Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 04, April-2020

Automatic Clustering Crime Region Prediction Model using Statistical Method in Data Mining

Shivani B. Mehta
Department of Computer Engineering
V.V.P. Engineering College, Rajkot
Gujarat, India

Rushabh D. Doshi
Department of Computer Engineering,
V.V.P. Engineering College, Rajkot
Gujarat, India

***Abstract*:- The most common social problem faced today all over the world are crimes. Quality of life, economic growth and reputation of nation are greatly affected by crimes. There has been increase in the crime rate since last few years and in order to reduce this crime rate government needs to take the preventive measures. In order to secure the society from such crimes there is a need to adopt new advanced system and approaches to improve crime analytics for protecting the community. Accurate real time crime predictions help up to a greater extent to reduce crime rate but it is a challenging problem as occurrences of crime depends on many factors. Various visualizing techniques and machine learning algorithms are followed for predicting the crime distribution in an area. The raw data sets were processed and visualized based on needs In the first step and later machine learning algorithms were used to extract knowledge out of the large datasets and find the hidden relationships among data which is further used to for reporting and discovering the crime patterns which is very important source of information for crime analyst to analyze these crime networks by means of various interactive visualizations for crime prediction and thus is very supportive in preventing crime.**

Keywords:- Crime Analysis, Crime prediction, Data Visualization, Crime Maps.

1. INTRODUCTION:

Crimes are one of the common problems faced by the country that affects the quality of life, economic growth and reputation of country but it also affects the various important decisions of an individual's life like moving to a new place for livelihood, roaming at right time avoid going to risky areas etc. crime broadly affects and defames the image of a community. Crimes also affect the economy of a nation as government has to arrange for additional police forces courts etc. which in case increases financial burden on government due to drastic increase in the crime rate in our country. We are at alarming stage to take necessary preventive measures to reduce them at faster rate as per the latest reports. There is a 13% increase in all police records offences across England and wales and even a greater increase in the violent affiances like knife crime and sexual offences and violence against person. The crime figures even show a 8% raise in the murder rate. Increase of 46 victims with 629 homicides recorded in 12 month to June. This figure are excluding the 35 people who got killed in terror attacks in London and Manchester however this figures can be reduced to a greater extent if we are able to analyze and predicate the crime occurrence the locations in advance and take prevention measures. The crime rates can be reduced by real time forecasting and mass surveillance which are helpful to save lives. Proper analysis of the previous year's crime data helps in reducing the crime rate consideration extent the steps included in analysis process are studying the crime reports and identifying the emerging patterns.[4]

Crimes can also be predicted easily as criminals are active and usually operate in their comfort zones once they are successful in committing the crime they replicate the crime under similar conditions. The occurrence of crime depends on different factors like intelligence of crime location security etc. criminals usually follow similar location and time while attempting the next crime. However it may not be true in all the cases. But the possibility of repeating the crime is high and thus predicating the crime becomes easy.

This paper proposes a web mapping and visualization based crime prediction tool which is built in R[1] using its various libraries such as Google maps[3], googleris[5] etc. The proposed framework uses different visualization techniques to show the trend of crimes and various ways that can predicate the crimes using machine learning algorithms.

The important phases of data analysis are data collection data pre processing data visualization and model building. In data collection phase data is obtained from the official site of u.k. police department. In the second step i.e. data pre processing it consists of cleaning and transformation of data. The visualization phase generates reports and maps for diagnoses and analysis process and finally in the model building phase classification of crime that can happen in a particular location is done using various machine learning algorithms.[8]

2. LITERATURE REVIEW:

2.1Crime Prediction & Monitoring Framework Based on Spatial Analysis:

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

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 04, April-2020

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.

2.2 An Intelligent Document Clustering Approach to Detect Crime Patterns:

In this paper Document Clustering is considered as one of the most commonly used methods in detecting topics/events or types of crime documents [6], and a method in which document clustering has three main processes [1-7].

