Analysis of New York City Reported Crime Data

Roua NSIRI
Higher School of Communication of Tunis
roua.nsiri@supcom.tn

Zahra GHARBI
Higher School of Communication of Tunis
zahra.gharbi@supcom.tn

***Abstract*—For the month of October 2021, the number of murders and shooting incidents continued to decline in New York City compared to last year. Murder decreased by 9.8% (37 v. 41) while shooting incidents decreased by 4.4 % (129 v. 135). There were 382 gun arrests for the month of October 2021, bringing the total number of gun arrests in 2021 to 3,808 – a 13.9% increase compared to 3,343 gun arrests year-to-date in 2020.Overall index crime in New York City increased by 11.2% in October 2021, compared with October 2020 (10,118 v. 9,095). While Burglary saw a 13.7% decrease for October 2021 (1,231 v. 1,427), Robbery increased by 15.8% (1,450 v. 1,252) and Felonious Assault increased by 13.8% (2,123 v. 1,865). Through October, overall index crime year-to-date has increased by 1.3% compared to 2020 (80,801 v. 79,772).**

The reduction in shootings and homicides during October 2021 continues a trend that has been carrying over from the summer of 2021, and coincides with the NYPD's ongoing work to build multiple long-term cases with its law enforcement partners with a focus on the drivers of violence. This is the precision-policing philosophy at work. Central to the NYPD's comprehensive public safety strategy is the curtailing of gun violence. This includes making gun arrests, and working with our criminal justice partners to address local concerns such as the interdiction of so-called ghost guns – whose proliferation presents a host of investigative challenges to law enforcement.

"The men and women of the NYPD have never wavered in their commitment to the collective public safety of all New Yorkers – as demonstrated by this ongoing, downward trend in violence," said Police Commissioner Dermot Shea. "While their devotion to service is commendable, effective crime fighting is predicated upon a collaborative effort from all aspects of the criminal justice landscape – as well as society as a whole. Additionally, our brave officers' work must be reinforced by meaningful consequences that send a consequential message to those who find themselves on the path toward criminality.".

***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

Security is an essential aspect of strengthening the roots of a country. In fact, crimes can make a significant impact on the economic growth of a country. Therefore, countries are spending a substantial amount of their (GDP) on law enforcement agencies to control crimes. Besides, advancement in technology and especially geographical information systems (GIS), assisted the researchers in presenting numerous crime detection and prediction techniques.

The enormous amount of data being available in the past few years have been a great motivation for the scientists to pursue research activities in the field of crime and criminal investigations. Studying as well as trying to understand and predict the crime trends and patterns have been the priority of the law enforcement agencies to make an effective policy by taking advantage of the historical data to lower the crime rates and make a peaceful community. Based on historical data, forecasting crimes has been a subject of interest that recently gained much attention in research, which resulted in proposing a significant number of different methods for the discovery of different aspects related to crime prediction. Crime can be considered as a location-oriented feature as some places can exhibit greater risk of crime to be committed than others due to several factors such as the degree of urbanisation and populated density, the greater rates of migration and population growth in urban populations or the variation of the demographic structure in different areas. It is an understood fact that in a particular area, no matter the size, crime is not distributed evenly, uniformly, or even randomly within that area or city.

Different types of crimes and the full consideration of the protection and safety of citizens in any society are significant components that play a vital and direct role in the quality of the life of residents.Numerous types of crimes can occur in an area with different frequencies. An area may be flagged for higher misdemeanor events while the other for felony.

The inclusion of spatial and temporal information in the crime data-sets using GIS has revolutionized the crime prediction systems. The spatio-temporal information helps the researchers to present more credible and accurate crime prediction systems that can be reliable in crime prevention.

II. Related Work

A. Smart Policing

Smart Policing represents an emerging paradigm in American policing that stresses crime reduction and promotes improvement of the evidence base for policing. Smart Policing emphasizes effectively using data and computational intelligence as well as improving analysis, performance measurement, efficiency and evaluation research along with encouraging innovation.

