

Prediction of Crime Rates in Chicago

WEEK 3 – PROJECT 1 – MILESTONE 2

ABDUL BASIT

Business Problem

Crime in Chicago is a very interesting topic for exploration for all kinds of reasons. Personally, I have been living in Chicago for more than 2 years now and crime here is always a topic of conversation with friends and family. Another reason is the availability of huge amounts of high-quality crime datasets open for data scientists to investigate.

In this project, I will explore crimes in Chicago from a perspective of a Chicago resident who wants to know more about the subject so he can better navigate his way through the city. Objective is to build an intelligent crime prediction model, where the model predicts different types: crime type, time of the crime, the place of crime.

Problem Statement

- How has crime in Chicago changed across years? Was 2016 really the bloodiest year in two decades?
- Are some types of crimes more likely to happen in specific locations or specific time of the day or specific day of the week than other types of crimes?

Dataset Explanation

- The Chicago Crime dataset contains a summary of the reported crimes occurred in the City of Chicago from 2005 onwards.
- Dataset has been obtained from the Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) system.

- This dataset contains attributes such as Case ID, time and date of the crime occurrence, report update date, codes for beats, communities, wards and districts, location co-ordinates of the crime occurrence and the types of crime based on Illinois Uniform Crime Reporting Code (IUCR).
- Number of Attributes: 23

Snapshot of Data [1]

	ID	Case Number	Date	Block	IUCR	Primary Type	Description	Location Description	Arrest	Domestic	...	Ward	Community Area	FBI Code	Coordinate	X	Cor
0	10000092	HY189866	03/18/2015 07:44:00 PM	047XXW OHIO ST	041A	BATTERY	AGGRAVATED: HANDGUN	STREET	False	False	...	28.0	25.0	04B	1144606.0	19	
1	10000094	HY190059	03/18/2015 11:00:00 PM	066XX S MARSHFIELD AVE	4625	OTHER OFFENSE	PAROLE VIOLATION	STREET	True	False	...	15.0	67.0	26	1166468.0	18	

Attribute Information:

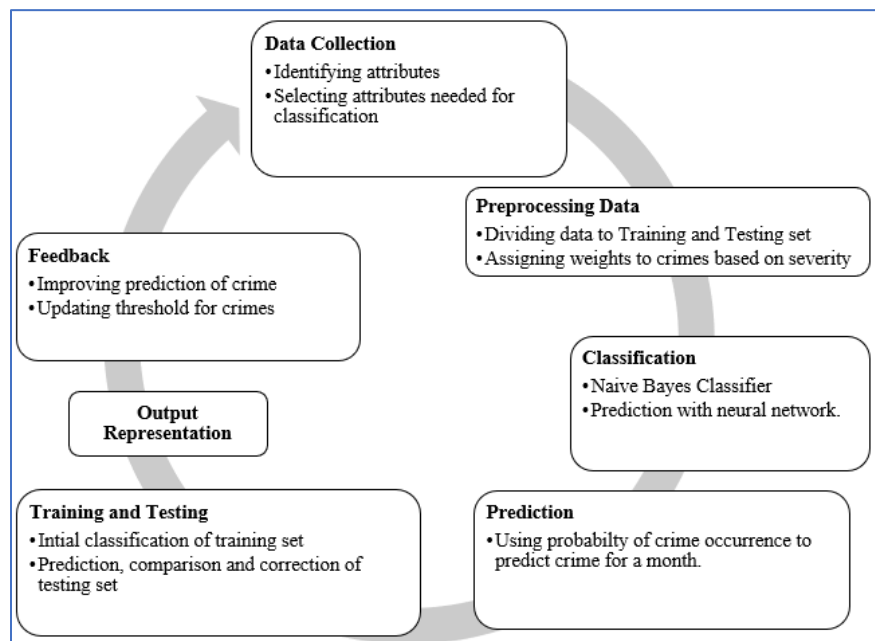
Here are some of the important fields used in the dataset:

- There are columns like ID and Case Number which helps us uniquely identify our crime record.
- Columns like Date and Year tells us when this crime happened.
- X/Y Coordinate, Latitude/Longitude, Location tells where exactly did the crime happened.
- Beat, District, Ward, Community Area tells us in which area of Chicago this crime happened.

