Severity Classification Prediction on Gun Violence Data

Jenifer Werthman
Computer Science
Kent State University
Kent, Ohio USA
jwerthma@kent.edu

Priyanka Vyas
Computer Science
Kent State University
Kent, Ohio USA
pvyas2@kent.edu

Sangeeta Kumawat
Computer Science
Kent State University
Kent, Ohio USA
skumawat@kent.edu

***Abstract*— Gun violence data classification refers to the collection and analysis of information on incidents involving firearms, including shootings, homicides, suicides, and accidental discharges. This data is used by researchers, policymakers, and law enforcement agencies to understand the scope and nature of gun violence, identify trends and patterns, and develop strategies to prevent and reduce its occurrence. This study presents the severity classification of gun violence data in three category low, mid, high. The classification is based on features such as number of victims, locations, date affected in the event. Several classification machine learning techniques is used and compared like Logistic Regression, K-Means, LSTM, RNN. In terms of application, this approach can prove to be useful, to create a flagging system for police deployment based on incidents severity level.**

Keywords—Logistic Regression, binary classification, kmeans, elbow, clustering, feature selection, visualization, lstm, rnn.

I. Introduction

Gun violence is a complicated and diverse problem with severe consequences for people, families, and society. Gun violence is the leading cause of death and injury in the US and other nations. Access to firearms, mental health conditions, and social and economic inequities are risk factors for gun violence. Studies, individuals, and law enforcement agencies all depend on data to address this issue as it helps them understand the scale and nature of gun violence, identify trends and patterns, and create effective preventative measures plans. Data on gun violence is collected and analyzed, but there are restrictions and difficulties, such as underreporting of some types of incidents and incompatible definitions and classifications of data.

Machine learning uses classification algorithms to determine the category or class of a parameter using a set of characteristics or predictions. In recent years, a number of classification methods have been developed, such logistic regression, k-means, LSTM, and RNN. In comparison to k-means, which is frequently utilized for clustering and unsupervised learning, logistic regression is an increasingly common method for binary classification. The neural network varieties LSTM and RNN have been frequently used to sequence data, covering time series analysis and natural language processing. Depending on the kind of data and problem at hand, each of the these classification systems offers advantages and disadvantages.

II. Literature Review

[1] The study "A CNN-RNN Combined Structure for Real-World Violence Detection in Surveillance Cameras" presents a distinctive structure that combines convolutional neural networks with recurrent neural networks. On a sizable dataset of actual instances of violence, the framework has a high accuracy of 91.67%. Other studies examined the use of techniques based on deep learning, including one-class support vector machine and auto encoder models for violence detection in surveillance camera data.

[2] The paper "Predicting Crime and Other Uses of Neural Networks in Police Decision Making" by Schutt and Zilinskas (1991) explores the use of neural networks in law enforcement for predicting crime and supporting police decision-making. The study found that a neural network outperformed traditional statistical methods in predicting crime, achieving a 70% accuracy rate. Other applications of neural networks in law enforcement include suspect identification and recidivism prediction. However, ethical considerations such as bias and privacy must be carefully addressed when implementing these technologies.

[3] The paper, "A Time-Series Analysis of Firearm Purchasing After Mass Shooting Events in the United States" by Kalesan et al. (2016) examines the connection between mass shooting incidents and gun buying behavior in the country. The analysis found that there was an average increase of 3.5 million more handgun purchases per year in the months after mass shooting incidents. The impact was stronger for handgun purchases and occasions that garnered a lot of media attention. The connection between mass shooting incidents and gun purchasing behavior in the US is modeled in the research using autoregressive integrated moving average (ARIMA) time-series analysis.

[4] The implementation of machine learning models to estimate firearm homicides in the US in near real-time is explored in the study "Development of a Machine Learning Model to Estimate US Firearm Homicides in Near Real Time" by Lee et al. (2020). Using information from the National Vital Statistics System, the study created a model that could correctly forecast firearm homicides up to 4 weeks in advance. The model was especially good in foretelling homicides with firearms committed in cities

and in the summer. The model for estimating firearm homicides in close to real-time is developed using a machine learning technique known as gradient boosting.

[5] In order to predict future terrorist activities, the study "A hybrid deep learning-based framework for future terrorist activities modelling and prediction" presents a hybrid deep learning framework that blends convolutional neural networks and long short-term memory networks. The Global Terrorism Database dataset used to train the framework for predicting the possibility of a terrorist attack in a certain area produced results with a high accuracy of 91.62%.