This introduction defines Smart Policing in historical and contemporary contexts and discusses several important and emerging characteristics in the local Smart Policing sites, namely, the need to improve the evidence base for policing, the police agency-research partnerships that are emerging in Smart Policing, the type of problems identified and approaches undertaken by the SPI sites, and future issues for Smart Policing.[3]

B. Crime Prediction

A previous work that inspired us to look more into crime prediction is crime prediction based on weather, crime data, and temporal data [1]. In the paper, the authors have proposed the utilization of weather information for crime forecasts. They employed feature selection techniques to determine the most significant features mainly the most occurred crimes and the correlation between the features, in forecasting crime calculations and rates in New York City over 5 years. They used both machine learning and deep learning techniques and provided benchmarking based on the prediction accuracy.

Another interesting work that motivated us is spatiotemporal crime forecasting using Amsterdam police Data [2]in which they focused on Crime history variables, Environmental variables, Demographic variables,Socio-economicvariables, and Proximity variables to provide more detailed and reasonable comprehension and prediction that highlights the reasons of the committed crimes.

III. METHODOLOGY

In this section, we will explain the strategy we choose to process historical data provided by **NYC Open Data Website** and then the different predictive models of monthly crime rates on the neighborhood level that we implemented to reduce crimes based on gender, age, time and location.

Figure 1 represents the pipeline of our work.

Fig. 1. Map of accidents by BORO.

A. Dataset Extraction

We downloaded the dataset on kaggle.In fact, it contains more than 7 millions complaints from 1198 to 2019. In our case we focused on 2019 complaints and we will explain in the next section how we proceeded.

B. Dataset exploration and Data cleaning

delimiting the period of the exploration in 2019. Downloading the complaints from 1 January 2019 00:00 until 31 December 2019 23:59. 2Then, dealing with null values, unnamed rows, converting dates into specific format:extracting day,month and hour of complaint in single columns based on CMPLNT_FR_DT and CMPLNT_FR_TM provided in the dataset. Droping some columns which are not important for the prediction such as PARKS_NM or STATION_NAME since most of the data provided is without name.

Encoding some features such as a column named CRM_ATPT_CPTD_CD,Which contains two categories : ATTEMPTED and COMPLETED, into 0 and 1.Convert categorical variables into dummy variables such as the following columns:KY_DT which contains the crime type: HOMICIDE, SEXCRIME, THEFTFRAUD, OTHERVIOLENT,DRUGS,OTHER. We did the same for features VIC_SEX,VIC_RACE,VIC_AGE_GROUP and BORO_NM

C. Feature extraction

Proceeding with different approaches such as correlation matrix to determine the correlation between the different features of the dataset, using map accident to locate the accident by BORO (Manhattan,Brooklyn,Queens,Bronx, Staten Island) as shown in Figure 2.

Fig. 2. Map of accidents by BORO.

Using folium plugins to show a heatmap of accidents in New York City using exact longitude and latitude from the dataset provided.

We used the **WWW (Where,When and Weather)** to have deeper insight into the distribution of the crimes,

D. Model building

Predicting the type of the crimes based on location, timing and the person description can be achieved by developing a classification solution and more precisely a supervised classification solution since the data is labeled. So we tried the following models :

- RandomForest Classification model: which is a meta estimator that fits a number of decision tree classifiers on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting and we used sklearn python library to implement the model.as a result the model gave us.
- K-nearest neighbors algorithm (kNN): Which is used for classification in our case. The class membership of the output is resulted from a plurality vote of its neighbors. In our study we set k to 30.

Fig. 3. Victims By Gender.

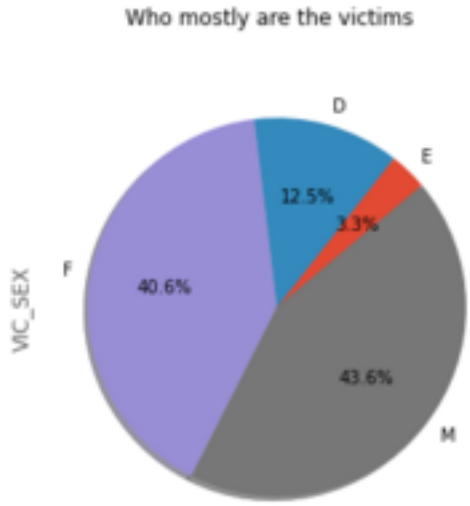

Fig. 5. Who mostly are the victims.

