CSE3021- Social and Information Networks

J Component – Final Project Report

Fall Semester 2022-23

Crime Detection System

By

Reg. No: 20BCE1176	Name: Balaji SV
Reg. No: 20BCE1368	Name: Harish B
Reg. No: 20BCE1644	Name: Arushi Anand
Reg. No: 20BCE1806	Name: Uday Singh Shergill

B.Tech CSE

Submitted to

Dr.A.Bhuvaneswari
Assistant Professor Senior,
SCOPE, VIT, Chennai

School of Computer Science and Engineering

November 2022

Abstract

TITLE: Crime Detection System

AUTHORS:

1- Balaji SV, VIT Chennai
2- Harish B, VIT Chennai
3- Arushi Anand, VIT Chennai
4- Uday Singh Shergill, VIT Chennai

In this paper, we tested the accuracy of classification and prediction using various test sets. classification is based on Bayes' theorem. We used this algorithm to train a large number of news articles and build a model. For testing, inputting some test data into the model will give better results. Our system takes the factors/attributes of a location and our prior algorithm provides common patterns for that location. This pattern is used to create decision tree models. For each location, train these common patterns to build a model. Crime patterns cannot be static because patterns change over time. Through training, we teach the system based on specific inputs. So the system automatically learns to change crime patterns by looking at crime patterns. Crime factors also change over time. By sifting through crime data, we need to identify new factors that lead to crime. Perfect accuracy cannot be achieved as it only considers a few limiting factors. To get better prediction results, we need to find more crime attributes of the location instead of setting specific attributes.

Literature Review

1) Crime investigation and criminal network analysis using archive call detail records

Kumar, M., Hanumanthappa, M. and Kumar, T.S., 2017, January. Crime investigation and criminal network analysis using archive call detail records. In 2016 Eighth International Conference on Advanced Computing (ICoAC) (pp. 46-50). IEEE.

GIST-Criminal activity is a major problem that the local community, the government, and individuals are all concerned about. In order to predict future misbehaviour with regard to location and timing, wrongdoing forecast makes use of historical data. Nowadays, criminal cases follow one another swiftly, making it a challenging task to accurately predict next crime with superior execution. The varied incorrect expectations and locations are examined in this essay. A useful framework for predicting misbehaviour accelerates the process of dealingwith infractions. The wrongdoing prediction framework uses recorded data, analyses it usinga few dissecting procedures, and then predicts the instances and patterns of wrongdoing using any of the methodologies listed below.

RESULT-There are various reasons why there is more wrongdoing today than ever before, such as the rise in poverty, depravity, unemployment, etc. The police experts take serious action if they believe that wrongdoing has increased to consider why this is the case as well as what can be done to reduce wrongdoing locally. Examining various misbehaviour forecast and identification approaches is themain goal of this article.

2) PerpSearch: An integrated crime detection system

Ding, L., Steil, D., Hudnall, M., Dixon, B., Smith, R., Brown, D. and Parrish, A., 2009, June. PerpSearch: an integrated crime detection system. In 2009 IEEE International Conference on Intelligence and Security Informatics (pp. 161-163). IEEE.

GIST- Data mining and social network analysis are two examples of information technology that have been extensively employed by law enforcement to solve crimes. According to recent study, spatial profiling also contributes significantly to making criminal investigations easier. But in actual use, those technologies are less useful due to a lack of integration. The integrated system we suggest in this work, dubbed PerpSearch, will accept as input a specified description of a crime, including the incident's location, nature, and the physical characteristics of the suspects. Using four integrated components—geographic profiling, social network analysis, crime patterns, and physical matching—the system will evaluate these inputs in order to find suspects. After that, use a score engine toanalyse the data and provide a ranked list of people for the investigators.

RESULT-We have created a prototype version of PerpSearch, which is now being distributed to active criminal investigators in Alabama. PerpSearch is a work in progress. Modus operandi (MO) descriptions will be added to the "crime pattern" component in the future, and with MO data we may add additional components like a "serial crimes link" to automatically find similar crimes in the system and aid cops in connecting such crimes. As mentioned above, we intend to look at using genetic algorithms to enhance our outcomes.

3)Using social network analysis to prevent money laundering

Colladon, A.F. and Remondi, E., 2017. Using social network analysis to prevent money laundering. Expert Systems with Applications, 67, pp.49-58.

GIST- This study investigates the potential uses of network analytical methods in the fight against money laundering. We used actual world data by studying the primary database of a factoring business that mostly operated in Italy during a 19-month period. This database included additional helpful data about the company's clients along with the financial processes related to the factoring business. To evaluate the risk profiles of clients participating in the factoring sector, we develop prediction models based on network metrics and suggest a novel method for sorting and mapping relational data. We discover that social network measurements may be used to forecast risk profiles. The most risky social players in our sample are involved in larger or more frequent financial transactions.

RESULT-Our research supports the notion that social network measurements are crucial factors to take into account when evaluating risk profiles. We also demonstrate how using different networks enhances the single measures' capacity for informing. In all networks, degree centrality emerges as a significant predictor, with more central nodes linked to riskier profiles.

4)The application of social network analysis algorithms in a system supporting money laundering detection

Dreżewski, R., Sepielak, J. and Filipkowski, W., 2015. The application of social network analysis algorithms in a system supporting money laundering detection. Information Sciences, 295, pp.18-32.

GIST- Criminal analysis is an extremely difficult operation that requires processing enormous quantities of data from many sources, including bills and bank account activities, in order to gather information that is helpful to an investigator. Devoted software tools are required to complement human analytic capabilities, and in our earlier research, the Money Laundering Detection System was suggested as one of these tools. This paper presents the system's social network analysis component. During a money laundering case investigation, the component enables the construction and analysis of social networks using data from bank statements and the National Court Register. The technology enables the examination of relationships between people in the network and the assignment of responsibilities to network members.

RESULT-In this study, we investigated a method for detecting money laundering using social network analysis. The technology described here creates social networks using data from the National Court Register and bank transactions. We can identify actual leaders and network weaknesses by giving responsibilities to individuals from the network. Once the network's participants' responsibilities have been identified, an analysis of their linkages may be done.

5) Crime rate detection using social media of different crime locations and Twitter part-of-speech tagger with Brown clustering

Vo, T., Sharma, R., Kumar, R., Son, L.H., Pham, B.T., Tien Bui, D., Priyadarshini, I., Sarkar, M. and Le, T., 2020. Crime rate detection using social media of different crime locations and Twitter part-of-speech tagger with Brown clustering. Journal of Intelligent & Fuzzy Systems, 38(4), pp.4287-4299.