III. Feature Selection

A. Data

The data was originally from the Gun Violence Archive, which is an online archive of gun violence events from media, law enforcement, and other sources. They are a non profit group with the goal is to provide near-real time data about gun violence events.

The reference we used came from Kaggle which contained 239, 678 events from January 2013 to March 2018. They have captured 29 data points, including date, number of killed, number of injured, source url, gun type, participant age groups, participant gender, arrests, and state district.

B. Summary

Feature Selection is the analysis of data to determine which features have the best relationships and will be the most useful in a predictive model. This will not only help improve accuracy but will also cut down the amount of data being processed reducing computational costs such as time and computer power.

C. Details

To analyze the gun violence data, we imported the data and parsed various data points to obtain number of participants, month, year, participant age ranges, arrests, and victims. We analyzed each combination to determine which features we should use for our classification. Type of gun used during an event contained a lot of missing information. Gender didn't provide much information, since we only have gender of participants but lack specific information on if the gender is pertaining to the victims or person(s) with a gun. We combined number of killed and number of injured to determine number of victims for each event. We found some strong correlations between state and number of victims.

D. Figures

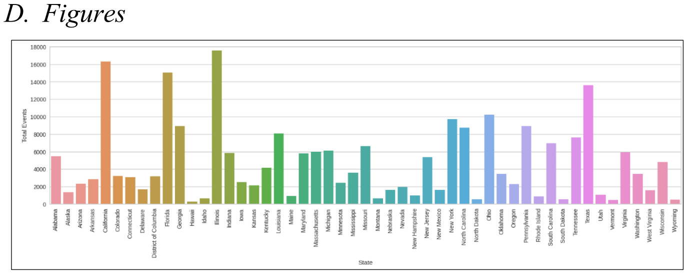

