Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 132 (2018) 696–705

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Computational Intelligence and Data Science (ICCIDS 2018)

Crime Prediction & Monitoring Framework Based on Spatial Analysis

Hitesh Kumar Reddy ToppiReddy[a,*], Bhavna Saini[a], Ginika Mahajan[a]

[a] *School of Computing and Information Technology, Manipal University Jaipur, Jaipur-303007, India.*

Abstract

Crimes are treacherous and common social problem faced worldwide. Crimes affect the quality of life, economic growth, and reputation of a nation. There has been an enormous increase in crime rate in the last few years. In order to reduce the crime rate, the law enforcements need to take the preventive measures. With the aim of securing the society from crimes, there is a need for advanced systems and new approaches for improving the crime analytics for protecting their communities. Accurate real-time crime predictions help to reduce the crime rate but remains challenging problem for the scientific community as crime occurrences depend on many complex factors. In this work, various visualizing techniques and machine learning algorithms are adopted for predicting the crime distribution over an area. In the first step, the raw datasets were processed and visualized based on the need. Afterwards, machine learning algorithms were used to extract the knowledge out of these large datasets and discover the hidden relationships among the data which is further used to report and discover the crime patterns that is valuable for crime analysts to analyse these crime networks by the means of various interactive visualizations for crime prediction and hence is supportive in prevention of crimes.

Peer-review under responsibility of the scientific committee of the International Conference on Computational Intelligence and Data Science (ICCIDS 2018).

* Corresponding author. Tel.: +91-7726054801.
E-mail address: hitesh95reddy@gmail.com

Peer-review under responsibility of the scientific committee of the International Conference on Computational Intelligence and Data Science (ICCIDS 2018).
10.1016/j.procs.2018.05.075

Keywords: Crime Analysis; Crime prediction; Data Visualization; Crime Maps.

1. Introduction

Crimes are common social problems that affect the quality of life, economic growth and reputation of a country. Crimes are one of the major factors that affect various important decisions of an individual's life like moving to a new place, roaming at right time, avoiding risky areas, etc. Crimes affect and defame the image of a community. Crimes also affect the economy of a nation by placing the financial burden on government due to the need for additional police forces, courts etc. As crimes are increasing drastically, we are at the alarming stage to reduce them at even faster rate. The latest figures show a 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 [9]. 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 [9]. These figures can be reduced if we are able to analyze and predict the crime occurrence, the locations and take preventive measures in advance. The 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. Proper analysis of previous crime data helps in predicting the crimes and thus supports in reducing the crime rate. The analysis process includes the study of crime reports and identifying the emerging patterns, series, and trends as quickly as possible. This analysis helps in preparing statistics, queries, and maps on demand. It also helps to see if a crime fits in a certain known pattern or a new pattern is necessary.

Crimes can be predicted as the criminals are active and operate in their comfort zones. Once successful they try to replicate the crime under similar circumstances [16]. The occurrence of crime depends on several factors such as intelligence of a criminal, security of a location, etc. Criminals generally find similar location and time for attempting next crime. Although it may not be true for all the cases, but the possibility of repetitions is high, as per studies, and this makes the crimes predictable.

This paper proposes a web mapping & visualization-based crime prediction tool which is built in R [1] using its various libraries such as RgoogleMaps [3], googleVis[5], etc. The proposed framework uses different visualization techniques to show the trend of crimes and various ways that can predict the crimes using machine learning algorithms. The work has followed the steps that used in Data Analysis [15], in which the important phases are the Data collection, Data pre-processing, Data Visualization and Model building which are discussed more in detail in the following sections. In brief, in data collection phase the data is obtained from the official site of U.K. police department [7]. The data pre-processing phase consists of cleaning and transformation of data. The visualization phase generates various reports and maps for diagnosis and analysis process and finally, in model building phase various machine learning algorithms are used for classification of crime that can happen in a particular location.

2. Related work:

Analysis and prediction of crime is an important activity that can be optimized using various techniques and processes. Lot of research work is done by various researchers in this domain. The existing work is limited to use the datasets to identify locations of crime. But none of them considered that the type of crime, date of crime as the factor. Yu, R et. al provides the static maps with no interactive features [8]. To overcome these limitations, the proposed framework provides the visualization techniques that consider the type of crime to identify the crime hotspots (shown in the fig.3 and fig.4) and helps to check these locations with the interaction features using Google maps (shown in fig .2).