GIST-Social organisations urgently need to develop enduring remedies and deterrentpunishments to address criminal challenges. Particularly, social media is crucial for detecting crime, which greatly lowers crime rates. It would work well as the intended medium. In this research, we examine Twitter data gathered from Twitter accounts for seven distinct places using India as a case study to highlight the effectiveness of the suggested work. In order totrack criminal activity, sentiment analysis has been employed to examine user behaviour and psychology in tweets. Online conversational text uses the Twitter part-of-speech tagger, which is a Markov Model of first-order entropy. When dealing with a large collection of unmarked tweets.

RESULT-The analyses show that the findings produced agree with the actual crime rate information. We think that these kinds of studies will make it easier to identify criminalpatterns and to determine the real-time crime rate for various places.

6)Using social network analysis to study crime: Navigating the challenges of criminal justice records

Bright, D., Brewer, R. and Morselli, C., 2021. Using social network analysis to study crime: Navigating the challenges of criminal justice records. Social Networks, 66, pp.50-64.

GIST- Over the past three decades, social network analysis has been increasingly popular for studying groups of criminals involved in illegal activities including drug trafficking and terrorism. However, with this expansion has come a number of difficulties for academicsusing information taken from criminal court databases. Using social network analytic techniques and existing empirical literature that uses information from the criminal justice system, we review these issues in this paper. First, we describe and debate the various formsof data that have been employed in this literature. Second, using a thorough literature study, we document the problems that have arisen in the realm of criminal networks. We specifically draw on the recorded experiences of field researchers, including our own, and outline "archaeological" methods that subsequent researchers can use to adapt to and get through these difficulties.

RESULT-There are a few potential drawbacks to using criminal court records, most notably in terms of correctness, validity, and reliability. Such material may contain mistakes, both deliberate (such as aliases, incorrect information) and unintentional (such as transcribing problems), as well as missing information. When determining the network boundary, using criminal justice records can be particularly problematic because the boundary set by law enforcement or prosecutors may not match the boundary established by network participants. We conclude by providing several suggestions for researchers on data gathering and preparation when using criminal justice data.

7) Social network analysis as a tool for criminal intelligence: understanding its potential from the perspectives of intelligence analysts

Burcher, M. and Whelan, C., 2018. Social network analysis as a tool for criminal intelligence: Understanding its potential from the perspectives of intelligence analysts. Trends in Organized Crime, 21(3), pp.278-294.

GIST- Over the past two decades an increasing number of researchers have applied social network analysis (SNA) to various 'dark' networks. This research suggests that SNA is capable of revealing significant insights into the dynamics of dark networks, particularly the identification of critical nodes, which can then be targeted by law enforcement and security agencies for disruption. However, there has so far been very little research into whether andhow law enforcement agencies can actually leverage SNA in an operational environment and in particular the challenges agencies face when attempting to apply various network analysis techniques to criminal networks.

RESULT- This paper goes some way towards addressing these issues by drawing on qualitative interviews with criminal intelligence analysts from two Australian state law enforcement agencies. The primary contribution of this paper is to call attention to the organisational characteristics of law enforcement agencies which, we argue, can influence the capacity of criminal intelligence analysts to successfully apply SNA as much as the often citied 'characteristics of criminal networks'.

8) Investigating Organized Crime Groups: A Social Network Analysis Perspective

Tayebi, M.A. and Glasser, U., 2012, August. Investigating organized crime groups: A social network analysis perspective. In *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (pp. 565-572). IEEE.

GIST- In order to discover organised crime structures and its constituent entities, we examine co-offending networks collected from a sizable real-world crime dataset in this research. In applying social network research tools and data mining techniques, we concentrate on systematic and analytical factors. Our research aims to advance computational co-offending network analysis as a useful technique for obtaining data on criminal organisations from sizable real-world crime datasets, particularly data from police-reported crimes. In our opinion, it would be nearly impossible to gather such data using conventional techniques for crime analysis. We offer an experimental assessment of our strategy with encouraging findings.

RESULT-Our analytic approach provides important insights into the ways in which co- offending networks shape and af- fect criminal behavior. Albeit, it should be noted that co- offending networks do not necessarily identify all individuals of an organization, simply because those operating in the background, who often give the orders, may not be visible inthe data. For obtaining a more holistic picture of criminal organizations, one may combine police-reported crime data with data from intelligence agencies. Further, the approach taken here primarily concentrates on organized crime groups with dense member relationships, which is not always the case, especially not for certain forms or criminal networks.

9) An advanced network visualization system for financial Crime Detection

W. Didimo, G. Liotta, F. Montecchiani and P. Palladino, "An advanced network visualization system for financial crime detection," 2011 IEEE Pacific Visualization Symposium*, 2011, pp. 203-210, doi: 10.1109/PACIFICVIS.2011.5742391.*

GIST- We introduce VISFAN, a brand-new method for the visual examination of financial activity networks. It provides the analyst with useful tools for identifying financial crimes including fraud and money laundering.

In contrast to other methods and systems now in use for

The following are the primary innovations that VISFAN delivers in its research of criminal networks: For the visual exploration of complex networks,

(i) it combines bottom-up and top-down interaction paradigms

(ii) it enables the mixing of automatic and manual clustering;

(iii) it enables the analyst to interactively adjust the dimensions of each cluster region and apply various geometric constraints to the layout.

In addition to clustering, VISFAN uses a number of other technologies for social network research. For instance, it calculates a number of indices to gauge each actor's importance in the network. We provided a solution for the visual examination of financial data called VISFAN. Networks of activities meant to help people find financial crimes.

RESULT-We intend to expand VISFAN in the near future to analyze other types of financial activity networks, such as those resulting from regardless of whether SARs or STRs exist, from all transactions made by a bank institution's customers. To manage this kind of larger networks, from an algorithmic perspective, it might be necessary to incorporate geometric constraints into multi-scale force-directed drawing algorithms. Additionally, it would be quite interesting to upgrade our system to include both methods for automatically extracting data from documents and alternative views (to graph views), such as those found in JIGSAW. Finally,further experimental studies involving FIU analysts will be helpful for a better validation of our system.

10) Criminal Network Community Detection Using Graphical Analytic Methods: A Survey

Sangkaran, T., Abdullah, A. and JhanJhi, N.Z., 2020. Criminal network community detection using graphical analytic methods: A survey. EAI Endorsed Transactions on Energy Web, 7(26), pp.e5-e5.

GIST- As network analysis became more and more prominent among professionals and scholars, it drew a large number of researchers to study criminal networks. We have provideda thorough analysis of community discovery techniques using graph analysis in this work. Both the idea of community and the techniques for finding communities inside a network were thoroughly examined. A complete overview of detection algorithms that have been created, used, and assessed by numerous authors in social network analysis was also covered, along with a broad categorization of community detection techniques. The strength and limitations of criminal network community detection approaches were shown by a thorough assessment of studies based on the identification of community in a criminal network. Therefore, it is clear from this study that additional research is needed and necessary in orderto further grow this research area.