Figure 3.1: Number of Events by State

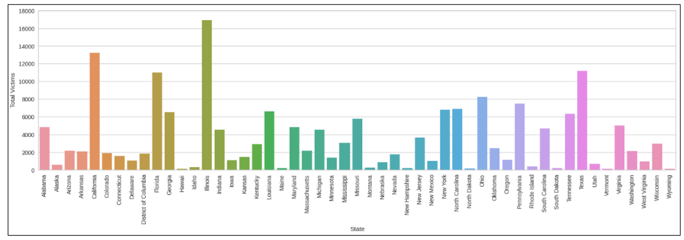

Figure 3.2: Number of Victims by State

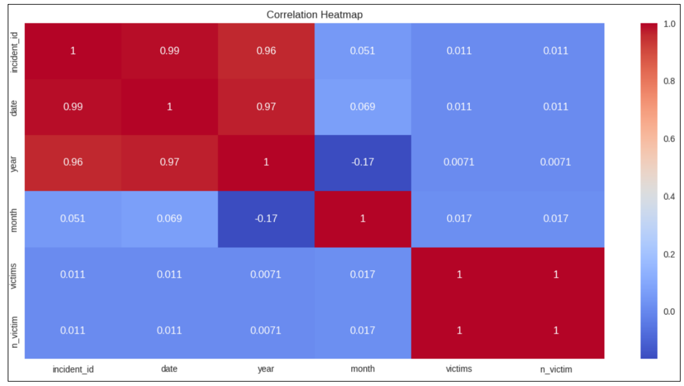

Figure 3.3: Feature Correlation Heatmap

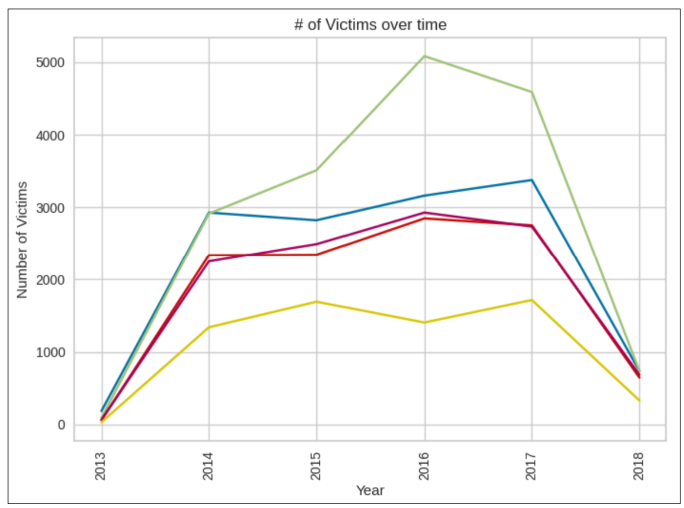

Figure 3.4: Victims by Year, by State

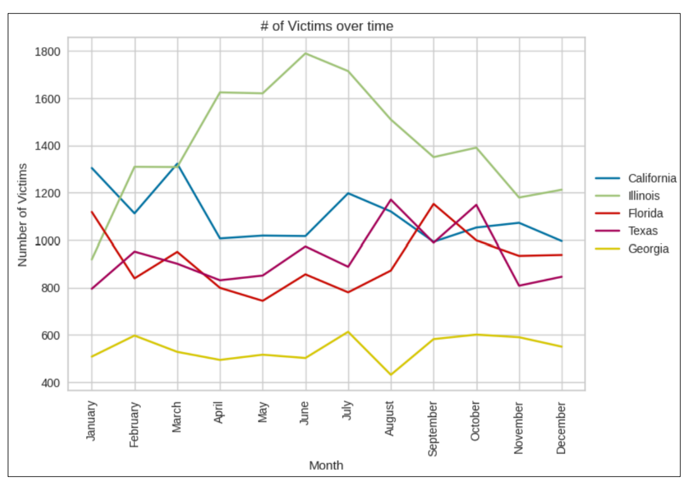

Figure 3.5: Victims by Month, by State

IV. Methodology

Method 1: Clustering

A. Summary

Clustering is the process of grouping the similar or relative data points. This allows clusters of data to be classified together. There are many types of clusters; however, there is no criteria for a cluster. Some data forms natural clusters and others they density, hierarchical, or shape references.

B. Details

In this study k-means and the Elbow Method is used to determine how many clusters we should use for this data set. K-means clustering looks for a fixed number of clusters in a dataset. K-means was originally used for signal processing and seemed to be a good fit for sensor data as well. Applying k-means to the data it illustrates that the data splits nicely into 5 clusters.

C. K-Means Clustering

K-Means Clustering can also be used as an unsupervised model to predict values. Using the clusters that were previously established, the encoded data is separated into train and test data then the predicted values are evaluated.

For error handling, the Accuracy Score method is used, which uses the mean intra-cluster distance. Using this method, the closer to 1 the better.

D. Figures and Diagrams

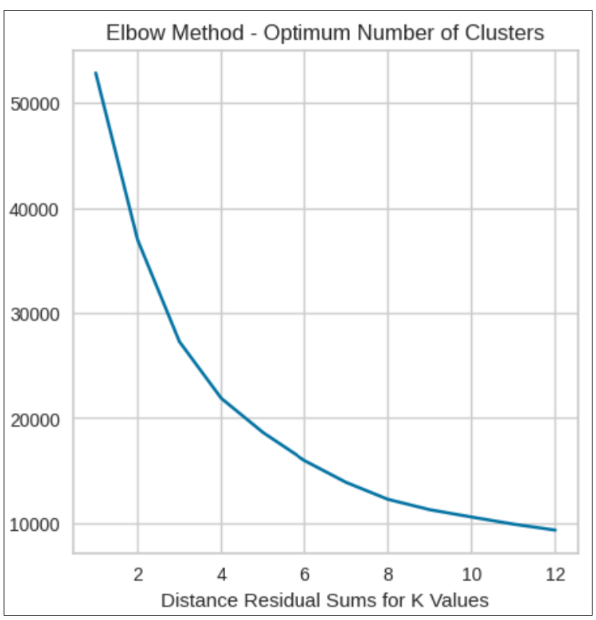

Figure 4.1.1: Elbow Method

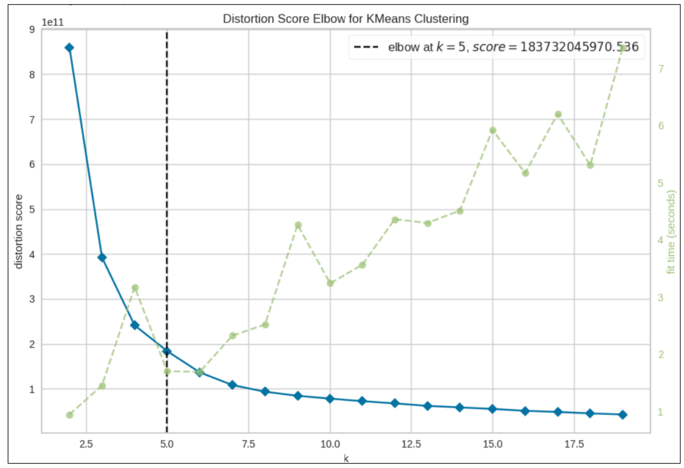

Figure 4.1.2: Elbow Method (k=5)