Few papers focused on usage of decision trees for crime prediction [4] [13][14]. Ahishakiye et. al and Iqbal et. al, used the attributes population of country, Median Household income, percentage of people who are unemployed with age greater than 16, type of crime, etc. which only predicts whether in an area there will be high, medium or low percentage of violent crimes that can happen in future. The methods proposed by them didn't predict the type of crime that can happen [4] [14]. Nasridinov et. al also proposed a method for classifying the crime rate as high, medium or low. None of them has classified the type of crime that can happen and its probability of happening.

Further, all used the decision trees which provide the information of which parameters of the dataset are important for the study. Further, it predicts the crime of a location if the information of location is available in the dataset. For example, if a crime happened at a location with latitude=-2.44 and longitude=50.35, the previous related provides the information of future happening of crime at this location only but cannot predict at a location with latitude=-2.3 and longitude=50.4. By using the nearest neighbour approach (k-nn), this paper is able to overcome the mentioned problem.

3. METHODOLOGY:

For optimum analysis and prediction of crime incidents, a Crime Prediction & Monitoring Framework Based on Spatial Analysis is introduced. In this framework, various visualization techniques are used to analyze the data in a better way. This framework is implemented in a GUI based tool using R programming and its various libraries. The methodology and various phases are described as follows.

3.1 DATA COLLECTION & PREPROCESSING:

The dataset used for the work is reliable, real and authentic as data is acquired from the official site of the U.K. Police department [7]. 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. In this phase, the history of crimes from the year 2015-17 was considered as the training dataset. In the pre-processing phase, removal of the inconsistent data (such as missing values, redundant information, etc.) and transformation of the data is done that is required for the predicting the crime in the following modules.

3.2 DATA VISUALIZATION:

Data visualization is an art and science. It is a form of visual communication. It involves creation and study of the 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. The work provides the generation of crime density maps which helps the crime analysts to analyze the crime patterns. Understanding patterns of criminal activities are important for law enforcement and intelligence agencies to investigate and prevent crimes.

As crimes occur in an area, analyzing them through location and maps helps a lot of understanding. This paper provides a novel tool for visualizing the previous crime data on maps and predict the future crimes that can happen. 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. As compared to textual data, visualization of information provides a better understanding, we have developed a tool to explore the dataset that provides various visualization modules. The various modules of the tool are developed in R [1] by using various R libraries mainly RgoogleMaps [3], googleVis [5], ggplot2[6] and ggmap [6]. In following sections, the various modules are described.

3.2.1 Module 1: Visualization of Crime Data Using Google Maps

This module extracts the recent crime data from the dataset and based on longitude-latitude it tags the specific location of the city. This tagging also displays the crime location name, the type of crime that happened. This information is useful for an individual in knowing dangerous and risky areas and it thus can help them to avoid such areas. The picture can help the law enforcement to improve the security in the areas. Fig.1 shows that locations, where crimes occurred, are very near to each other. From this, we can analyze that if a location feasible to a criminal attack, then the nearby locations are also feasible for the crime to occur. This module also provides the facility to enquire about a specific location to show what type of crime is feasible to happen in that location.

Fig.1.Visulaization of Crime Data Using Google Maps.

3.2.2 Module 2: Visualization of Exact Location of Crime with 3D View

This module visualizes the area where the crime has happened exactly. This helps the law enforcement to analyze the security measures of an area. The module provides the 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 also helps the police for the clear understanding of the cause of crime and helps them to investigate the location by not visiting the location again and again. By just clicking the tag in Fig.1, it provides the realistic 3D interactive image of the location and helps in navigation around the location as shown in Fig.2.

Fig.2. Visualization of Exact Location of Crime with 3D View.

3.2.3 Module 3: Visualization based on type of Crime

The type of crime is also an important factor as safety measures are majorly taken based on the type of crime. This module helps visualize the crimes that had happened based on category over different areas as shown in Fig.3. This helps the law enforcement to analyze what type of crimes are frequently happening in an area and helps them to improve security measures based on the type of crimes.