The first process is preprocessing of documents to remove unimportant words and symbols from the document of crime.[9]

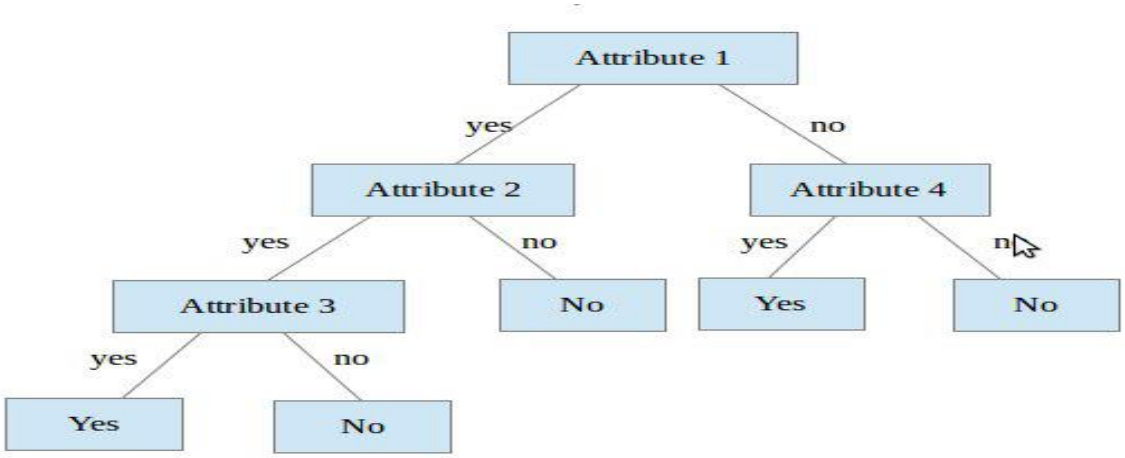
The second process is representation of documents crimes to extract the most important information from the document of crime and show the similarity among these documents.

The last process of document clustering consists in applying the clustering algorithm to the groups of documents of topics/events or types crime based on the similarities among the documents.[12]

2.3 Crime Analysis and Prediction Using Data Mining:

In this paper, we investigate the problem of clustering is solved by naive Bayes algorithm.in this paper we first collect the data than classify according to their data. The advantage of using Naive Bayes Classifier is that it is simple, and converges quicker than logistic regression. Compared to other algorithms like SVM (Support Vector Machine) which takes lot of memory the easiness for implementation and high performance makes it different from other algorithms. Also in case of SVM as size of training set increases the speed of execution decreases.[8] Using Naive Bayes algorithm we create a model by training crime data related to vandalism, murder, robbery ,burglary, sex abuse, gang rape, arson, armed robbery, highway robbery, snatching etc. By training means we have to teach them on particular inputs such that we can test them for unknown inputs.[10]

Figure 2.3.1 Attribute decision tree



2.4 Survey on Crime Analysis and prediction using data mining:

Sharma [1] proposed a concept which depicts zero crime in the society. For detecting the suspicious criminal activities, he has concentrated on the importance of data mining technology and designed a proactive application for that purpose. In his paper, he proposed a tool which applies an enhanced Decision Tree Algorithm to detect the suspicious e-mails about the criminal activities. An improved ID3 Algorithm with an enhanced feature selection method and attribute-importance factor is applied to produce a better and faster Decision Tree based on the information entropy which is explicitly derived from a series of training data sets from several classes.[10] He proposed a new algorithm which is a combination of Advanced ID3 classification algorithm and enhanced feature selection method for the better efficiency of the algorithm.

2.5 Using Machine Learning Algorithms to Analyze Crime Data:

we implemented the Linear Regression, Additive Regression, and Decision Stump algorithms using the same finite set of features, on the communities and crime un normalized dataset to conduct a comparative study between the violent crime patterns from this particular dataset and actual crime statistical data for the state of Mississippi that has been provided by neighborhoodscout.com [6]. The crime statistics used from this site is data that has been provided by the FBI and had been collected for the year 2013 [6]. Some of the statistical data that was provided by neighborhoodscout.com such as the population of Mississippi, population distribution by age, number of violent crimes committed, and the rate of those crimes per 100K people in the population are also features that have been incorporated into the test data to conduct analysis.

