CRIME PREDICTION USING WEB MAPS

BHAVNA SAINI
School of computing & IT
Department of Information Technology
Manipal University Jaipur.
bhavnasaini10@gmail.com

GINIKA MAHAJAN
School of computing & IT
Department of Information Technology
Manipal University Jaipur.
ginika.mahajan@jaipur.manipal.edu

T. HITESH KUMAR REDDY
School of computing &IT
B. Tech(CSE)
Manipal University Jaipur.
hitesh95reddy@gmail.com

Abstract: *Crimes are one of the dangerous and important common social problem facing by every country. Crimes effect the quality of life, economic growth and reputation of a country. There has been enormous increase in crimes in the recent times. So, the law enforcements need to take the preventive measures to reduce the crimes, hence they need the advanced systems and new approaches for improving the crime analytics for protecting their communities. Accurate real-time crime prediction helps to reduce the crime rate but remains challenging problem for the scientific community. Crime occurrences depend on many complex factors. In this work, we adapt the various visualizing techniques and machine learning algorithms to predict the crime distribution over an area. First step we process the raw datasets and visualize these datasets based on the need. The machine learning algorithms extract the knowledge out of these large datasets and discover the hidden relationships among the data by using the mathematical equations and helps in prediction and prevention of crime.*

INTRODUCTION: *Crimes are common social problems that effects the quality of life, economic growth and reputation of a country. Crimes are the one of the effecting factor that determines the people to move to a new place, to roam at right time, to avoid the areas that have risk factors. Crimes damage the image of a community. Crimes effect the economy by placing the financial burdens on government because of need of increase in police forces, courts etc. Crimes are increasing drastically and there is a need for reducing these. The latest figures show an 13% increase in all police-recorded offences across England and Wales, and even greater rises for violent offences including knife crime, sexual offences and violence against the person. The crime figures show an underlying 8% rise in the murder rate, an increase of 46 victims, with 629 homicides recorded in the 12 months to June, excluding the 35-people killed in the London and Manchester terrorist attacks. These figures can be reduced if we can predict and take preventive measures. Crime rates can be significantly reduced by the real-time crime forecasting and mass surveillance, which are helpful in saving lives that is the most valuable thing. Previous Crime data will be helpful in reducing and predicting the crimes. Crime Analysis helps in Crime Forecasting. Crime Analysis helps to study the crime reports and identifies the emerging patterns, series and trends as quickly as possible. Crime Analysis helps in preparing statistics, queries and maps on demand. Crime Analysis helps to see if a crime fits in a certain known pattern or a new pattern is necessary. Crimes can be predictable because the criminals tend to operate in their own comfort zones. Once successful they try to replicate the crime under similar circumstances. The happening of crime depends on several factors such as intelligence of a criminal, security of a location, time etc. The criminals look for similar location and time for attempting crime for next time. Once a criminal is successful, he comes to the vulnerabilities and tries to use the same mechanism he followed for the previous times. Although it may not be true for all the cases but frequent times there is a possibility of happening the same. This makes the crimes predictable. This work provides different visualization techniques that shows the trend of crimes and various*

ways that can predict the crimes using machine learning algorithms.

METHODOLOGY:

I.DATASET PREPROCESSING

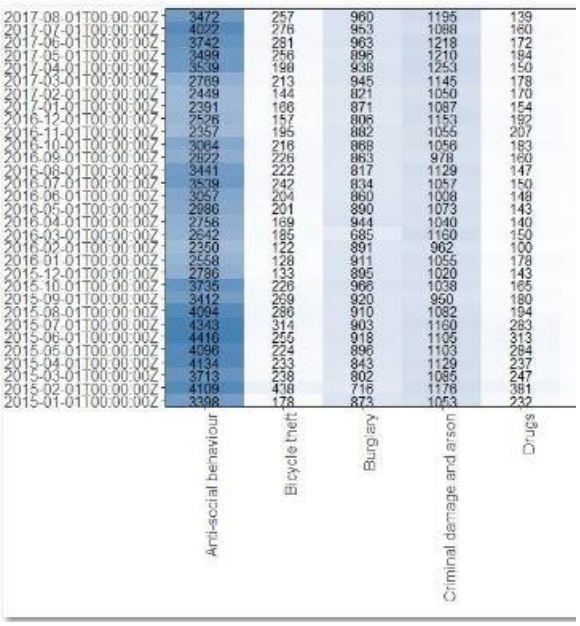
The dataset used for the work is real and authentic. The data is acquired from the official site of the U.K. police department. The data set contains a total of 11 attributes out of which 5 attributes were considered for the study, they are crime type, location, date, latitude and longitude. For the training dataset we consider the history crimes from the year 2015-17.

II.DATA VISUALIZATION:

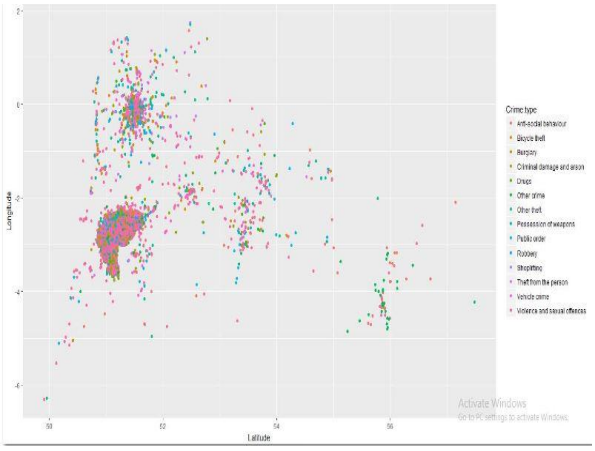
Data visualization is an art and science. It is a form of visual communication. It involves creation and study of visual representation of data. The primary goal of data visualization is to communicate data clearly and effectively via statistical graphics and plots. The effective visualization helps us to analyze and reason about data and evidence. As visualization compared to only data affect the general understanding, we explored the dataset to show various visualizations. With the attributes from the dataset the work can shows different visualizations based on the need. This project helps to visualize the Crime density over an area. The crime density maps help the crime analysts to analyze the crime patterns. Understanding patterns of criminal activities is important for law enforcements and intelligence agencies to investigate and prevent crimes. The interactive and visual features can be helpful in discovering and analyzing the crime Networks. Crime map plots can help the investigators to explore relationships between criminals in the social network. The following figures are generated with the help of the data in datasets and by various R packages. Each figure in the following can be used for various analysis.

The above figure shows the places where recently the crimes had happened. This can make the user know which areas are dangerous and it can help them to avoid those areas. The picture can help the law enforcements to improve the security in the areas. we can see from the picture that locations crimes happened are very near to each other. From this we can analyze that if a location feasible to a crime attack, then the locations near to those are also feasible for the attack.

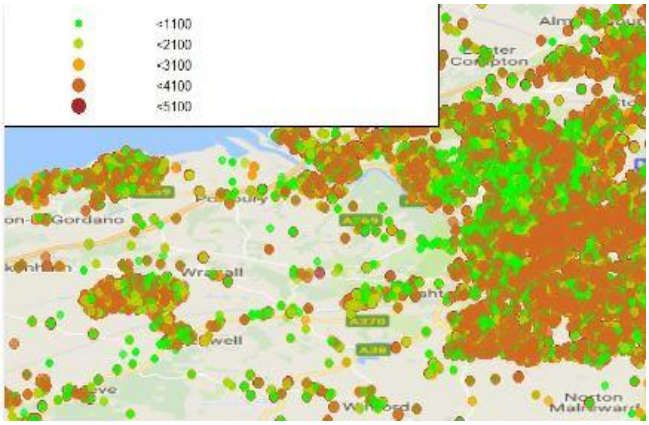
The picture visualizes the area where the crime has happened. This help the law enforcements to analyze the security measures of an area. This is an interactive image which takes help of Google Maps to navigate around the crime location and it can help the analyst to analyze the security of an area, also what locations can be the target for next attack.