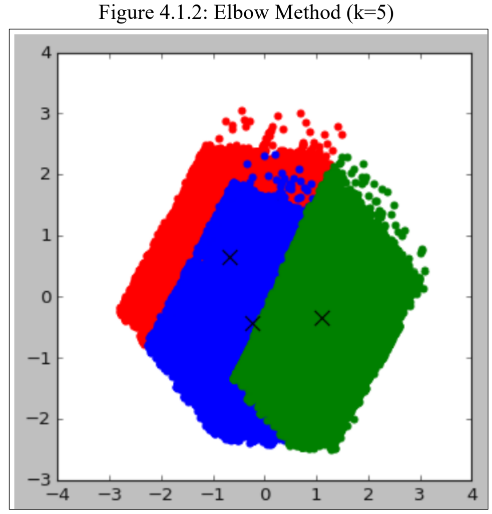

Figure 4.1.3: Clustered classification data – When looking solely at features

Method 2: Binary Classification

Binary classification is a type of supervised learning statistical classification which separates the data into two groups for each class. Categories are predefined and probabilistic observations are used to categorize the data into the two groups. Some methods used for binary classification are decision trees, random forests, Bayesian networks, logistic regression, and linear genetic programming. There are four types of combinations, True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

E. Implementation

First we standardized the data and separated date, state, and city features from the number of victims. Then the data is split into training and test data sets and then the binary labels are separated into their own columns for each class High Severity, Moderate Severity, and Low Severity. Finally each binary classifier is trained for each label using Logistic Regression and after calculating the predicted probabilities for each class, the probabilities are combined into a single array and the ROC curve and AUC score for each class is computed.

F. Results

Applying binary classification to the data produced an accuracy score of 40.74%. When the data is split and dropped the accuracy seemed to do worse then using the entire data set. We also tried keeping the classes in a single column verses multiple columns, however the single column produced lower accuracy as well.