RESULT-Criminal networks analysis has attracted several numbers of researchers as network analysis gained its popularity among professionals and researchers. In this study, we have presented a comprehensive review of community detection methods based on graph analysis. The concept of community was vividly discussed as well as the algorithms for detecting communities within a network. Broad categorization of community detection algorithms was also discussed as well as a thorough review of detection algorithms which has been developed, implemented and evaluated by several authors in social network analysis. Most importantly, a strict review of researches based on the detection of a community in a criminal network was carried out revealing the strength and limitations of criminal network community detection methods. Thus, it becomes obvious through this study that more research activities is necessary and expected in order to further grow this research area.

Proposed Work

Algorithm used:

1. PRESENTED DATA SET COMPRISES DOCUMENTS FROM DISTINGUISHED CLASSES.

2. Calculate the prior probability of class A by multiplying the number of class A objects by the total number of repeat objects across all classes.

3. FIND Ni, THE TOTAL NUMBER OF WORD FREQUENCY OF EACH CLASS.

4. FIND CONDITIONAL PROBABILTY OF KEYWORD OCCURRENCE GIVEN N A CLASS
P(WORD1/CLASS(A)) =WORDCOUNT/Ni(A)
REPEAT FOR ALL KEYWORDS.

5. AVOID ZERO FREQUENCY PROBLEMS BY APPLYING UNIFORM DISTRIBUTION

6. CLASSIFY A NEW DOCUMENT C BASED ON PROBABILITY P(C/W)
P(A/W) = P(A)*P(WORD1/CLASS A)*P(WORD2/CLASS A).......... *P(WORDn/CLASS A)
REPEAT FOR ALL CLASSES

7. ASSIGN DOCUMENT TO CLASS THAT HAS HIGHER PROBABILITY.

Implementation Details:

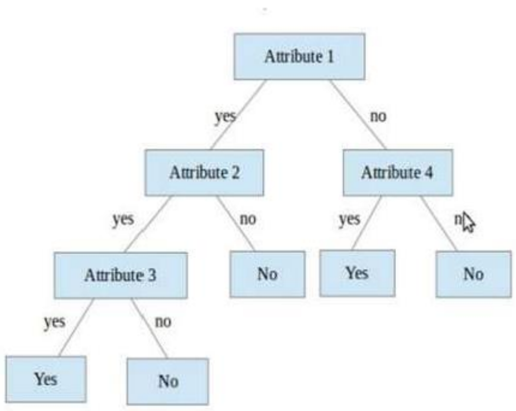
A. We gather data during the data gathering stage from a variety of websites, including news websites, blogs, social networking platforms, RSS feeds, etc.

B. A naive Bayes classifier is a probabilistic classifier that, given an input, returns a probability distribution over all classes rather than providing a single output. Algorithms classify news articles into crime types that best fit them. From the classification, "What is the probability that criminal document D belongs to a particular class C?"

The advantage of using the Naive Bayes classifier is that it is simple and converges faster than logistic regression. It differs from other algorithms in that it is easy to implement and has high performance compared to other algorithms like SVM (Support Vector Machine) which require a lot of memory. SVMs also run slower as the training set size increases.

C. The third step is the phase where we have to find trends and patterns in criminal activity.

D. We are applying the decision tree approach for prediction. A decision tree is similar to a graph in that each branch reflects the result of a test, and each internal node represents a test on an attribute. The decision tree's key benefit is that it is easy to comprehend and interpret. Other benefits include its ability to handle enormous data amounts and robustness. This characteristic enables the algorithms to choose variables more wisely.



E.A heat map, which displays amount of activity and often uses darker colours to indicate low activity and brighter colours to indicate high activity, can be used to visually represent the crime-prone locations.

Heat maps provide the following advantages over other types of representational mechanisms:
numeric and category-based colour images;
gradient colour range;
the ability to analyse only the data we are interested in;
and the automatic erasure of out-of-range data.

Dataset:

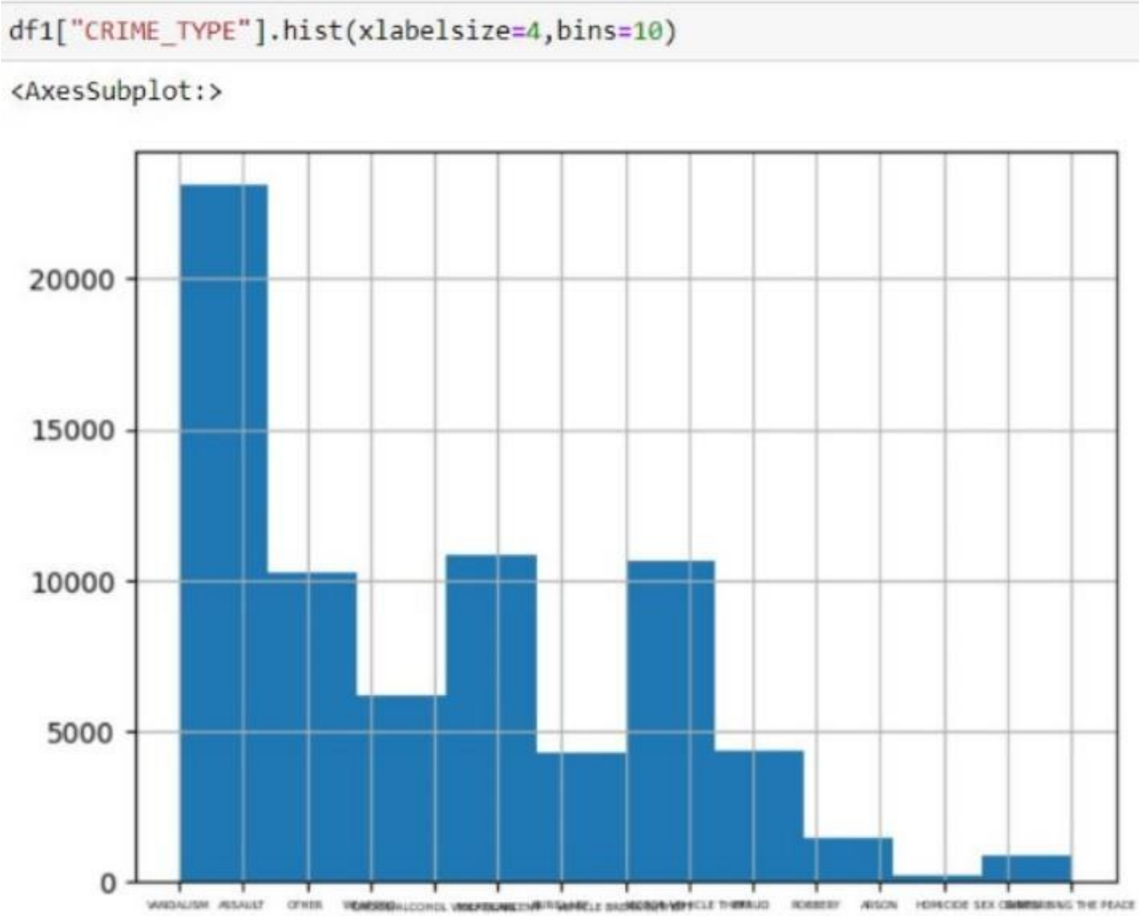
https://data.louisvilleky.gov/datasets/louisville-metro-ky-crime-data-2021/explore