This picture visualizes the no of crimes that happened in every month. This can help to visualize the frequency of crimes based on different category. This can help the public to take safety measures and helps the crime analysts to check which type of crimes have increased.



The picture visualizes the crimes that had happened based on category over different areas. This helps the law enforcements to analyze what type of crimes are frequently happening in an area. The law enforcements can improve security measures based on the type of crimes.



The picture visualizes the crime hotspots i.e. which areas the more no of crimes had happened. This map can be helpful for exploring the reasons why more crimes are happening in an area.

III. Algorithms:

Several algorithms can be used for prediction. The work provides the use of the following algorithms.

A) K-Nearest Neighbor:

K-NN is a method used for classification. In K-NN classification, the output is a class membership. An object is classified by a majority vote of its neighbor, with the object being assumed to the class most common among its k-nearest neighbors. This algorithm can be applied to the crime dataset. We can think of crime happened in a house suppose a theft, then the house next to it is also vulnerable for the theft as the criminal estimates the security is less and can try for the theft at same locations again. Hence, we can say the close the areas to the crime scene the more probable the crime happens in those areas again.so location is one factor to be considered. Date can also be considered as a factor. We need to compute the distance factor for classification, for this factor we compute the distances between the area that we want to test and the areas in the training set. For this we can consider the latitude and longitude as the

coordinates and compute the distance factor as

$$\boldsymbol{d_i} = \sqrt{(\boldsymbol{x_i} - \boldsymbol{x})^2 + (\boldsymbol{y_i} - \boldsymbol{y})^2}$$

If date is also considered as a factor, then we need to compute the no of days (Z_i), and calculate the distance factor as

$$\boldsymbol{d_i} = \sqrt{(\boldsymbol{x_i} - \boldsymbol{x})^2 + (\boldsymbol{y_i} - \boldsymbol{y})^2 + (Z_i)^2}$$

The problem with the K-NN is the computation. Every time it computes the Euclidean distance which involves squaring and square root. Of course, computing distance with every training set can be parallelized using OpenMp parallel processing techniques. To avoid the squaring and square root we compute the Manhattan distance i.e.

$$d_i = |x_i - x| + |y_i - 1| + |z_i|$$

This can also be computed parallelly. After computing the distances, we find the nearest ones by using any effective sorting techniques and we assign the crime attribute that has maximum voting in the k neighbors.

sno	DATE	crime_type	latitude	longitude	location
1	2015-01-01	Anti-social behaviour	51.41087300	-2.50993000	On or near Ludlow Close
2	2015-01-01	Anti-social behaviour	51.41294100	-2.51215300	On or near Heathfield Close
3	2015-01-01	Burglary	51.40871700	-2.51581600	On or near Caroline Close
4	2015-01-01	Criminal damage and arson	51.40996600	-2.51176100	On or near Caernarvon Close
5	2015-01-01	Other theft	51.41613700	-2.50912600	On or near St Francis Road
6	2015-01-01	Vehicle crime	51.41935700	-2.51507200	On or near Stockwood Hill
7	2015-01-01	Anti-social behaviour	51.42023200	-2.49776700	On or near Chandos Road
8	2015-01-01	Anti-social behaviour	51.42318100	-2.51042700	On or near Durley Lane
9	2015-01-01	Anti-social behaviour	51.42318100	-2.51042700	On or near Durley Lane
10	2015-01-01	Anti-social behaviour	51.41669200	-2.50142500	On or near Parking Area

The picture shows the first 10 elements of training dataset in the database

```
lat<-50
lon<- (-0.4)
test<-data.frame(Latitude=lat,Longitude=lon)
```

The picture shows the data that is to be tested i.e. finding what crime can happen at a given location.

```
[1] Anti-social behaviour
attr(,"prob")
[1] 0.7142857
```

The picture shows the output of k-NN showing the crime that can happen at an area and the probability of happening.