.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 04, April-2020

3. METHODOLOGY:

For optimum analysis and prediction of crime incidents a crime prediction and monitoring framework used on spatial analysis has been introduced. Various visualization techniques are used to analyze the data in better way. This framework is implemented in GUI based tool using R programming and various phases are described as follows:

3.1) Phase 1: Data Collection:

The dataset which is used for the work is real, reliable and authentic as it is acquired from the official site of the U.K. police department.[7] The dataset contains a total of 11 attributes out of which 5 attributes are considered important for the study. This attributes 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 datasets. In this stage removal of irrelevant data such as missing and transformation of data which is required for predicting the crime is done.

3.2) Phase 2: Data Visualization:

Data visualization is a form of visual communication. It is an out as well as science. It involves creation and study of visual representation of data. The primary aim of data visualization is to communicate data clearly and effectively to its users through statically Graphics and plots. The effective are visualization helps to reason about data and evidence. Data visualization includes generation of crime density maps which helps the crime analysis to analyze the crime patterns. It is important to understand the pattern of criminal activities for the effective enforcement of law and also for intelligent agencies. It investigates and prevents crimes.

When a crime occurs in an area analyzing them through location and maps helps. A lot of understanding data visualization provides a novel tool for visualizing the previous crime data on maps and predicate the future crimes that can happen. The interactive and visual features are helpful in discovering and analyzing the crime networks. Crime map plots are also a useful tool for the investigators to explore relationships between criminals in the social network. Studying visualized information provides a better understanding than textual data. There are various tool to explore the data set that provides various modules of the tool are developed by using various R libraries [1] R libraries mainly includes R goggle maps [3], ggplot [5] and ggmap [6].In following sections the various modules are described.

3.3) Phase 3: Visualization of crime data using Google maps:

This module extracts the recent crime data from the dataset and based on longitude and latitude it tags the specific location of city. This tagging also shows the crime location name and type of crime that happened. This information helps an individual in knowing dangerous and risky areas and also helps to avoid this area. It also helps the government in enforcing the laws and increasing security of crime zone areas.

By visualization of crime data using Google maps we can analyze if a location is feasible to a criminal attack then it's nearly 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 happen in that location.

[Fig.1. Visualization of Crime Data Using Google Maps]

3.4) Phase 4: Visualization of exact location of crime with 3D view:

This module helps to visualize the area where the crime has exactly happened. It also helps the government to enforce more laws for the additional security of an area. The module also provides intercalative image which takes helps of Google maps to navigate around the crime location and it also helps the analyst to analyze what can be the next target for crime attack. This also helps the police for clear understanding of the cause of crime and investigates the location by not visiting the location again and again.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 04, April-2020

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

3.5) Phase 5: Visualization based on type of crime:

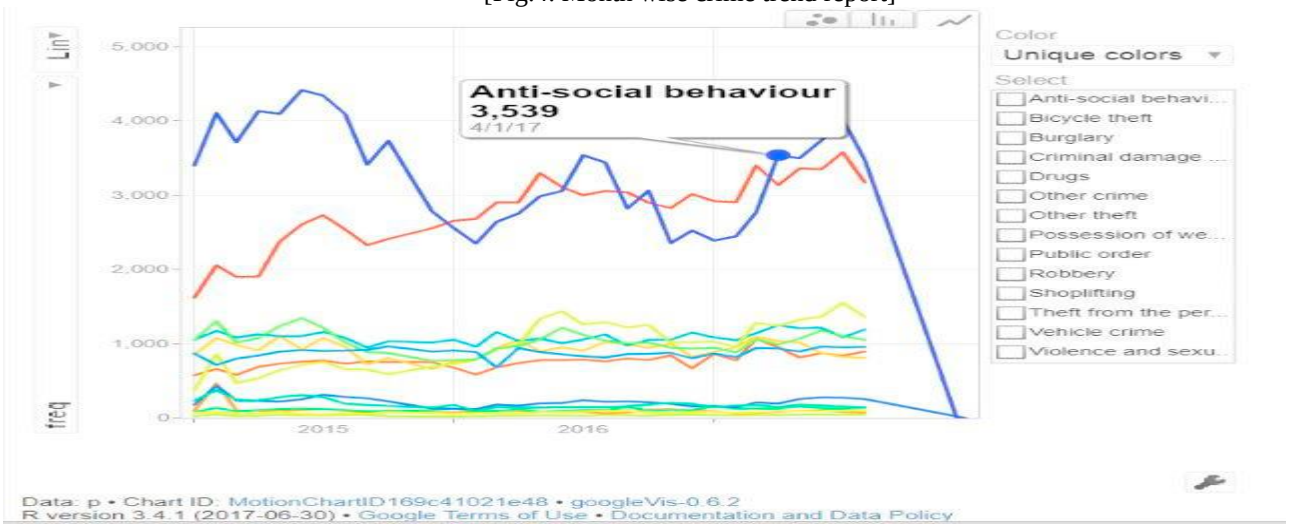
The type of crime is also an important factor to consider as safety measures to be taken are dedicated based on the type of crime. This module also helps to visualize the crime that had happened based on categories of different areas. This module helps the government to analyze the type of crime which is frequently happening in a particular are and to take necessary security measures for the improvement.