Code:

```
import numpy as np
import pandas as pd
df1 = pd.read_csv('crime.csv')
df1
from sklearn import preprocessing
sti = preprocessing.LabelEncoder()
df = df1.apply(sti.fit_transform)
df
featured_cols = ['CRIME_TYPE', 'LMPD_DIVISION', 'LMPD_BEAT', 'PREMISE_TYPE',
'City', 'ZIP_CODE']
x = df[featured_cols]
```

```
y = df.UOR_DESC
x
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.3)
from sklearn.naive_bayes import GaussianNB
rain = GaussianNB()
rain.fit(x_train, y_train)
x_test
y_pred = rain.predict(x_test)
result = pd.DataFrame({'Actual': y_test, 'Predicted': y_pred})
result
print(len(pd.unique(df['CRIME_TYPE'])))
print(len(pd.unique(df['LMPD_DIVISION'])))
print(len(pd.unique(df['LMPD_BEAT'])))
print(len(pd.unique(df['PREMISE_TYPE'])))
print(len(pd.unique(df['City'])))
print(len(pd.unique(df['ZIP_CODE'])))
pd.DataFrame({'CRIME_TYPE': [8], 'LMPD_DIVISION': [5], 'LMPD_BEAT': [54],
'PREMISE_TYPE': [40], 'City': [476], 'ZIP_CODE': [52]})
otpt = rain.predict(inpt)
otpt[0]
df1[df['UOR_DESC'] == otpt[0]]['UOR_DESC'].iloc[0]
```

Crime Type Distribution Analysis

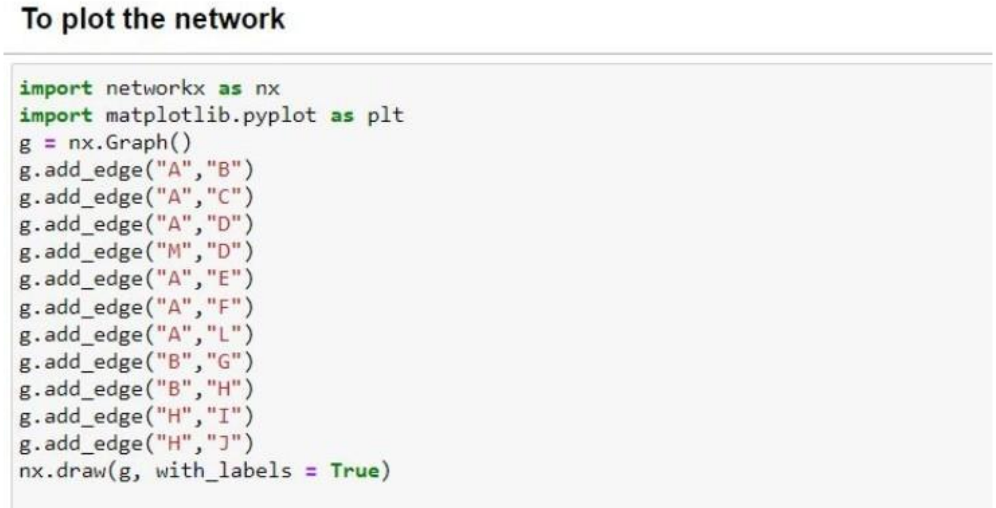


```
df1["CRIME_TYPE"].hist(xlabelsize=4,bins=10)
```

```
<AxesSubplot:>
```

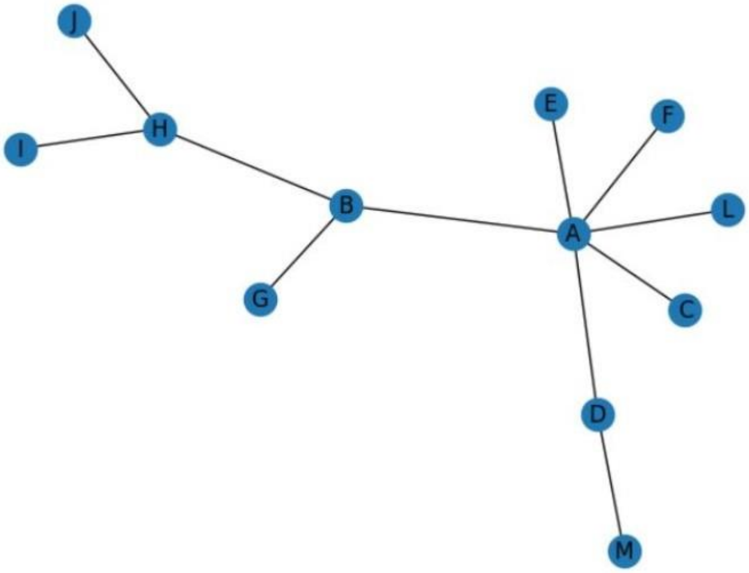
This graph basically gives us an distribution bar graph of the types of cirmes we have taken into consideration for this project. For example vandalasim, assault etc. This is to study the type of crimes and its distributions. We check for variations in the distribution of the crime type. So from the above graph we can say that vandalism has the highest frequency.