Fig. 4. Which day mostly these crimes happen.

- Decision Tree Classification algorithm: Which is a tree-structured classifier, where internal nodes represent the features of a data-set, branches represent the decision rules and each leaf node represents the outcome.
- Support Vector Machine (SVM): This model tries to find a hyper-plane in an N-dimensional space (N is the number of features) that distinctly classifies the data points.
- Multi-layer Perceptron classifier: This model optimizes the log-loss function using LBFGS or stochastic gradient descent.To implement this model also we used sklearn python library

E. Model Evaluation

Using different metrics to evaluate the performance of the model on the dataset. After training the previous models and evaluate them we adopted Random Forest model that gave us the following evaluation metrics values

TABLE I
MODEL EVALUATION

	RandomForest model Evaluation metrics		
	Accuracy	***Recall***	***Precision***
Results	0.72	0.72	0.73

F. Dashboard and UI Building

As to visualize the final results of our research, we created a graphical user interface in the form of a web application using web mapping techniques in which we plotted the New York map and we gave the user the possibility to choose a spot in the map for which he intends to visit, to enter his data such as his age,gender and the time in which he will be visiting the spot and then as an output we displayed the most likely type of crime that will be commited against him. Behind the scenes or as we may call it in the server-side of our application, we integrated our ready to use Machine Learning model and we created an API that takes the data provided by the user, apply the necessary transformations on it, predicts the type of crime using our model and then sends back the result to the client side of our application.

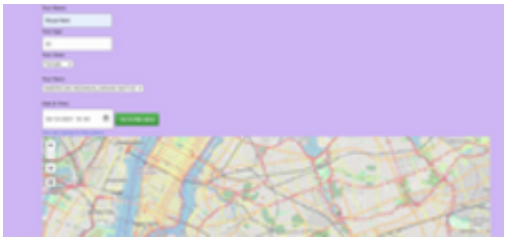

Fig. 6. User Interface.

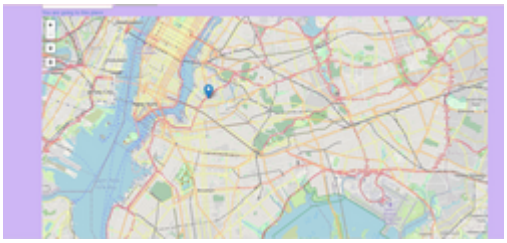

Fig. 7. Dashboard.

CONCLUSION

In this work, we used the data-set provided by the NYPD.

As a first step we explored the data in order to understand it's pattern and produce insights.Then We kept the most essential features by performing techniques of feature selection and feature extraction,

After that, we applied various Machine Learning models and compared their performances in order to predict the type of crime expected. as our random forest model gave the best results, we opted to use it as our final model.

In the future and as a next stage for this research, we want to include population density based on the location with the current features and observe if this factor plays a significant role in predicting crimes

REFERENCES

[1] Elluri, Lavanya Mandalapu, Varun Roy, Nirmalya. (2019). Developing Machine Learning Based Predictive Models for Smart Policing. 10.1109/SMARTCOMP.2019.00053.

[2] Rummens, Anneleen Hardyns, Wim Pauwels, Lieven. (2017). The use of predictive analysis in spatiotemporal crime forecasting: Building and testing a model in an urban context. Applied Geography. 86. 10.1016/j.apgeog.2017.06.011.

[3] Coldren JR, Huntoon A, Medaris M. Introducing Smart Policing: Foundations, Principles, and Practice. Police Quarterly. 2013;16(3):275-286. doi:10.1177/1098611113497042.

[4] Umair, Areeba and Sarfraz, Muhammad Shahzad and Ahmad, Muhammad and Habib, Usman and Ullah, Muhammad Habib and Mazzara, Manuel,Spatiotemporal Analysis of Web News Archives for Crime Prediction,Applied Sciences,2020;8220.doi:10.3390/app10228220.