G. *Figures and Diagrams*

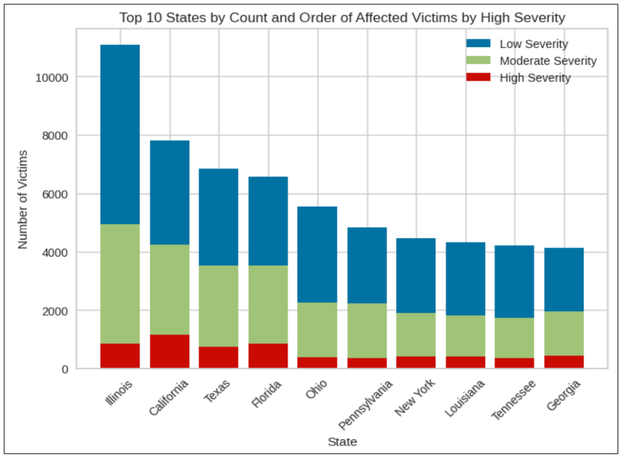

Figure 4.2.1: Top 10 States with highest victim count from high severity events

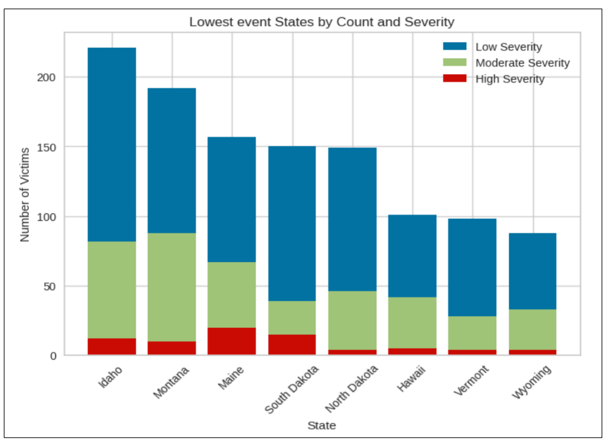

Figure 4.2.3: Top States with lowest victim count

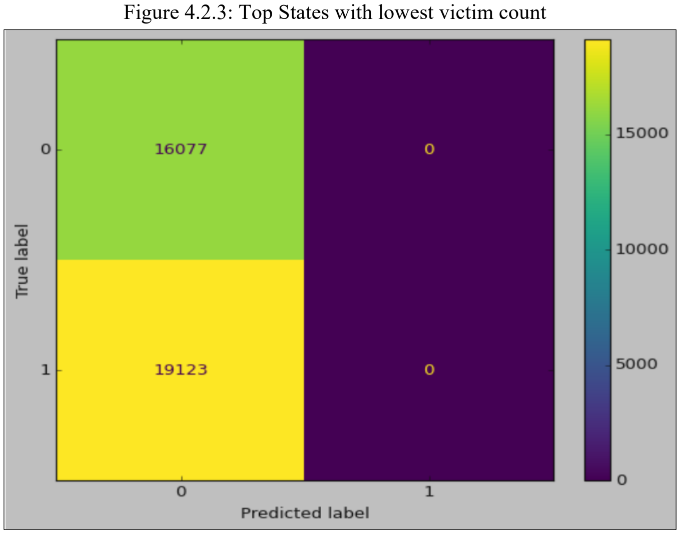

Figure 4.2.4: Results Confusion Matrix

METHOD 3: RECURRING NEURAL NETWORKS (RNN)

Recurrent Neural Networks (RNN) are a type of unsupervised deep learning which uses sequential data or time series data. RNNs are typically used for language translation, natural language processing, and speech recognition. Like CNNs, they are based on the brains neural network and use training data to learn; however, they are differing by using previous inputs to influence new inputs by using a feedback loop.

To implement the RNN, feature selection is used to determine what the X values would be. For the Y values, the Severity Level is assigned to each row based on number of victims per event. The data is then transformed the string values by encoding them.

The RNN was setup to have 4 simple layers using the sigmoid and ReLU activation layers, Adam optimizer, with a dropout of .25 and a single dense fully connected layer. The fully connected layer is where each neuron transforms the input vector through a weight's matrix. All possible neurons are connection layer to layer. The result is that every input will influence the output. This allows the network to learn non-linear combinations for each feature.

The output of the RNN is a matrix/array the same batch size as the input but other information will/may change. Accuracy is measured by looking at Mean Square Error. Mean Square Error assesses how close the regression line matches the data points. This is calculated by taking the average of errors squared. For Mean Square, the closer to 0 the better.

H. *Outcome*

When RNN is used with a single layer or 4 rnn layers and one dense layer, the same result is produced of .48 accuracy but when a dropout layer of .25 was added the accuracy went down to .46.

METHOD 4: LSTM

Long Short-Term Memory Networks (LSTM) is a type of unsupervised deep learning which is a variation of RNN that is capable of learning long term dependencies. LSTMs are typically used for speech recognition, machine translation, robot control, handwriting recognition, and healthcare. LSTMs partially solve the vanishing gradient problem; however, the exploding gradient problem is still possible.

I. *Outcome*

When LSTM is used with a single LSTM layer and 1 dense layer using ReLU and Sigmoid activation, and adam optimizer and no dropout layers. This produced an accuracy of .47.

METHOD 5: MULTI-CLASS BINARY CLASSIFICATION

In this approach, the model performs a classification task for predicting severity of Gun violence incidents and the severity of incidence is classified into three classes based on the number of victims i.e., people affected by Gun violence incidents. First, the dataset is loaded using pandas read_csv function then dataset is preprocessed by dropping the irrelevant columns, filling the missing values with 0, creating a new column for the number of victims, converting the date column to numerical representation, and encoding the state and county columns using Label Encoder. The dataset doesn't have a pre-defined label therefore, severity labels namely High Severity, Moderate Severity and Low Severity are created based on the number of victims and as per the below logic:

- High Severity: n_victim ⩾ 7
- Moderate Severity: 4 ⩽ n_victim < 7
- Low Severity: n_victim < 4

where 7, is mean of total victims.

The features used from the dataset for modeling is date, state, city or county fields, once the feature is standardized using Standard Scalar, the dataset is split into training (80%) and testing (20%) sets, random state is set to 42 to ensure reproducibility and a separate binary classifier is trained for each label/severity level using Logistic regression, one for each severity level. The lambda function is used in the condition to return 1 if it meets the certain condition or range defined for the severity level otherwise it returns 0. After the model fit on train data, the predicted probability for each class is calculated on test data and later are combined into a single array so as to determine which severity class each sample belongs to. Finally, model and its performance is evaluated using ROC-AUC curve and Confusion matrix. The ROC curve is a plot between true positive rate and false positive rate at various thresholds and it takes two arguments true labels and predicted probabilities for a single class.