How to prevent from Crime using Network Analysis:

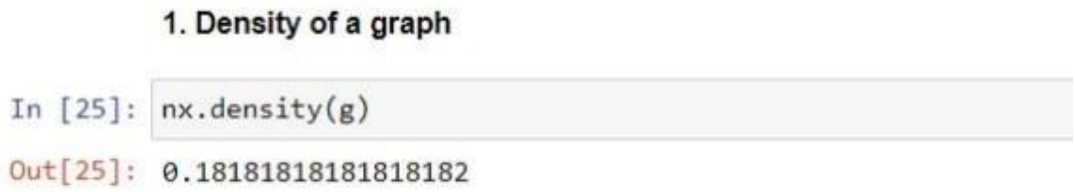


To plot the network

```
import networkx as nx
import matplotlib.pyplot as plt
g = nx.Graph()
g.add_edge("A","B")
g.add_edge("A","C")
g.add_edge("A","D")
g.add_edge("M","D")
g.add_edge("A","E")
g.add_edge("A","F")
g.add_edge("A","L")
g.add_edge("B","G")
g.add_edge("B","H")
g.add_edge("H","I")
g.add_edge("H","J")
nx.draw(g, with_labels = True)
```



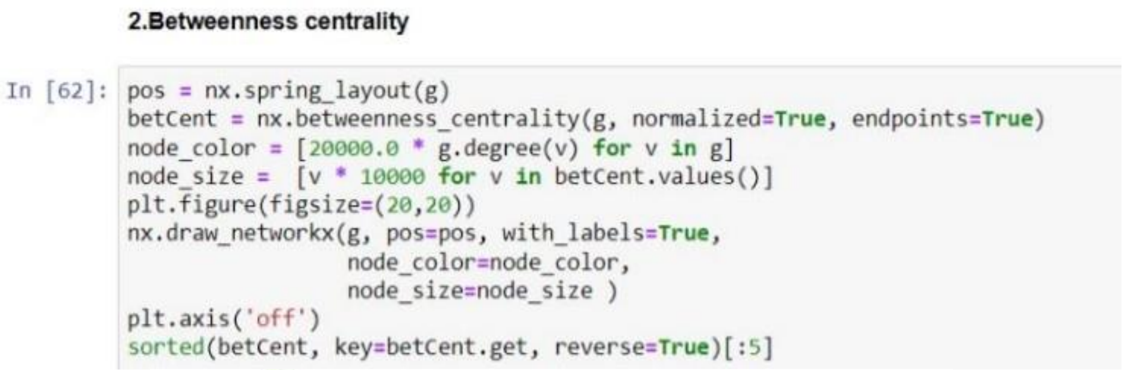
Say we have a gated community or colony of houses and they are connected as shown above. So above we have a network of houses connected with the nearest neighbours. If say crime happens in any one of the house say “F”, we can send a signal to “A” which in turn sends an alert message to its neighbours making the other houses in the network safe or keeping them prepared.



1. Density of a graph

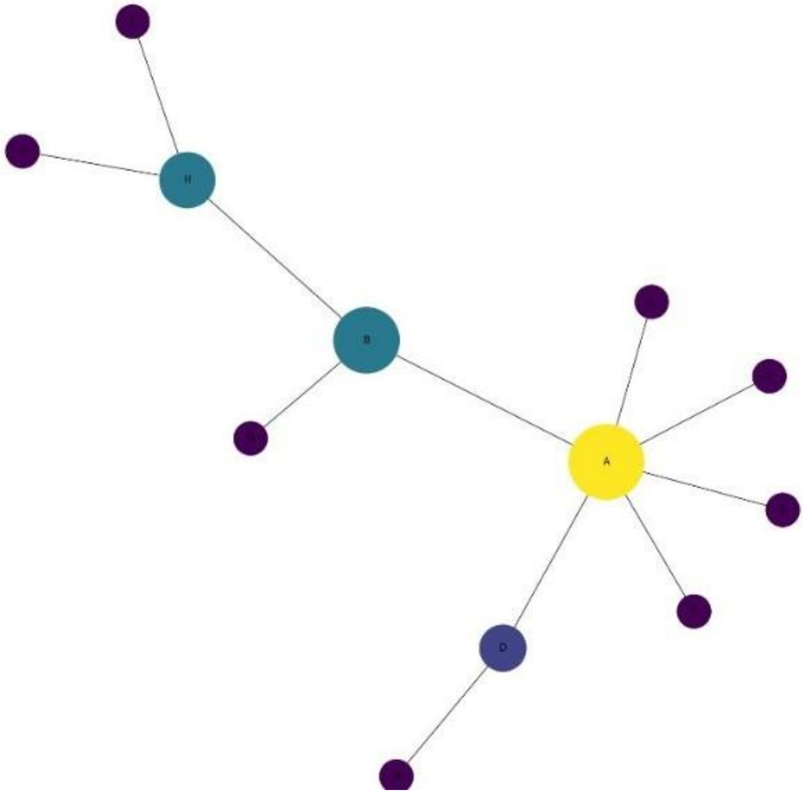
```
In [25]: nx.density(g)

Out[25]: 0.18181818181818182
```

So here we can get the density of the colony/neighbourhood. Using which we can decide which houses are more close together and where is the possibility of a crime to happen, as crime happens only in a empty or crowd less places, so this density measures can give us an approximate measure of possibility of the crime to happen in this neighbourhood.

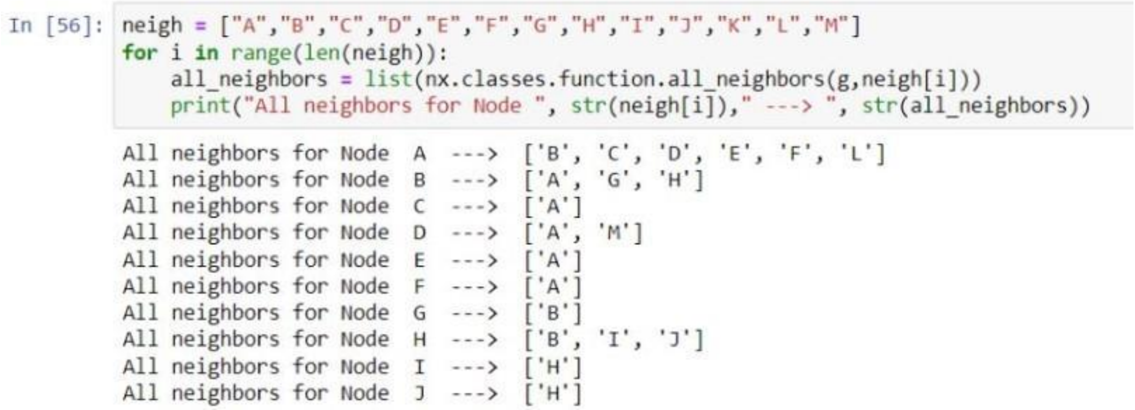


2.Betweenness centrality

```
pos = nx.spring_layout(g)
betCent = nx.betweenness_centrality(g, normalized=True, endpoints=True)
node_color = [20000.0 * g.degree(v) for v in g]
node_size =  [v * 10000 for v in betCent.values()]
plt.figure(figsize=(20,20))
nx.draw_networkx(g, pos=pos, with_labels=True,
                 node_color=node_color,
                 node_size=node_size )
plt.axis('off')
sorted(betCent, key=betCent.get, reverse=True)[:5]
```



Using betweeness centrality, this will give us an exact network having important nodes. These important node will help us to send us alert signals to all the neighbouring nodes. Say if there happened a crime at house “D”, it will first send the signal to most important node(house) or house with many neighbours that is “A” so that that house will send an alert signals to its neighbouring houses immediately, as “A” has many neighbours asking “A” to send an alert to all the other neighbours will be more effective way to alert all other houses and is time friendly. Similarly the loop goes on until all the houses in the neighbourhood gets an alert.