Data source: <https://www.kaggle.com/currie32/crimes-in-chicago>

Methodology

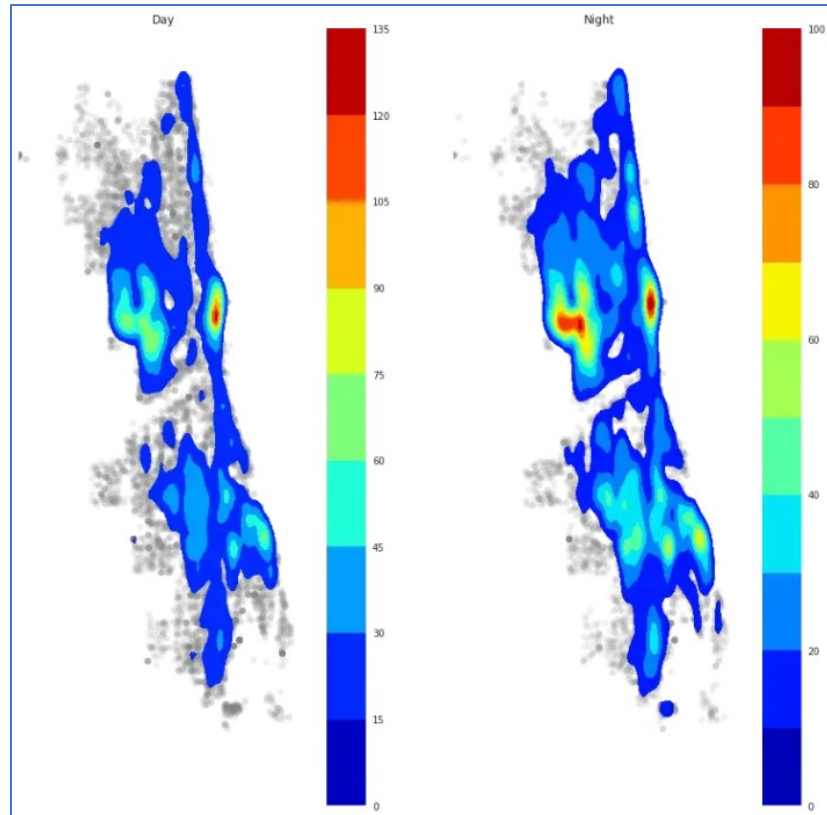
This project approaches the analysis of dataset based on dividing it into training and testing sets. The training set is used for classification based on both the Naïve Bayes classifier and the Neural Network. The performance of both the algorithms is analyzed and the best algorithm is used for further classification of the testing set. The classification of the dataset is done by dividing the time into various months of the year and the crime patterns are analyzed for each month of the year. [2]



Analysis

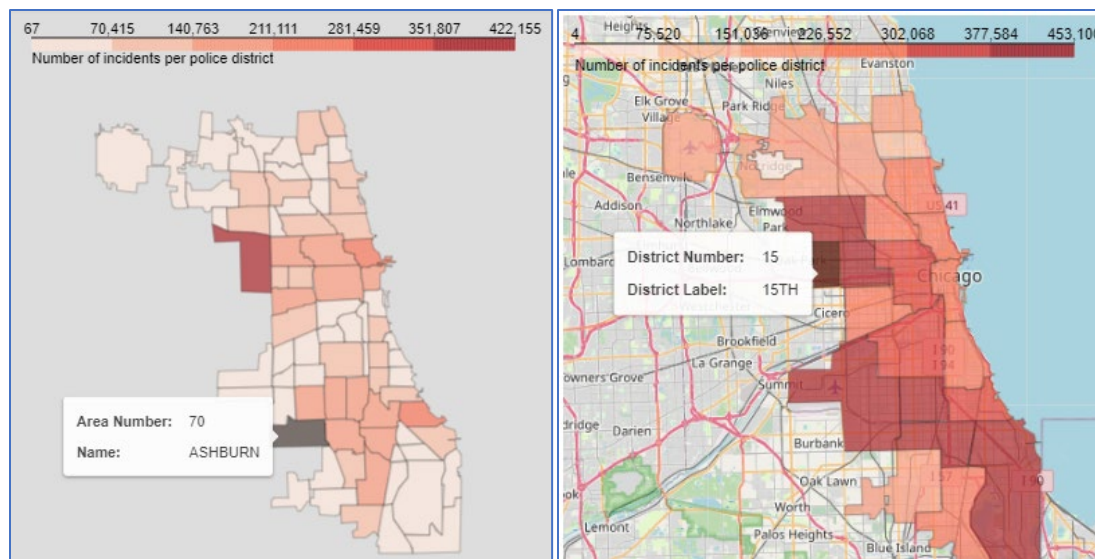
The prediction algorithm finds the probability of occurrence of a particular crime in the month of the year for a given location. This prediction can be verified within the test set to measure accuracy based on classification of actual crime occurrences. The feedback from this process is used to update the thresholds of crime prediction.[3]

CRPC



Heatmaps of crimes happening in day(left) vs crimes happening in night(right)

Finally, the output of the prediction of crime hotspots is depicted graphically in the map of City of Chicago. This prediction can be depicted for the past years for showing the algorithm's prediction efficiency.[4]



K-means clustering provides a way to group data points together in a way that minimizes differences between the data points in the same group. By applying these methods, we can take n data points and partition them into k clusters.[5]

Ethical Problems

Predictive policing relies on a large database of previous crime data and forecasts where crime is likely to occur. Since the program relies on old data, those previous arrests need to be unbiased to generate unbiased forecasts. One way crime data can be biased is if it lacks necessary information and context. Many cities including Los Angeles, New York, and Chicago have stopped using predictive policing due to violation of the ethical framework of justice and fairness because they perpetuate systemic racism through the use of biased data.