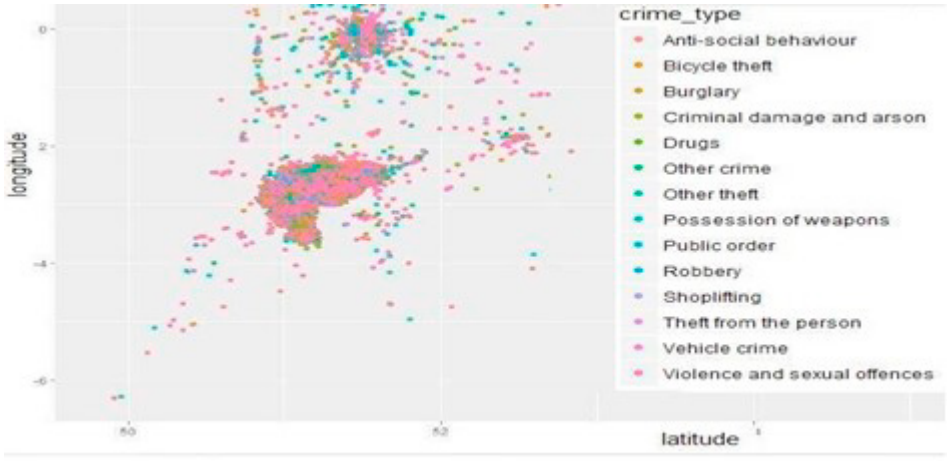

Fig.3. Visualization based on type of Crime.

3.2.4 Module 4: Visualization of Crime Hotspots

The number of crimes happened in an area makes sense of how dangerous the area is. This module helps to visualize the crime hotspots as shown in the Fig.4. The areas on the map that have high crime density are called the crime hotspots [10]. Developing maps that contain hotspots are becoming a critical and influential tool for policing. These are used by the researchers and analysts to examine the occurrence of hotspots in certain areas and why they happen and help them to build the theories. This also allows researchers to explain why crime occurs in certain places and why crime does not in other places. Crime analysts can use these to make better decisions, target resources, formulate strategies and help the law agencies.

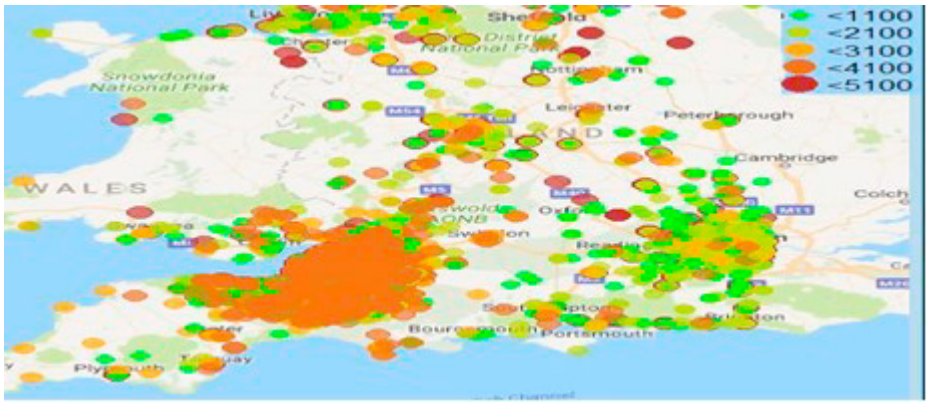

Fig.4. Visualization of Crime Hotspots

3.2.5 Module 5: Crime Frequency Report

This module helps to generate the crime report based on the number of crimes that happened in every month and on different categories of crimes as shown in Fig.5. This can help the public to take safety measures and helps the crime analysts to check which type of crimes are increased or decreased.