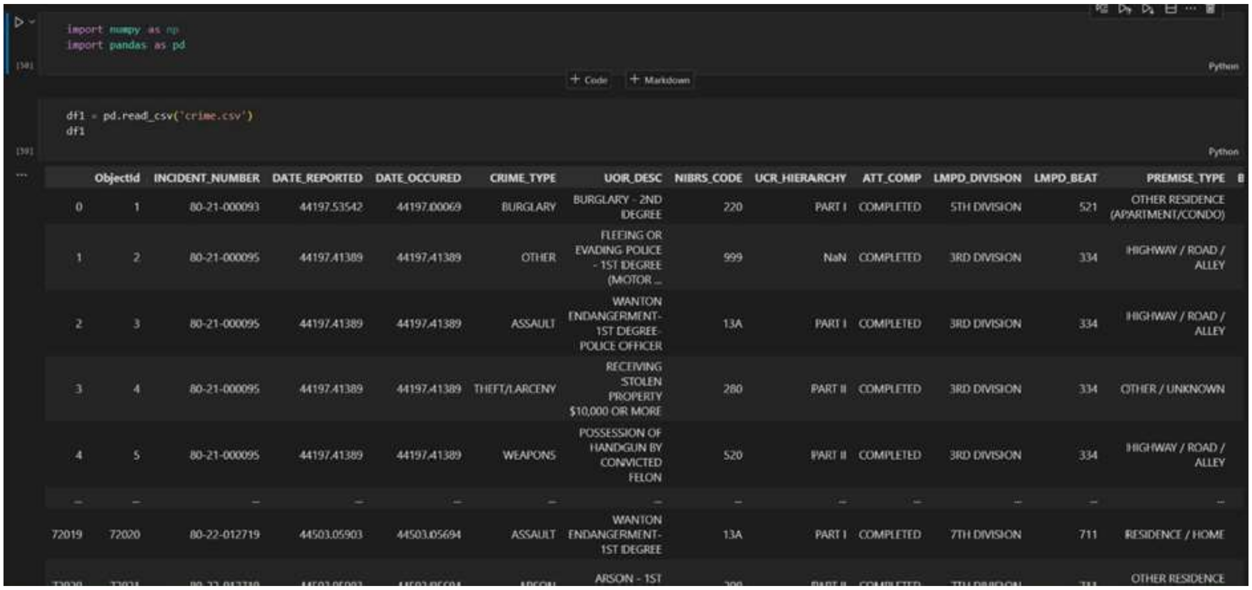
```
In [56]: neigh = ["A","B","C","D","E","F","G","H","I","J","K","L","M"]
         for i in range(len(neigh)):
             all_neighbors = list(nx.classes.function.all_neighbors(g,neigh[i]))
             print("All neighbors for Node ", str(neigh[i])," ---> ", str(all_neighbors))
```

```
All neighbors for Node  A  --->  ['B', 'C', 'D', 'E', 'F', 'L']
All neighbors for Node  B  --->  ['A', 'G', 'H']
All neighbors for Node  C  --->  ['A']
All neighbors for Node  D  --->  ['A', 'M']
All neighbors for Node  E  --->  ['A']
All neighbors for Node  F  --->  ['A']
All neighbors for Node  G  --->  ['B']
All neighbors for Node  H  --->  ['B', 'I', 'J']
All neighbors for Node  I  --->  ['H']
All neighbors for Node  J  --->  ['H']
```

So here we have an list of all the neighbours of each house in the community. This will also give us a clear picture of the network and the important node in the network.

So doing such we can prevent and stop the occurrence of crimes in any locality. As sending alerts to all other neighbours can help themselves to catch the burglar or criminal before he escapes. Doing so, will decrease the rate of crime in the particular locality or neighbourhood.

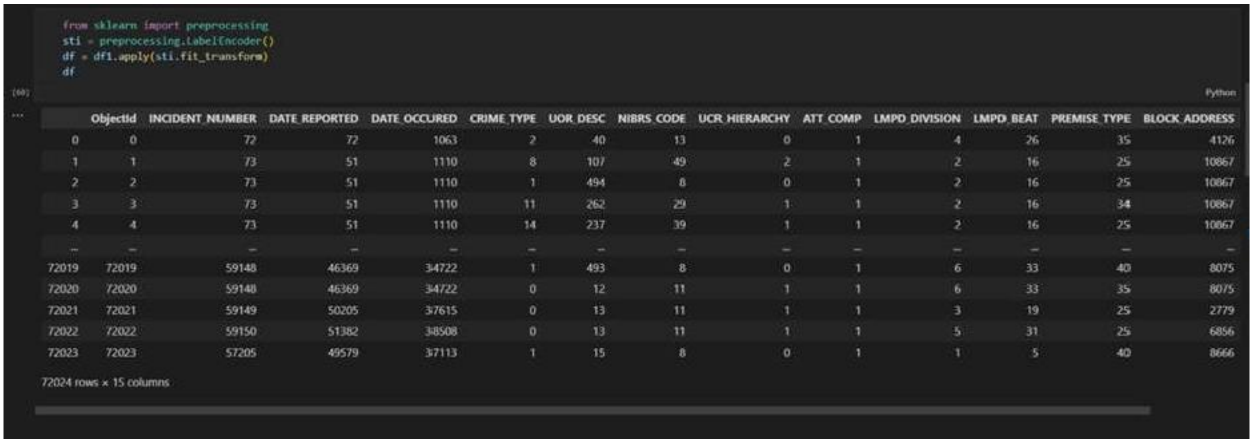
Results



```python
import numpy as np
import pandas as pd
```

```python
df1 = pd.read_csv('crime.csv')
df1
```

