model:**

   The selection and implementation of a machine learning algorithm play a vital role in the result of the predictive model. Given the nature of data, deep learning models such as neural networks are best suitable for the predictive goal of the project. Architectures like convolutional neural networks are capable of recognition of complex patterns in data. This model is capable of adapting to complex relationships within the data such as geospatial and demographic data which is very important to recognizing the patterns in gun violence hotspots.

   Furthermore, the continuous learning capability of neural networks ensures that the model can evolve and improve in the future and can adapt to the changes in socio-economic factors, and demographics which is very crucial for maintaining relevance as the occurrence of gun violence incidents is very dynamic and sensitive.

DATA FLOW DIAGRAM

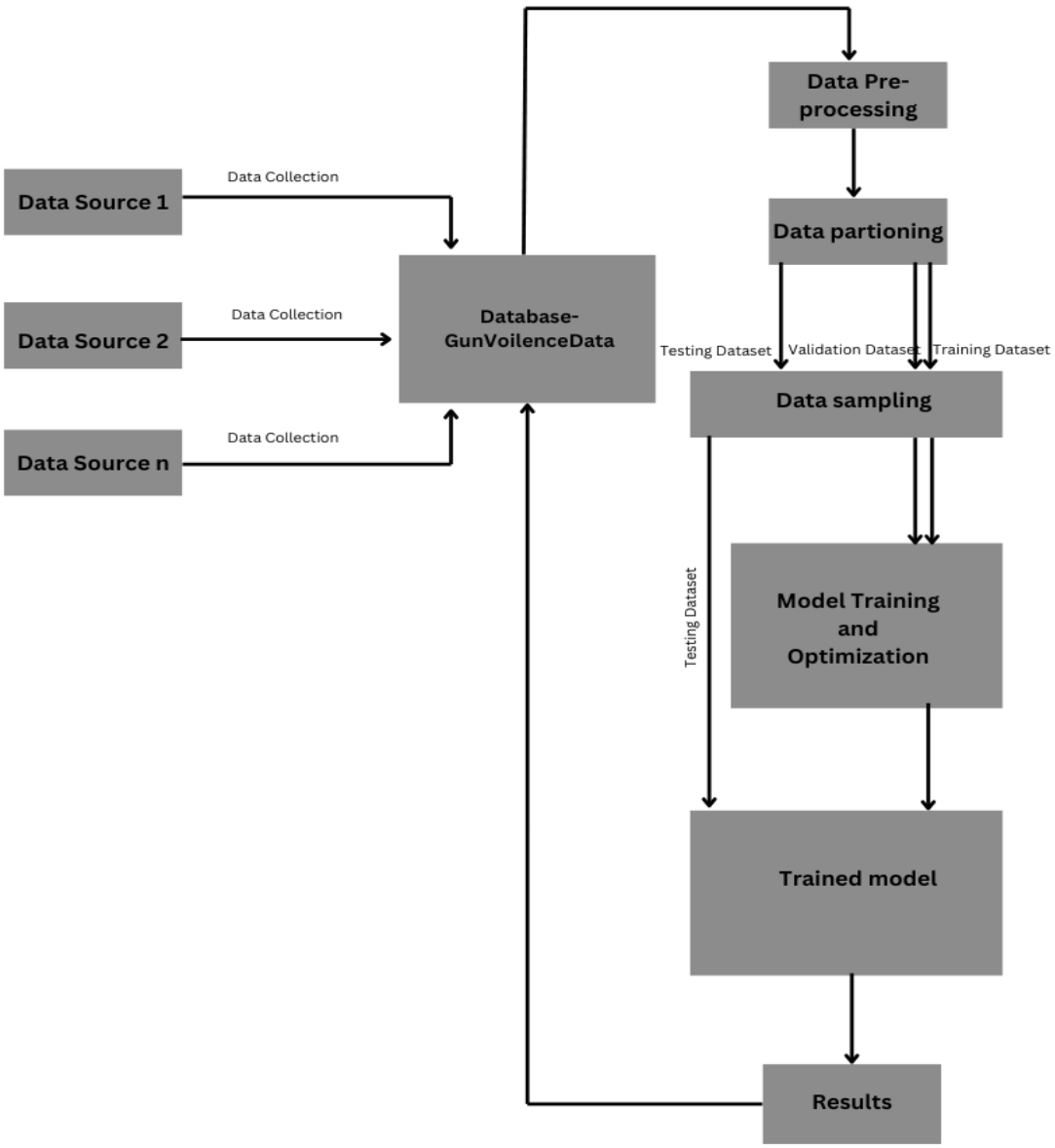

Fig 4. Data Flow Diagram

3. Ethical Guidelines and Privacy Protocols:

One of the major factors regarding this project is incorporating ethical guidelines and considering privacy protocols. As a model that has been trained and tested with sensitive information. The information in the database should be secured. In this process, the implementation of robust security is involved which includes the implementation of encryption techniques to protect data. Additionally, a strong authentication system is implemented to control the access to the database.

4. Real-world Implementation:

The task of implementing this project in the real world is another challenge in the collection of data from various sources. Moreover, the data needs to be updated time and again. So, the most important part of this is collaboration with different authoritative and governmental bodies or public safety departments. Establishing a partnership with law enforcement agencies and governmental bodies is crucial in ensuring the legitimacy and reliability of the collected data and results in the successful implementation of this project.

5. Evaluation Metrics:

The correct evaluation of this framework is critical in measuring the success of this predictive model. As sensitive project as this, it is very crucial to have effective evaluation metrics to navigate the success of the project and work continuously on its improvement. So, some metrics that can be used for evaluation are precision, recall, and F1 Score to correctly identify gun violence hotspots and minimize false positive results.

Additionally, we can conduct community impact assessments and analyze the feedback given by the community can contribute to evaluating the model's effectiveness and ethical consideration. These approaches will help in creating a project with the highest accuracy within its ethical boundaries.

6. Continuous Improvement:

Continuous improvement of a system is another crucial part as updating a system with time makes it relevant with time. With regular updates of the database and the incorporation of new data sources, our model can adapt to the evolving dynamic of socio-economic factors, and community, and legislative changes. Updating our database with the new results can help increase the effectiveness.

Outcome:

The success of this project highly relies on the robust database. In this section, we delve into the outcome of the project emphasizing the pivotal role of databases in supporting information systems and business operations.

- Accelerate strategic decision-making ability:

The database serves as a backbone to this project in the development of the model. By amalgamation of all the data from various sources, the developed machine

learning model will identify the patterns and correlations for predicting incidents. This enables the allocated safety department/governmental body or any concerned party to make quick and strategic decisions such as resource allocation, law enforcement, providing strict security, etc. Hence, considering the long term, the community/state can implement rules and regulations focusing on public safety. From this, we can also identify the main cause of incidents of this kind such as socio-economic factors, and focus on the improvement to minimize the possibility of hazard. This decision-making heavily relies on the insight obtained from the created database.

- Support the Business:

  Operational systems, databases, and data warehousing are integral components of this project in development, improvement, and providing accurate results for this project. The operational system of this project ensures the continuous collection of data from various sources and trains the model with the updated dataset to keep the model updated and adapt to the evolving change in the community. The database serves as a centralized repository for storing and retrieving the data. Furthermore, the data warehouse keeps a record of a large volume of historical data which is crucial for training the model and it is also helpful to reflect results that have changed over time. It also helps to reflect the effectiveness of the actions taken with the help of the prediction of this model.

- Cloud Technology, Distribution, and Virtualization:

  The Gun Violence Hotspot Predictive Model leverages cloud technology for enhancement, scalability, flexibility, and accessibility. Cloud-based infrastructures allow a system to efficiently store the data and retrieve the data as well. It also helps in the smooth handling of a large volume of information. It ensures the distribution of data across the cloud servers enhances the system performance and helps execute the query processes and model training faster.

  Virtualization ensures in creating of a dynamic and adaptable environment for the working of the project where it makes sure that the system is efficiently allocating computing power and storage based on workload. Virtualization ensures optimal performance during the high computing power demand such as updating the database and training the machine learning model.