[Fig.3. Visualization of Crime Hotspots]

3.6) Phase 6: Visualization of crime hotspots:

The number of crimes happening in an area makes sense. Of how dangerous an area is. This module helps to visualize the crime hotspots developing maps that contain hotspot are becoming a critical and influential tool for policing. These maps are used by analyst and researchers to examine the occurrence of hotspot in certain areas and why does it happen and also helps to crime theories it also helps researchers to explain why crimes occur in particular areas. And why it does occur in data to make better decision, target, resources formulate strategies and help the law agencies.

[Fig.4. Month wise crime trend report]

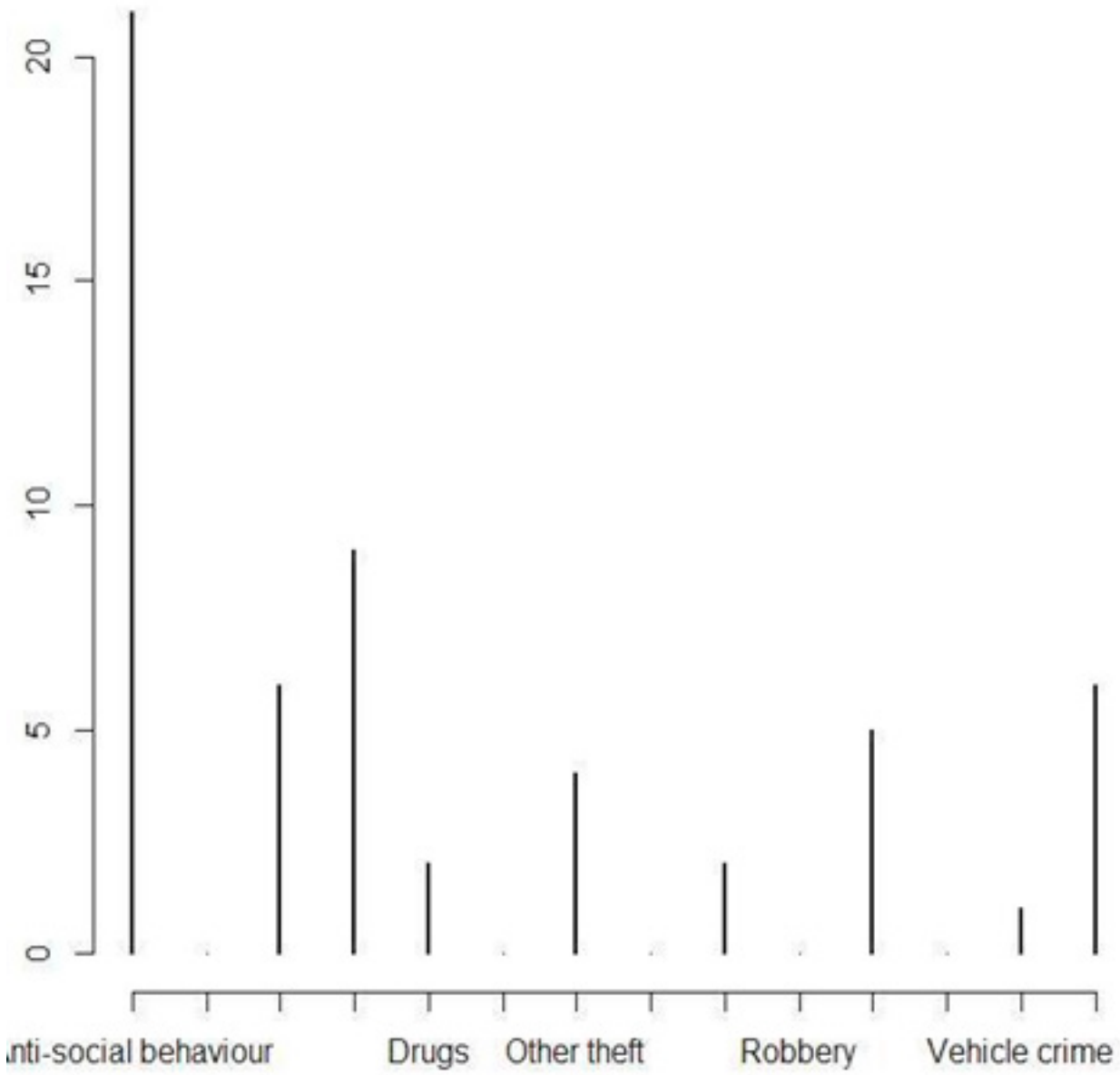


International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 04, April-2020

3.7) Phase 7: Crime frequency report

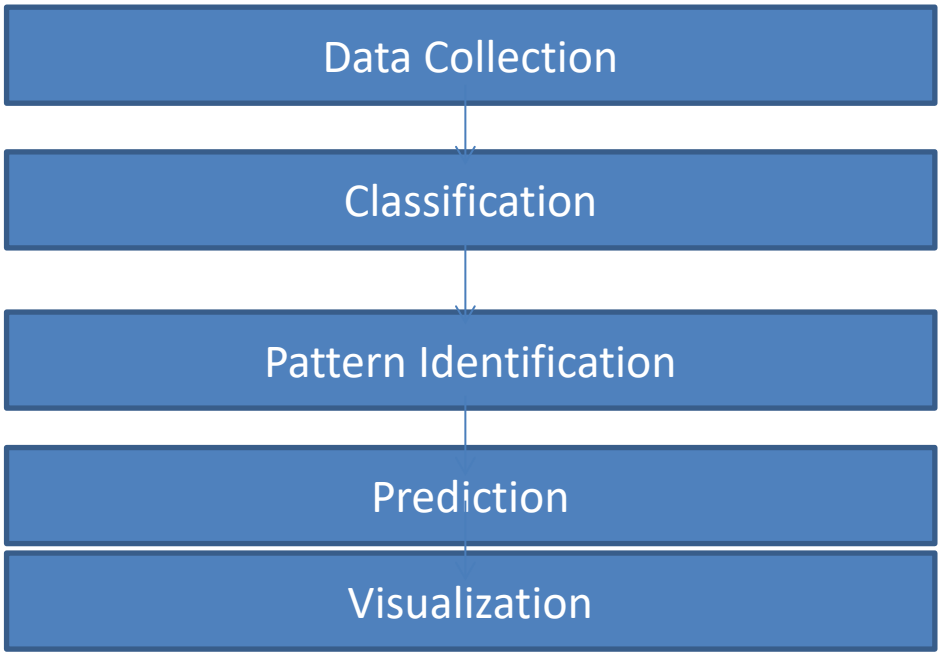
This module helps to generate the crime report based on the number of crime report based on the crime that happened in every in different categories of crime. It also helps the public to take safety measures and helps the crime analyst to check which type of crimes have increased or decreased.

[Fig.5. Graphical representation of probabilities]



International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 04, April-2020

Flow Diagram of the proposed methodology:



Data Collcetion

In data collection step we are collecting data from different web sites like news sites, blogs, social media, RSS feeds etc. The collected data is stored into database for further process. Since the collected data is unstructured data we use Mongo DB. Crime data is an unstructured data since the no of field, content, and size of the document can differ from one document to another the better option is to have a schema less database. Also the absence of joins reduces the complexity.

Classification

For classification we are using an algorithm called KNN -classifier which is a supervised learning method as well as a statistical method for classification. KNN classifier is a probabilistic classifier which when given an input gives a probability distribution of set of all classes rather than providing a single output. The algorithm classifies a news article into a crime type to which it fits the best.