Challenges/Issues

Crime has become the focus of a challenge for both the police department and law enforcement agencies to reduce the spread of crime and use modern techniques to predict and reduce the spread of crime. Many police departments use artificial intelligence algorithms and use big data tools to help them to predict where crimes are occurring. For example, police

CRPC

departments in places like Seattle, Los Angeles, and Atlanta have experimented with predictive police programs which try to identify the geographic area where crime is likely to occur in the future. At the same time, the Chicago Police Station used an algorithm based on a heat list that tries to identify people as criminals for violent crimes or repeated abuse.

Conclusion

Crime is one of the most dangerous phenomena for any country. To reduce and nonproliferation of crime, it requires new techniques that can deal with the vast amount of data, where the data cannot be analyzed with traditional analysis techniques. Therefore, Deep Learning algorithms techniques were employed through using repeated neural network (RNN) and the employment of one of its types (LSTM). Our contribution is to build an intelligent crime prediction model, where the model predicts different types: crime type, time of the crime, the place of crime).

Future Work

Clustering data and analyzing it is an important step to many other machine learning algorithms. In particular, prediction of crime points can be achieved by multiple methods:

- Minimizing the distance between a crime occurrence and the centroid of a cluster
- Performing regression analysis on the identified clusters and fitting crimes to the best fit line

We could also implement the K-nearest neighbors' algorithm (which finds the most similar points given a specific points), with the condition that they be in the same cluster as we

CRPC

predict our data point will be in. This would provide a list of k possible crime types that are most likely to occur in Chicago on January 1 at midnight.

Recommendations

The goal was to deploy the officers based on the crime data and socio-economic indicators. Following are the recommendations:

- For particular community areas, the crime count was found to be high but the arrest rate was lower than the average Chicago arrest rate. For these areas the deployment of police officers needs to be increased. [6]
- It was given that few community areas have the high number of vacant housing units. Thus, more than average deployment of police officers is required in the weeks where we have predicted the crime count is going to be highest. Also, based on exploratory data analysis we should have more than average deployment of police officers between 12 PM and 12 AM
- Community areas where there are highest number of single parents with child. Thus, police officers need to be deployed in these areas during the weeks when the crime count is going increasing
- Based on clustering, we found out that there are community areas where there is high number of domestic abuses, high vacant housing units, high single parent with child and so on. These community areas need to monitored more than average during the weeks when the crime count is going to be high
- Special task forces that can handle extremely violent crimes should be deployed on the specific days of week when the violent crimes are high in number based on the EDA

performed. As the crime count was not predicted on daily basis, we can look at the weekly predicted numbers to deploy the special task forces

- Building new parks will help to engage people socially and might help in the reduction of crime as it is found from our analysis that community that have a greater number of parks have less crime

Questions an audience may ask

- How has the number of various crimes changed over time in Chicago?
- How have the number arrests corresponding to the crimes changed over time in Chicago?
- Are there any trends in the crimes being committed?
- Which crimes are most frequently committed?
- Which locations are these frequent crimes being committed to?
- Are there certain high crime locations for certain crimes (etc sexual harassments)?
- How has the number of certain crimes (etc homicide) changed over the years in Chicago?

References

- [1] *IRJET-International Research Journal of Engineering and Technology*.
<https://www.irjet.net/archives/V5/i9/IRJET-V5I9192.pdf>.
- [2] “Precise Event-Level Prediction of Urban Crime Reveals Signature of Enforcement Bias.”
Home, 11 Feb. 2021, <https://www.researchsquare.com/article/rs-192156/v1>.
- [3] Jain, Naman. “Hands on Machine Learning with Chicago Crime Data.” *Medium*, Medium, 2 Mar. 2020, <https://medium.com/@namanjain2050/hands-on-machine-learning-with-chicago-crime-data-3657b713d62c>.
- [4] SAFAT, WAJIHA, and SOHAIL ASGHAR. “Empirical Analysis for Crime Prediction and Forecasting Using Machine Learning and Deep Learning Techniques.”

<https://ieeexplore.ieee.org/Stamp/Stamp.jsp?Arnumber=9424589>, IEEE,
<https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9424589>.

- [5] Ravi, Anjana. "Crime Prediction and Analysis Using Big Data." *JETIR*, JETIR(Www.jetir.org), <https://www.jetir.org/view?paper=JETIR2107201>.
- [6] Yerpude, Prajakta, and Vaishnavi Gudur. "Predictive Modelling of Crime Dataset Using Data Mining." *International Journal of Data Mining & Knowledge Management Process*, 30 July 2017, <https://scholar.archive.org/work/sp3q5cavqzda3l7inieufy3noy>.

<https://www.kaggle.com/currie32/crimes-in-chicago>

<https://www.kaggle.com/femiogunbode/eda-of-crime-in-chicago-from-2012-2016/discussion>

<https://www.kaggle.com/djonafegnem/chicago-crime-data-analysis>

<http://pandas.pydata.org/>

<http://pandas.pydata.org/pandas-docs/stable/tutorials.html>