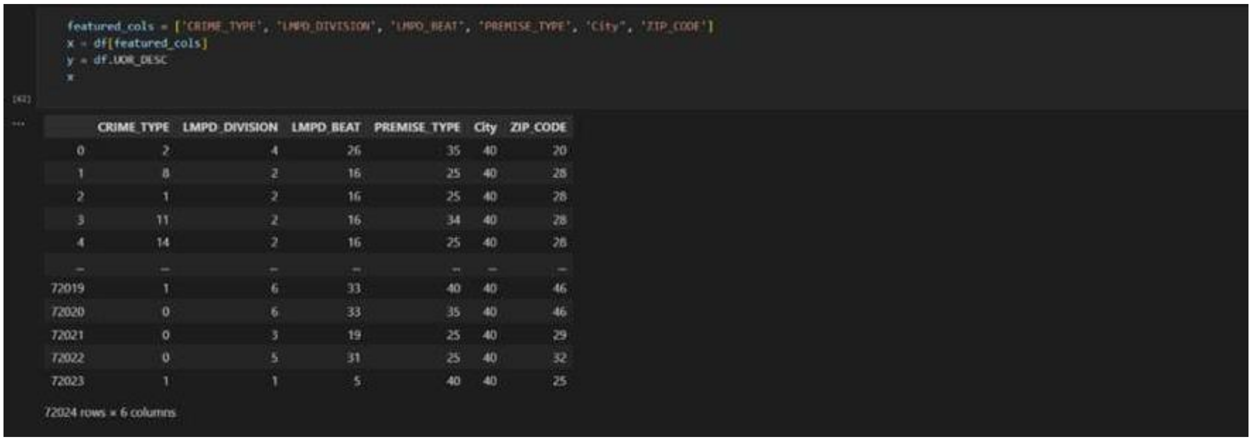
	ObjectId	INCIDENT_NUMBER	DATE_REPORTED	DATE_OCCURED	CRIME_TYPE	UOR_DESC	NIBRS_CODE	UCR_HIERARCHY	ATT_COMP	LMPD_DIVISION	LMPD_BEAT	PREMISE_TYPE	B
0	1	80-21-000093	44197.53542	44197.00069	BURGLARY	BURGLARY - 2ND DEGREE	220	PART I	COMPLETED	5TH DIVISION	521	OTHER RESIDENCE (APARTMENT/CONDO)	
1	2	80-21-000095	44197.41389	44197.41389	OTHER	FLEEING OR EVADING POLICE - 1ST DEGREE (MOTOR ...	999	NaN	COMPLETED	3RD DIVISION	334	HIGHWAY / ROAD / ALLEY	
2	3	80-21-000095	44197.41389	44197.41389	ASSAULT	WANTON ENDANGERMENT-1ST DEGREE-POLICE OFFICER	13A	PART I	COMPLETED	3RD DIVISION	334	HIGHWAY / ROAD / ALLEY	
3	4	80-21-000095	44197.41389	44197.41389	THEFT/LARCENY	RECEIVING STOLEN PROPERTY $10,000 OR MORE	280	PART II	COMPLETED	3RD DIVISION	334	OTHER / UNKNOWN	
4	5	80-21-000095	44197.41389	44197.41389	WEAPONS	POSSESSION OF HANDGUN BY CONVICTED FELON	520	PART II	COMPLETED	3RD DIVISION	334	HIGHWAY / ROAD / ALLEY	
...	...	...	...	...	...	...	...	...	...	...	...	...	
72019	72020	80-22-012719	44503.05903	44503.05694	ASSAULT	WANTON ENDANGERMENT-1ST DEGREE	13A	PART I	COMPLETED	7TH DIVISION	711	RESIDENCE / HOME	
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	ARSON - 1ST	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	OTHER RESIDENCE	



```python
from sklearn import preprocessing
sti = preprocessing.LabelEncoder()
df = df1.apply(sti.fit_transform)
df
```

	ObjectId	INCIDENT_NUMBER	DATE_REPORTED	DATE_OCCURED	CRIME_TYPE	UOR_DESC	NIBRS_CODE	UCR_HIERARCHY	ATT_COMP	LMPD_DIVISION	LMPD_BEAT	PREMISE_TYPE	BLOCK_ADDRESS
0	0	72	72	1063	2	40	13	0	1	4	26	35	4126
1	1	73	51	1110	8	107	49	2	1	2	16	25	10867
2	2	73	51	1110	1	494	8	0	1	2	16	25	10867
3	3	73	51	1110	11	262	29	1	1	2	16	34	10867
4	4	73	51	1110	14	237	39	1	1	2	16	25	10867
...	...	...	...	...	...	...	...	...	...	...	...	...	...
72019	72019	59148	46369	34722	1	493	8	0	1	6	33	40	8075
72020	72020	59148	46369	34722	0	12	11	1	1	6	33	35	8075
72021	72021	59149	50205	37615	0	13	11	1	1	3	19	25	2779
72022	72022	59150	51382	38508	0	13	11	1	1	5	31	25	6856
72023	72023	57205	49579	37113	1	15	8	0	1	1	5	40	8666

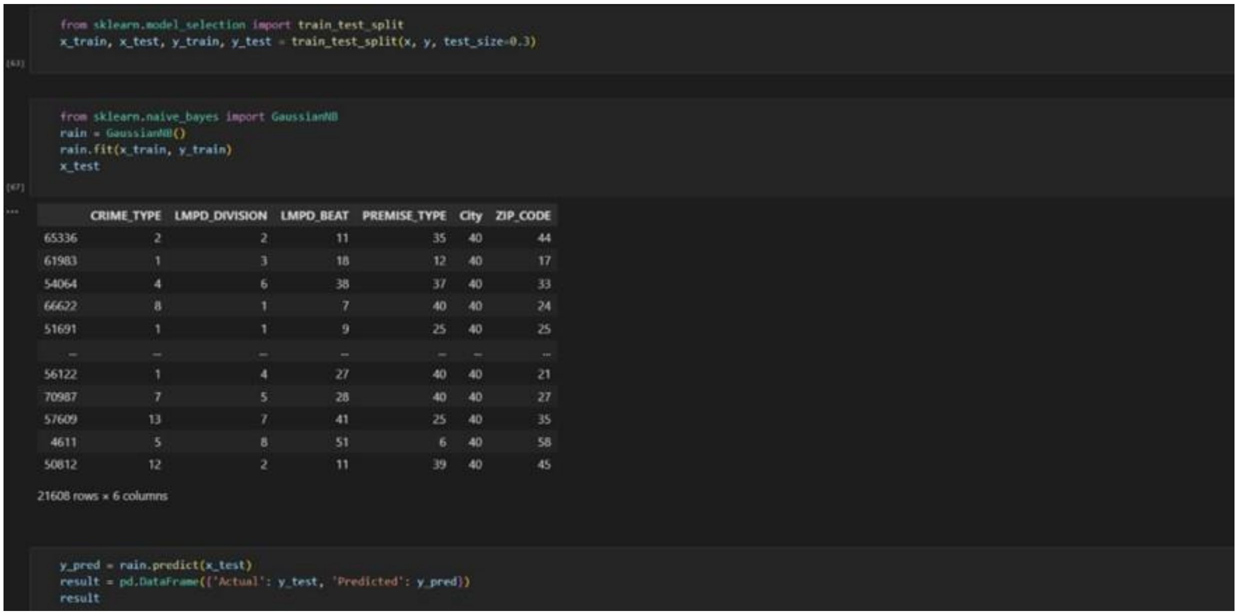
72024 rows × 15 columns



```python
featured_cols = ['CRIME_TYPE', 'LMPD_DIVISION', 'LMPO_BEAT', 'PREMISE_TYPE', 'City', 'ZIP_CODE']
x = df[featured_cols]
y = df.UOR_DESC
x
```