B) Naïve Bayes:

It is based on Bayes theorem which describes the probability of an event based on the prior knowledge of conditions that might be related to the event. Mathematically it can be stated as

$$P\left(h/x\right) = \frac{P\left(x/h\right).P(h)}{P(x)}$$

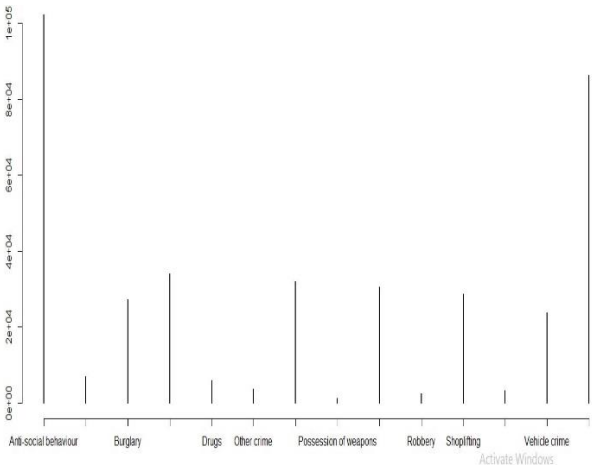
The Naïve Bayes classifier classifies a new instance X by assigning the most probable target value i.e. the maximum likelihood. i.e.

$$Y = \max_{d_i \in d} \left(p(d_i) \cdot \prod_{k=1}^{n} p\left(\frac{x_k}{d_i}\right)\right.$$

(since Naïve Bayes assumes the independency of the attributes.)

With the data available in the datasets Naïve Bayes classifier can be applied on the Latitude, Longitude (or location), Date attributes to classify the crime type that can occur.

```
Anti-social behaviour                Bicycle theft
            0.3333333                    0.0000000
                Drugs                Other crime
            0.0000000                    0.0000000
        Public order                      Robbery
            0.0000000                    0.0000000
       Vehicle crime Violence and sexual offences
            0.1666667                    0.0000000
```



The pictures show the computed probabilities and their graphical representation.

IV. Conclusion:

The project helps to make a framework for visualizing the crime networks and analyzing them by various machine learning algorithms using the google Maps and various R packages. The project helps the crime analysts to analyze these crime networks by means of various interactive visualizations. The interactive and visual feature applications will be helpful in reporting and discovering the crime patterns. A number of classification models can be considered and compared in the Analysis. It is evident that law enforcing agencies can take a great advantage by using machine learning algorithms to fight against the crimes and saving humanity. For better results we need to update data as early as possible by using current trends such as web and Apps.

References:

[1] Ihaka, Ross (1998). R: Past and Future History. Interface98 (Technical report). Statistics Department, The University of Auckland, Auckland, New Zealand.

[2] Bao Wang, Duo Zhang, Duanhao Zhang, P. Jeffrey Brantingham, and Andrea L. Bertozzi,' Deep Learning for Real Time Crime Forecasting'. July 2017.

[3] Markus Loecher, RgoogleMaps, https://cran.r-project.org/web/packages/RgoogleMaps/RgoogleMaps.pdf , 2016

[4] Swetha. G,' Crime Data Investigation and visualization using R', IJETIE, Volume I, Issue 5, May 2015

[5] U.K. Crime data*, https://data.police.uk/data/*

[6] Markus Gesmann, Diego de Castillo, R Interface to Google Charts, https://cran.r-project.org/web/packages/googleVis/googleVis.pdf , 2017

[7] David Kahle, Hadley Wickham, 'Spatial Visualization with ggplot2', https://cran.r-project.org/web/packages/ggmap/ggmap.pdf ,2016

[8] Run Yu, Manlin Song, Enhao Cui, 'San Francisco Crime Analysis and Classification',2017.