To compute the ROC curve and AUC score, the code is comparing the predicted probabilities generated by the logistic regression models against the binary labels for each class (1 if the number of victims falls in the corresponding severity range, and 0 otherwise). The apply() method with a lambda function is used to convert the original numeric labels into binary labels for each class. The thresholds output from roc_curve() are an array of probability thresholds used to generate the fpr and tpr values. By default, the thresholds are set to include 0 and 1, which correspond to the lowest and highest possible probabilities, respectively. These thresholds are used to plot the ROC curve. The AUC score is a measure of how well the model can distinguish between positive and negative classes, with a score of 1 indicating perfect classification and 0.5 indicating random guessing.

V. Experimental Evaluation

The result in method 5, is evaluating the performance of the three binary classifiers that have been trained to predict the severity level of gun violence. The performance is evaluated using the ROC curve and the AUC score, where the AUC (Area Under Curve) is a metric that measures the ability of the model to distinguish between positive and negative classes.

The plot in Fig 5.1 below, Receiver Operating Characteristic (ROC) curve, shows the trade-off between the true positive rate (sensitivity) and the false positive rate (1 - specificity) of a binary classifier as the classification threshold is varied. It contains the false positive rate, true positive rate, and threshold values at various points on the ROC curve. The ROC AUC score (0.99) is a single scalar that summarizes the performance of the binary classifier over all possible threshold values. The plot also includes a dashed line representing random guessing and a legend showing the AUC score for each class.

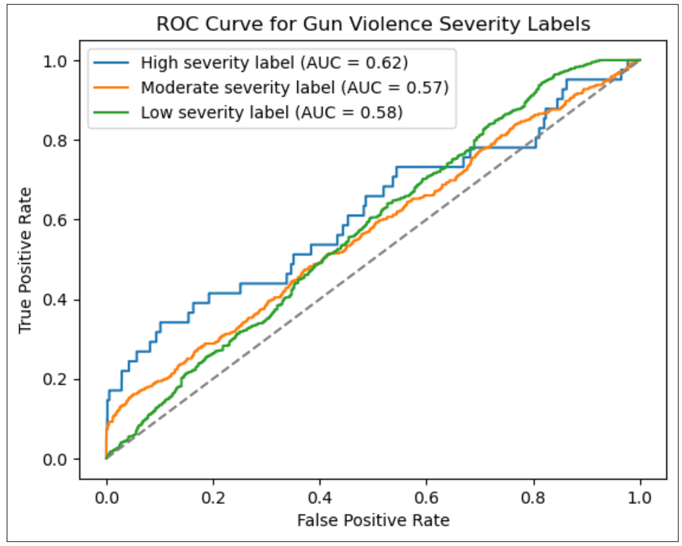

Fig 5.1: ROC Curve for Gun Violence Severity Labels

The AUC (Area Under the ROC Curve) scores for the three levels are reported: 0.62 for High Severity, 0.57 for Moderate Severity, and 0.58 for Low Severity. AUC measures the model's ability to distinguish between positive and negative samples, with a higher AUC indicating better performance. For the high severity label, the AUC score is 0.62, which indicates that the model is better than random but has room for improvement.

The random guessing line is the diagonal line from the bottom-left corner to the top-right corner of the ROC plot. This line represents the performance of a classifier that makes random guesses and has an area under the curve (AUC) of 0.5. A good classifier should have an ROC curve that lies above the random guessing line, indicating that its performance is better than random guessing. The higher the AUC, the better the classifier's overall performance.

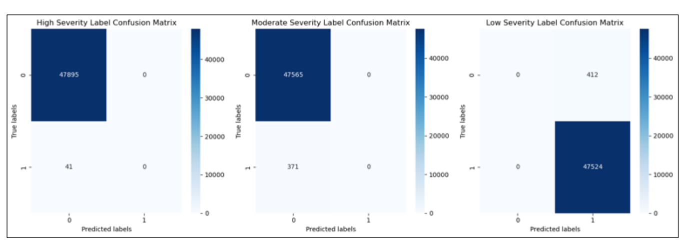

Fig 5.2: Confusion Matrix for Severity Levels

In Fig 5.2 for the High Severity label, the confusion matrix shows that out of 47,895 actual instances, the model predicted all of them as negative (0). The model did not predict any positive (1) instances for this label. This indicates that the model is not performing well for this label. For the Moderate Severity label, the confusion matrix shows that out of 47,936 actual instances, the model predicted 37 of them as positive (1) and the rest as negative (0). The accuracy score for this label is 0.9922, which indicates that the model is performing well for this label. For, the Low Severity label, the confusion matrix shows that out of 47,936 actual instances, the model predicted 47524 of them as negative (0) and the rest as positive (1). The accuracy score for this label is 0.9914, which indicates that the model is performing well for this label.