Pattern Identification

Third phase is the pattern identification phase where we have to identify trends and patterns in crime. For finding crime pattern that occurs frequently we are using KNN- classifier algorithm. KNN-classifier can be used to determine association rules which highlight general trends in the database. The result of this phase is the crime pattern for a particular place. Here corresponding to each location we take.

Prediction

For prediction we are using the KNN-classifier concept. A decision tree is similar to a graph in which internal node represents test on an attribute, and each branch represents outcome of a test. The main advantage of using decision tree is that it is simple to understand and interpret. The other advantages include its robust nature and also it works well with large data sets. This feature helps the algorithms to make better decisions about variables [4].

Visualization

The crime prone areas can be graphically represented using a heat map which indicates level of activity, usually darker colors to indicate low activity and brighter colors to indicate high activity.

CONCLUSION

They Required Initial information of crime area from dataset. It also Express Crime scene and represent the location. The Static Approach is used to predicate the Crime in Areas. We will going to implement system in near future is by android application so it is search by location wise and by use of GPS we can easily find the location of crime prone areas.

REFERENCES

[1] "Crime prediction & Monitoring Framework Based on Spatial Analysis"
[2] "Crime Analysis and Prediction Using Data Mining"
[3] "Survey on Crime Analysis and Prediction Using Data Mining"
[4] "Crime prediction Using Decision Tree and Classification Algorithm"
[5] "An Intelligent Document clustering to Detect Crime Patterns"
[6] https://dataaspirant.com
[7] https://medium.com
[8] Loecher, M. (2014). RgoogleMaps: overlays on Google map tiles in R. See http://cran. r-project.org/web/packages/RgoogleMaps/index.
[9] https://en.wikipedia.org/wiki/Data_analysis.
[10] Chandra. A, Gupta. B and Gupta. C, A multivariate time series clustering approach for crime trends prediction. Proceeding of International Conference on Systems, Man and Cybernetics, SMC; 2008. p. 892-896.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 04, April-2020

[11] Can hui Wang, Min Zhang, Liyun Ru, Shaoping Ma An Automatic Online News Topic Key phrase Extraction System, IEEE conference, 2006.
[12] Somchai Chatvienchai Automatic metadata extraction classification of spreadsheet Documents based on layout similarities, IEEE conference, 2005.
[13] Kahle, D., & Wickham, H. (2013). ggmap: Spatial Visualization with ggplot2. *R Journal*, 5(1)
[14] https://www.slideshare.net/socialmediadna/predictive-policing-the- Roleof-crime-forecasting-in-law-enforcement-operations
[15] https://en.wikipedia.org/wiki/Data_analysis
[16] 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.
[17] Crime Petrol https://en.wikipedia.org/wiki/Crime_hotspots
[18] U.K. Crime data, https://data.police.uk/data/ [8] Yu, R., Song, M., & Cui, E. San Francisco Crime Analysis and Classification.
[19] Fodeh, Punch and Ning Tan. On ontology-driven document clustering using core semantic features, Journal of Knowl Inf Syst, Springer- Verlag London; 2011.
[20] Al-Shammari, patent application publication of lemmatizing, stemming and query expansion method and system, Pub.no.:US 2010/0082333 A1,
[21] Li, Kuo, Tsai. An intelligent decision-support model using FSOM and rule.
[22] Farnstrom and Lewis, Fast. Single-pass K-means algorithms, www.citeulike.org/user/zador/article/1772993 ; 2007.
[23] Bouras C, Tsogkas V. Assigning Web News to Clusters. Proceedings of Conference on Internet and Web Applications and Services; 2010. p. 1-6.
[24] Taeho J, Clustering News Groups using Inverted Index based NTSO, NDT. First International Conference on Networked Digital Technologies; 2009. p. 1-7.
[25] Dai, He, Sun. A Two-layer Text Clustering Approach for Retrospective New Event Detection. International Conference on Artificial Intelligence and Computational Intelligence, IEEE computer security; 2010. p. 364-368.
[26] Velmurugan .T. and Santhanan .T. A survey of partition based clustering algorithms in data mining: An experimental approach, Information.