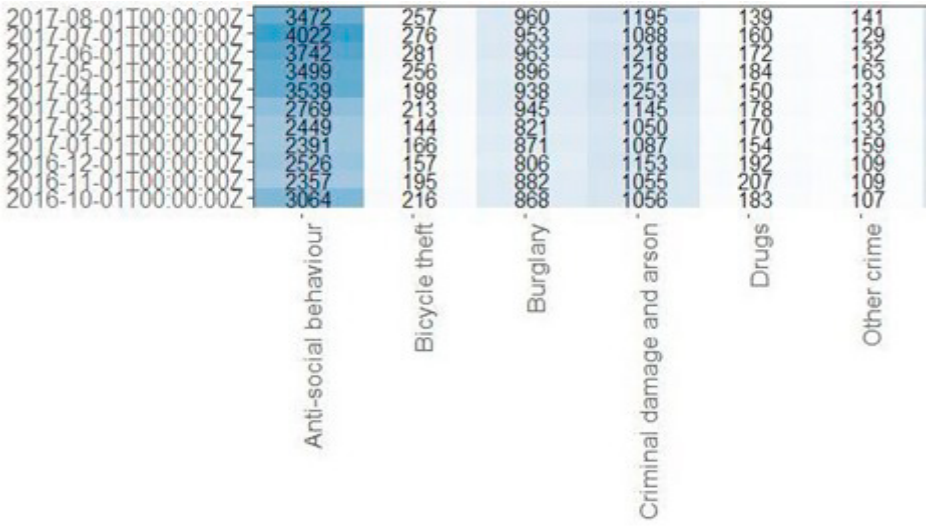

Fig.5. Crime Frequency Report

3.2.6 Module 6: Interactive Crime Frequency Report Using Graph and Bar chart:

This module helps to generate a video representation of the trend of the type of crimes in every month that is extracted from the datasets. This helps to visualize that which type of crime has increased or decreased compared to previous months. Fig.6 represents the bar-chart, when the video is played it shows how the frequency of each crime is changing in every month (the bars move up and down depending on the crimes have increased or decreased) and Fig.7 represents the graphical representation in the bar chart. This module helps the analysts to understand the trend of every crime that has happened in an area. Fig.6 shows that there are 3523 Anti-social behaviour crimes and around 1000 burglary crimes are reported on 13-April-2017. Fig.7. represents the graphical representation of the bar chart. This module helps the analysts to understand the trend of every crime that has happened in an area. It shows that there are 3539 Anti-social behaviour crimes (the blue line) are reported on 1-January-2017. It also displays that Anti-social behaviour is the most frequent crime in every month. The Burglary cases (the red line) have increased majorly in every month, and remaining crimes are tending to be constant.

Fig.6. Bar chart between crime frequency and time (month).

