

Big Data Analytics for Crime against Women in the state of Tamil Nadu

D. Usha¹, K. Ganesan²

¹Associate Professor, Dr.M.G.R. Educational and Research Institute

²Assistant Professor (Senior), VIT School of Law, VIT, Chennai Campus

Article Info

Volume 82

Page Number: 6459 - 6462

Publication Issue:

January-February 2020

Abstract

In this growing technology field, crime rates are increasing and challenge investigative people's capabilities. Crime data generation is also increasing nowadays. With the use of conventional analytical techniques, data generated today cannot be managed effectively. It would be useful to use Big Data Analytics for that huge data instead of using standard data analysis techniques. Primarily collected data will be stored and frequently occurred data will be displayed. Based on the strong attribute the frequent data is analyzed using Big Data Analytics. Finally, the analyzed data are given as input into the Rattle tool and retrieved the actual result as output. The above pattern helps the police department to identify the criminals in the short span of time.

Keywords: Big data, Crime dataset, Rattle tool, Visualization

Article History

Article Received: 18 May 2019

Revised: 14 July 2019

Accepted: 22 December 2019

Publication: 01 February 2020

1. Introduction

Big data is basically a large, enormous data that can be structured or unstructured from various sources. In processing these voluminous data, the older processing system may not be efficient. Big Data Analytics (BDA) uses advanced tools and techniques to analyze massive, voluminous sets of data and draw valid conclusions from them. The rapidly growing population of our country causes an extend in violence and, in effect, to produce large amounts of data that can be evaluated for the government to create important and necessary law and order decisions. This becomes very essential with the growing concern about the crime rate.

Crime mapping is used to evaluate, chart and model instances of crime or crime pattern in order to have an understanding to forecast the frequency of crime. Therefore, crime mapping allows both the security and police to coordinate their resources to prevent crime appropriately. In earlier periods, the mapping of crime could be performed by some groups who had special tools. All researchers and professionals now have the potential to map the crime using existing criminal statistical information and using advanced technology. Therefore, crime mapping is applied specifically to minimize crime against women by recognizing the hotspots (the areas where crime might occur at a higher rate).

Traditional data mining methods are not so capable of analyzing such an enormous amount of data either in semi-structured or unstructured format. For that purpose, data Big Data Analytics is used. Data generation is exponentially increasing, and traditional infrastructure is somewhat unable to handle such huge data. These huge data that includes unstructured or semi-structured data can be handled using Big Data Analytics. Since Hadoop's input may be semi-structured or unstructured, but the result generated from Hadoop would result in structured data. The mapper and reducer will include the algorithm for prediction and map reduction will be used to manage these data and produce the results in half the time that is taken by traditional data mining methodology.

R tool is used to distribute the data geographically. This tool is capable of generating geospatial representation of data geographically distributed data. Different packages are available with this tool which needs to be installed in order to perform the data distribution. Data analysis as well as different visualization patterns of distributed data can be obtained from this tool.

2. Problem Statement

Data mining is delicate to the reliability of input data with incomplete information (noise, redundant data) which may be unreliable. It could be complicated in its own ways to map real data to data mining characteristics.

Bigdata are commonly used to turn huge structured or unstructured raw data into critical and significant data that helps for the judiciary and legislature by creating a stable decision-making support system to maintain law and order and keep crime under control and make strategic decisions for the security and well-being of society. Population growth leading to a rapid increase in crime rate is the major issue about managing, preserving and processing enormous amounts of data produced each year over a limited period.

3. Existing System

In the existing system, data mining techniques have been used like frequent pattern mining algorithm and rule mining algorithm for analyzing crime data and identify the frequently occurred pattern and took the strongest rule combination. The existing system focused on application of Frequent Pattern Mining in Crime Pattern Bulk amount of data is stored as a dataset which contains information about suspects, victims and witnesses. Frequent Pattern Mining algorithm is used to mine the most frequent and relevant data from the dataset and compare with the current data. In Crime pattern detection, it will help the investigating department to trace out the suspects with their identification and other related features at the short span of time.

4. Proposed Work

Using R tool, Big Data Analytics is used to perform the crime pattern detection. This mainly consists of three stages – the spatial distribution of data and the creation of clusters, cluster analysis of generated clusters and crime forecasting. Geographically, the allocation of data is the first step of the distribution of available data across geographic areas. Here the data available are crime-related. This can be achieved with the geospatial products using the R tool.