	CRIME_TYPE	LMPD_DIVISION	LMPD_BEAT	PREMISE_TYPE	City	ZIP_CODE
0	2	4	26	35	40	20
1	8	2	16	25	40	28
2	1	2	16	25	40	28
3	11	2	16	34	40	28
4	14	2	16	25	40	28
...	...	...	...	...	...	...
72019	1	6	33	40	40	46
72020	0	6	33	35	40	46
72021	0	3	19	25	40	29
72022	0	5	31	25	40	32
72023	1	1	5	40	40	25

72024 rows × 6 columns



```
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.3)
```

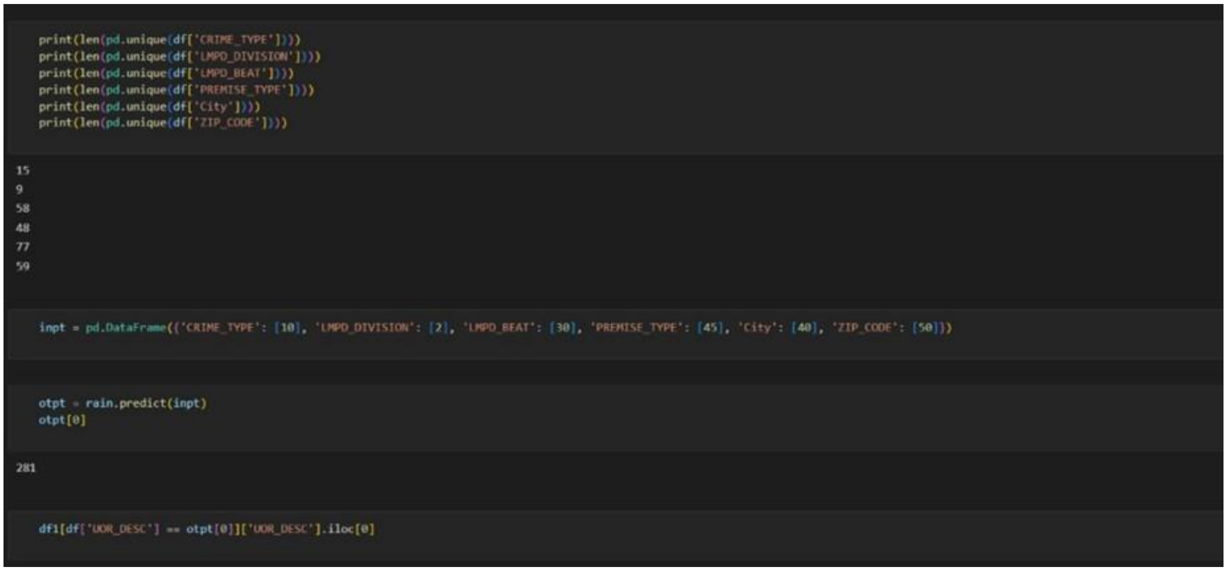
[63]

```
from sklearn.naive_bayes import GaussianNB
rain = GaussianNB()
rain.fit(x_train, y_train)
x_test
```

[67]

```
       CRIME_TYPE  LMPD_DIVISION  LMPD_BEAT  PREMISE_TYPE  City  ZIP_CODE
65336           2              2         11            35    40        44
61983           1              3         18            12    40        17
54064           4              6         38            37    40        33
66622           8              1          7            40    40        24
51691           1              1          9            25    40        25
  ...         ...            ...        ...           ...   ...       ...
56122           1              4         27            40    40        21
70987           7              5         28            40    40        27
57609          13              7         41            25    40        35
 4611           5              8         51             6    40        58
50812          12              2         11            39    40        45

21608 rows × 6 columns
```

```
y_pred = rain.predict(x_test)
result = pd.DataFrame({'Actual': y_test, 'Predicted': y_pred})
result
```



```
print(len(pd.unique(df['CRIME_TYPE'])))
print(len(pd.unique(df['LMPD_DIVISION'])))
print(len(pd.unique(df['LMPD_BEAT'])))
print(len(pd.unique(df['PREMISE_TYPE'])))
print(len(pd.unique(df['City'])))
print(len(pd.unique(df['ZIP_CODE'])))
```

```
15
9
58
48
77
59
```

```
inpt = pd.DataFrame({'CRIME_TYPE': [10], 'LMPD_DIVISION': [2], 'LMPD_BEAT': [30], 'PREMISE_TYPE': [45], 'City': [40], 'ZIP_CODE': [50]})
```

```
otpt = rain.predict(inpt)
otpt[0]
```

```
281
```

```
df1[df['UOR_DESC'] == otpt[0]]['UOR_DESC'].iloc[0]
```

CONCLUSION

In this paper, we tested the accuracy of classification and prediction using various test sets. classification is based on Bayes' theorem. We used this algorithm to train a large number of news articles and build a model. For testing, inputting some test data into the model will give better results. Our system takes the factors/attributes of a location and our prior algorithm provides common patterns for that location. This pattern is used to create decision tree models. For each location, train these common patterns to build a model. Crime patterns cannot be static because patterns change over time. Through training, we teach the system based on specific inputs. So, the system automatically learns to change crime patterns by looking at crime patterns. Crime factors also change over time. By sifting through crime data, we need to identify new factors that lead to crime. Perfect accuracy cannot be achieved as it only considers a few limiting factors. To get better prediction results, we need to find more crime attributes of the location instead of setting specific attributes.

REFERENCES

1. P. Duijn, P.P.H.M. Klerks
Social network analysis applied to criminal networks: recent developments in Dutch law enforcement
Networks and Network Analysis for Defence and Security (2014), pp. 121-159
In Lecture Notes in Social Networks

2. A. Framis
Illegal networks or criminal organizations: structure, power, and facilitators in cocaine trafficking structures
Carlo Morselli (Ed.), Crime and Networks, Routledge, New York (2014), pp. 131-147

3. M.E. Beare et al. Money Laundering in Canada: Chasing Dirty And Dangerous Dollars 2007

4. H. Chen et al. Crime data mining: a general framework and some examples IEEE Comput.(2004)

5. J. Dajda et al. Component-based architecture for systems, services and data integration in support for criminal analysis J. Telecommun. Inform. Technol.(2012)