The accuracy score of the model is 0.999, which is very high, but can be misleading in imbalanced datasets where the majority class dominates. In a confusion matrix, 0 and 1 typically represent the two classes of a binary classification problem. The confusion matrix is a table that summarizes the performance of a classification algorithm by comparing the predicted labels with the actual labels for a set of data. In a binary classification problem, the predicted labels can be either 0 or 1, where 0 represents the negative class and 1 represents the positive class.

Similarly, the actual labels can also be either 0 or 1. The four possible outcomes of a binary classification problem are:

- True Positive (TP): The model correctly predicts the positive class.
- False Positive (FP): The model incorrectly predicts the positive class.
- True Negative (TN): The model correctly predicts the negative class.
- False Negative (FN): The model incorrectly predicts the negative class.

	Predicted 0	Predicted 1
Actual 0	TN	FP
Actual 1	FN	TP

Fig 5.3: Confusion matrix: Actual v/s Predicted.

As shown above in a fig 5.3 confusion matrix, these outcomes are arranged in a table where the rows represent the actual labels and the columns represent the predicted labels.

VI. Implementation Difficulties

The main difficulty in this model is that the labels are imbalanced, with a majority of the gun violence incidents being low severity.

The dataset is highly imbalanced, with a large number of instances belonging to the "low severity label" class and a small number of instances belonging to the "high severity label" class and the same can be seen from the fig (6.1) below. This can lead to bias in the model and affect the performance of the model in predicting the "high severity label" class accurately.

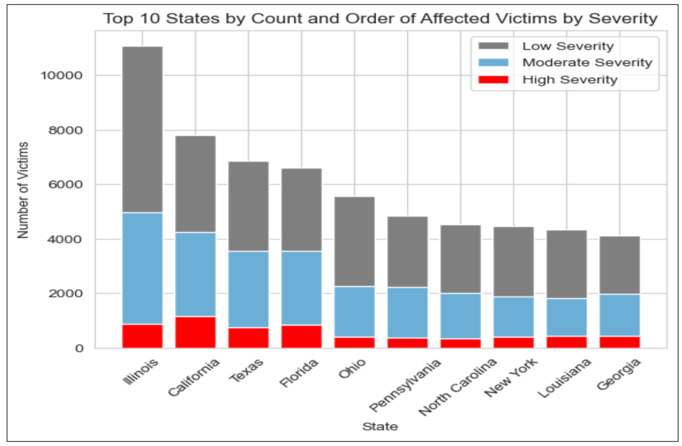

Fig: 6.1 Location v/s Number of victims Severity level based mapping

Another major difficulty was to handle the missing values and feature selection. By dropping the missing values, the model would perform ineffectively and lead to loss of information. Though we tried to fetch more information by parsing the data, but the dataset also contains many columns that are not relevant for the classification task, such as address or the source url, which doesn't provide a detailed information that can be used for modeling.

Also, it is learned that there could be several reasons why the model is failing to identify false negatives in the respective severity label. One possibility is that the data is imbalanced, meaning that there are far more negative examples than positive examples in the dataset. In this case, the model may be biased towards predicting the negative class, as it is the dominant class in the data. Another possibility is that the model's features and/or hyperparameters are not well-suited for identifying the positive class. For example, the model may not be capturing the relevant patterns and relationships in the data that are associated with the positive class. Additionally, the threshold for predicting the positive class may be set too high (in this model it is varying from 0 to 1), resulting in many false negatives. It is important to carefully consider these factors in future research and evaluate the model's performance and taking these factors into account while interpreting the results and improving the model's performance.