- Network Infrastructure and Security:

  The information system for the Gun Violence Hotspot Predictive Model focuses on robust network infrastructure and security measures. A secure and reliable network ensures seamless and secure communication between different components of the system such as data source, database, and machine learning model. This part is time-consuming and is very prone to data breaches and different possible threats, therefore, the efficient and secured flow of data is very crucial in safeguarding sensitive information of these kinds. The security measures include encryption of data and prevention of the access of data to an unauthorized person.

Summing up, this prediction system is solely dependent on the effectiveness of the database and the information system. The diverse nature of the collected data enables strategic decision-making, ensuring the scalability, flexibility, and security of the system. This project holistically attempts to solve the current problem as well as it is prepared to adapt to future development and change in the information in the context of the complex and sensitive nature of the project.

Future Works

For future enhancement, one of the major works will be the addition of a feature that will automatically retrieve the information from different data sources and update it to our database. Another work includes exploring different machine learning algorithms and enhancing the accuracy of our model. To add more to this, we can also explore ensemble learning techniques for combining different models to help increase the accuracy. As sensitive project as this, community involvement is very important, so in the future feedback can be taken from the community and implemented in our project to make it a more human-centric project.

Summary

This term paper is an in-depth analysis of the use of the information system to develop predictive modeling for reducing gun violence by sourcing various relevant data. The design of a sophisticated database system and machine learning algorithm to identify the hotspot of the possible incident can help stop the hazard or can help in the enforcement of the law. This paper discusses the current issues that our communities have been facing, and companies and organizations that are working simultaneously on possible solutions. It

discusses the methods of how this project can be implemented in practice. This paper focuses on the crucial part of the project i.e. database system which performs as a backbone of this project and how it helps make strategic decisions, integrate cloud computing, and ensure data security for this project.

Furthermore, this paper also discusses the strategic plan to keep this project relevant with time, continuously update the database and the system, and adapt to the change in the socio-economic sector of the community. Overall, this term paper is an amalgamation of information on how a robust database system is crucial in the development of a successful project.

Reference

Braga, A., Papachristos, A., & Hureau, D. (2010). The Concentration and Stability of Gun Violence at Micro Places in Boston, 1980–2008. Journal of Quantitative Criminology, 26(1), 33–53. https://doi.org/10.1007/s10940-009-9082-x

Fowler, K. A., Dahlberg, L. L., Haileyesus, T., & Annest, J. L. (2015). Firearm injuries in the united states. Preventive Medicine, 79, 5-14. https://doi.org/10.1016/j.ypmed.2015.06.002

Gramlich, J. (2023, April 26). What the data says about gun deaths in the U.S. Pew Research Center. https://www.pewresearch.org/short-reads/2023/04/26/what-the-data-says-about-gun-deaths-in-the-u-s/

Papachristos, A. V. (2022). The promises and perils of crime prediction. *Nature Human Behaviour, 6*(8), 1038-1039. https://doi.org/10.1038/s41562-022-01373-z

Polcari, Ann M., Hoefer, Lea E., Zakrison, Tanya L., Cone, Jennifer T., Henry, Marion C. W., Rogers, Selwyn O., Slidell, Mark B., Benjamin, Andrew J. (2023 July). A novel machine-learning tool to identify community risk for firearm violence: The Firearm Violence Vulnerability Index. *Journal of Trauma and Acute Care Surgery, 95*(1), 128-136, DOI: 10.1097/TA.0000000000003992

Rotaru, V., Huang, Y., Li, T., & Chattopadhyay I. (2022). Event-level prediction of urban crime reveals a signature of enforcement bias in US cities. *Nat Hum Behav 6,* 1056–1068 https://doi.org/10.1038/s41562-022-01372-0

Scylla Technologies Inc. . (n.d.). Scylla Gun Detection System. https://www.scylla.ai/gun-detection/

Wheeler, A. P., Worden, R. E., & Silver, J. R. (2019). The Accuracy of the Violent Offender Identification Directive Tool to Predict Future Gun Violence. Criminal Justice and Behavior, 46(5), 770-788. https://doi.org/10.1177/0093854818824378