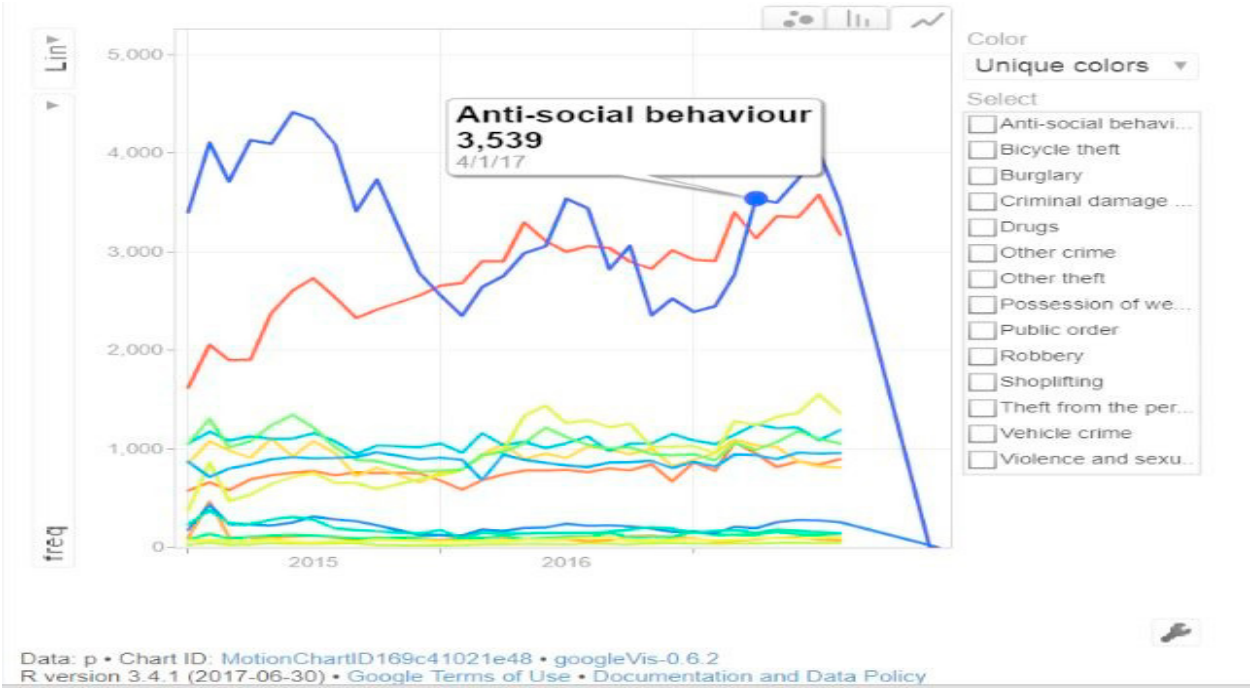

Fig.7. Month wise crime trend report

3.3 Crime Prediction:

Data mining involves exploring the datasets and extracting the fruitful information to transform into an understandable form for the further use. The data mining techniques were applied to the crime data for the crime prediction based on theories in Criminology. The criminology mainly focuses on the Rational Choice Theory [12] and Routine Activity Theory [11]. The Rational choice theory focuses on the understanding of crimes from offender's perspective which is directly concerned with thinking process of the offender and how they evaluate their opportunities.

The Routine Activity Theory states that for a crime to occur, a likely offender finds a suitable target with capable guardians absent and states that crimes are unaffected by social causes such as poverty, inequality, and unemployment. The criminals repeat their activities by choosing the targets which are under similar conditions. Based on this information the work provides the use of the following algorithms.

3.3.1 K-Nearest Neighbour:

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 neighbour, with the object being assumed to the class most common among its k-nearest neighbours. This algorithm can be applied to the crime dataset. Suppose a theft has happened in a house, 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, the areas nearby the previous crime location are more probable for crime occurrence. Therefore, the location is one factor to be considered. The date can also be considered as a factor. The distance factor for classification, hence the distances between the testing areas and training areas are computed. For this, the latitude and longitude as the coordinates and compute the distance factor as

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

If the date is also considered as a factor, the no of days (Z_i) needs to be computed and then calculate the distance factor as

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

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, the Manhattan distance was computed i.e.

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

This can also be computed parallelly. After computing the distances, the nearest ones were identified by using effective sorting techniques and are assigned the type of crime attribute that has maximum voting in the k-neighbours.

```
> test
   latitude longitude
1  53.74933 -2.000177
2  52.85565 -2.799866
3  53.98675 -2.548269
4  53.97359 -2.828916
5  53.72188 -2.161456
6  53.84521 -2.546358
7  52.12237 -2.252581
8  52.25856 -2.534608
```

Fig.8. shows the data that is to be tested i.e. finding what crime can happen at a given location.

```
> tested
  latitude longitude             prediction probability                          textAddress
1 53.74933 -2.000177                  Drugs   0.6666667          Walker Ln, Hebden Bridge HX7, UK
2 52.85565 -2.799866          Shoplifting   0.4000000            B4397, Shrewsbury SY4 5ST, UK
3 53.98675 -2.548269          Shoplifting   0.4000000     Whitendale Road, Clitheroe BB7 3BL, UK
4 53.97359 -2.828916          Shoplifting   0.4000000          Hillam Ln, Lancaster LA2 0DX, UK
5 53.72188 -2.161456          Shoplifting   0.6666667          Dark Red, Todmorden OL14 7ER, UK
6 53.84521 -2.546358          Shoplifting   0.4000000                                      NA
7 52.12237 -2.252581 Criminal damage and arson 0.6666667 115 B4424, Callow End, Worcester WR2 4TH, UK
8 52.25856 -2.534608          Shoplifting   0.4000000       Bromyard Rd, Tenbury Wells WR15, UK
```

Figure.9. shows the output of k-NN showing the crime that can happen in an area and the probability of happening.

3.3.2 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.

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

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

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

Anti-social behaviour	Bicycle theft	Burglary	Criminal damage and arson
0.37500000	0.00000000	0.10714286	0.16071429
Drugs	Other crime	Other theft	Possession of weapons
0.03571429	0.00000000	0.07142857	0.00000000
Public order	Robbery	Shoplifting	Theft from the person
0.03571429	0.00000000	0.08928571	0.00000000
Vehicle crime	Violence and sexual offences		
0.01785714	0.10714286		