VII. Conclusion

A. Summary

In conclusion, the RNN algorithm did not perform very well with a means square error of 5.78 compared to the k-means clustering model which had a silhouette score of .87. Compared to all the methods, specifically method 5 proved effective, as the model has been evaluated on three different severity levels: High Severity, Moderate Severity, and Low Severity. For each severity level, the result includes the Area Under the Curve (AUC) value, which is a measure of how well the model is able to distinguish between positive and negative cases. An AUC score of 0.5 means that the model is random, while an AUC score of 1.0 means that the model perfectly separates the positive and negative classes. The accuracy score of the model is 0.999, which is very high, but may be, the result can be misleading due to imbalance in data where the majority class dominates and also this area can be further explored through extensive research and by adding more features to the model.

B. Future Enhancements

- The current model only uses the date, state, and county as features for the models. There may be other useful features that could be extracted from the data or external sources to improve the models' performance.
- The current model only evaluates the models' performance using the area under the ROC curve and confusion matrix. Other evaluation metrics could be explored to gain a better understanding of the models' strengths and weaknesses.
- Continue to experiment with the RNN model and find a way to improve the score, by perhaps adding batches.
- Implementing LSTM model by using more important features.
- It may be interesting to explore which features are most important for predicting each severity level. This could provide insight into which factors contribute most to gun violence severity.

VIII. Roles and Responsibilities

A. Summary

We met a few times in group sessions to strategize on topics and approach. Researching data, articles, and solution methods. Together we determined we should use Gun Violence data and apply LSTM, RNN, and Binary Classification to the results. We also collaborated and resolved issues and gaps experienced by each other while working and exploring applicable models and visualizations, combining codes and outlining the slides for presentation.

Priyanka explored the data to understand the severity levels defined in line the location and number of victims or affected people (i.e. killed+injured). She also did a majority of the initial research on binary classification, modeling (method 5), ROC-AUC curve and confusion matrix for result visualization and finding, which was shared and discussed with team for necessary extension and other model exploration. The modeling for method 5 and the result findings were explained by her through slides in the final presentation.

Sangeeta researched half of the supplementary papers and articles used to provide an understanding of the problem we are solving and how others have solved it. She produced and defined a third of the slides presented and helped to develop the agenda and goal of the project. Sangeeta also worked on Kmeans clustering of the result data and produced visualizations. Sangeeta also completed half of the report, including the abstract, intro, and literature sections.

Jenifer setup shared code, slides, and report document to help us work as a team while working separately on various independent pieces. She also contributed to creating a third of the slides and half of the written report. Jenifer initially worked on parsing the data to analyze participants, victims, gender, and age groups, along with state and city information. Then she applied feature selection and kmeans clustering to determine which features would have the strongest relationships and produce best results. Jenifer also researched half of the literature. Jenifer also ran some experimentations around the Binary Classification, RNN, and LSTM to try to produce better accuracy.

References

For this study, the references used are course slides provided by Dr JungYoon Kim and Wikipedia for basic information on data preprocessing and modeling that can be implemented. Below are references used in this study:

[1] *Paper 1: CNN-RNN Combined Structure for Real-World Violence Detection in Surveillance Cameras. Soheil Vosta, Kim-Choong Yow https://doi.org/10.3390/app12031021*

[2] *Paper 2: Walczak, Steven. (2021). Predicting Crime and Other Uses of Neural Networks in Police Decision Making. Frontiers in Psychology. 12. 587943. 10.3389/fpsyg.2021.587943.*

[3] *Paper 3: Real World Violence Detection in Surveillance Cameras. Liu, Gina & Wiebe, Douglas. (2019). A Time-Series Analysis of Firearm Purchasing After Mass Shooting Events in the United States. JAMA Network Open. 2. e191736. 10.1001/jamanetworkopen.2019.1736.*

[4] *Paper 4: Swedo EA, Alic A, Law RK, et al. Development of a Machine Learning Model to Estimate US Firearm Homicides in Near Real Time. JAMA Netw Open. 2023;6(3):e233413.doi:10.1001/jamanetworkopen.2023.3413*

[5] *Paper 5: Firas Saidi, Zouheir Trabelsi,A hybrid deep learning-based framework for future terrorist activities modeling and prediction,Egyptian Informatics Journal,Volume 23, Issue 3,2022,Pages 437-446,ISSN 1110-8665, https://doi.org/10.1016/j.eij.2022.04.001.*

[6] *Data: https://www.kaggle.com/datasets/jameslko/gun-violence-data from http://www.gunviolencearchive.org/*

[7] *https://scikit-learn.org/stable/modules/generated/sklearn.metrics.silhouette_score.html*