Fig.10. computed probabilities

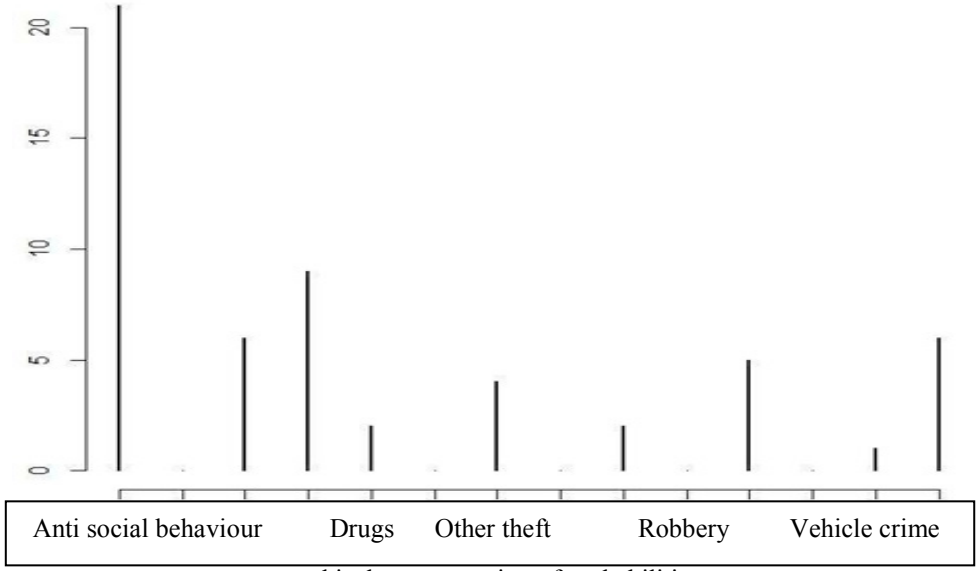

Fig.11. graphical representation of probabilities

Figures.10 and 11 show the computed probabilities and their graphical representation. It shows that there is 37.5% chance of reporting the Anti-social behaviour case and 10.7% chance of Burglary.

4. Conclusion:

The tool we have developed provides 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. Many classification models can be considered and compared in the Analysis. It is evident that law enforcing agencies can take a great advantage of 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.

5. Future Scope:

This paper presents the visualization techniques and classification algorithms that can be used for predicting the crimes and helps the law agencies. In future, there is a plan for applying other classification algorithms on the crime data and improving the accuracy in prediction. On other direction, we will be trying to build an Android App for the live capture of the realistic data and updating the results by using this new data frequently, that will be helpful in better prediction and providing the general information to the public for the awareness of trends in the crime.

References:

[1] Ihaka, R. (1998). R: Past and future history. *Computing Science and Statistics*, *392396*.

[2] Wang, B., Zhang, D., Zhang, D., Brantingham, P. J., & Bertozzi, A. L. (2017). Deep Learning for Real Time Crime Forecasting. *arXiv preprint arXiv:1707.03340*.

[3] Loecher, M. (2014). RgoogleMaps: overlays on Google map tiles in R. See http://cran. r-project.org/web/packages/RgoogleMaps/index. html.

[4] Ahishakiye, E., Taremwa, D., Omulo, E. O., Nairobi-Kenya, G. P. O., & Niyonzima, I. (2017). Crime Prediction Using Decision Tree (J48) Classification Algorithm. *analysis*, *6*(03).

[5] Gesmann, M., & de Castillo, D. (2011). Using the Google visualisation API with R. *The R Journal*, *3*(2), 40-44.

[6] Kahle, D., & Wickham, H. (2013). ggmap: Spatial Visualization with ggplot2. *R Journal*, *5*(1).

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

[8] Yu, R., Song, M., & Cui, E. San Francisco Crime Analysis and Classification.

[9] https://www.theguardian.com/uk-news/2017/oct/19/rising-at-increasing-rate-in-england-and wales-police-figures-show.

[10] Crime Petrol https://en.wikipedia.org/wiki/Crime_hotspots

[11] Routine Activity Theory https://en.wikipedia.org/wiki/Routine_activity_theory

[12] Routine Choice Theory https://en.wikipedia.org/wiki/Rational_choice_theory_(criminology)

[13] Nasridinov, A., Ihm, S. Y., & Park, Y. H. (2013). A decision tree-based classification model for crime prediction. In *Information Technology Convergence* (pp. 531-538). Springer, Dordrecht.

[14] Iqbal, R., Murad, M. A. A., Mustapha, A., Panahy, P. H. S., & Khanahmadliravi, N. (2013). An experimental study of classification algorithms for crime prediction. *Indian Journal of Science and Technology*, *6*(3), 4219-4225.

[15] https://en.wikipedia.org/wiki/Data_analysis

[16] https://www.slideshare.net/socialmediadna/predictive-policing-the-role-of-crime-forecasting-in-law-enforcement-operations