The Hadoop platform is used for the second phase of cluster analysis. In this phase, clusters generated in the primary phase are used as inputs and the appropriate clustering algorithm is to be applied for the purpose of the analysis. Hadoop can execute parallel processing on various clusters, the computation will be fast compared to traditional processing capabilities, resulting in less time consuming and perform faster than the normal analysis process of data mining cluster.

5. Technologies used

Hadoop

The Apache Hadoop is a platform that enables the distributed operation of huge datasets through device clusters utilizing simple programming methods, scaling up to thousands of computers from single nodes, each providing local computing and space. Apache Hadoop's two main components consist of HDFS and YARN.

HDFS and YARN

Hadoop File System has been built using distributed file system model which stores a very huge volume of data and makes it simpler to reach. The data are located in several devices to maintain stability in the case of unexpected crashes contributing to system rescue. Yet Another Resource Negotiator (YARN) is a technology of cluster management. YARN is a Hadoop OS that explicitly handles assets like the RAM, the CPU of all nodes (devices) in the Hadoopcluster. Apps such as Hive, Map Reduce, Spark demands YARN to assign resources such as computing energy and storage to serve the application's work.

6. Crime Dataset

The data is collected from articles and many other resources which would be converted into formal data sets. According to NCRB data, crime rate rises in 2017; women remain insecure.

The National Crime Records Bureau has released data on crime in 2017, with a massive lag of more than a year. Crime rose from 3,793 per million in 2016 to 3,886 per million in 2017. Nearly 100 more crimes took place per million people in 2017 compared to the previous year (2016), though heinous crimes such as murders and rapes came down, shows a recently-released national data on crime. The National Crime Records Bureau (NCRB) on Monday released data on crime in 2017 with a massive lag of more than a year. Crimes rose from 3,793 per million in 2016 to 3,886 per million in 2017. Crime rate under state laws that pertain mostly to prohibition, narcotics, excise, electricity-related ones and gambling rose faster than crimes under the Indian Penal Code (IPC). IPC crime incidence also rose, with crimes such as kidnapping, attempt to murder on the rise, per million population.

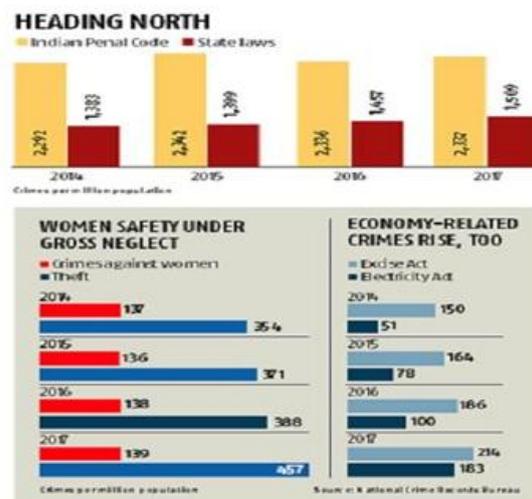


Figure 1: Summary of Crime against Women

The incidence of rapes per million populations has reduced, but women are becoming more unsafe over time, with the overall crime rate rising every year. Among

states, Karnataka, Kerala, Tamil Nadu, Punjab, Rajasthan and West Bengal showed a reduction in crime rates, while most others witnessed a rise. Theft are rising the fastest. The crime rate in Delhi rose by 8 percent in a year. Nearly three million IPC crimes and two million crimes under state laws were recorded in 2017. But this is an understatement, the report itself notes. The actual count of each crime per head may be underreported. This is because among many offences registered in a single FIR, only the most heinous crime (maximum punishment) will be considered as a counting unit,” the report notes. In a first, crimes pertaining to communal violence were not compiled by the bureau this time. Incidence of rioting reduced from 53 crimes per million people in 2014 to 46 per million people in 2017. Incidence of kidnapping and abduction, on the other hand, rose from 62 per million to 74 per million. The absolute number of reported crimes do not take the rise in population into account. An absolute rise may mean reduction in crime rate if the population has risen faster in the same period. In a disturbing trend, crime rate against women went up again, continuing the trend observed in the past. Cases of rapes reduced from 63 per million people in 2016, to 52 per million in 2017. Crimes under state laws are rising fast, courtesy a massive rise in crime rate under state electricity and excise laws. Crime rate under the electricity laws mostly pertain to electricity theft and line sabotage.

7. Implementation using R

R is a programming language used to analyze data statistically. R can manipulate data, analyze data, and visualize data. The analyzed data is used graphically in this paper by R-tool.

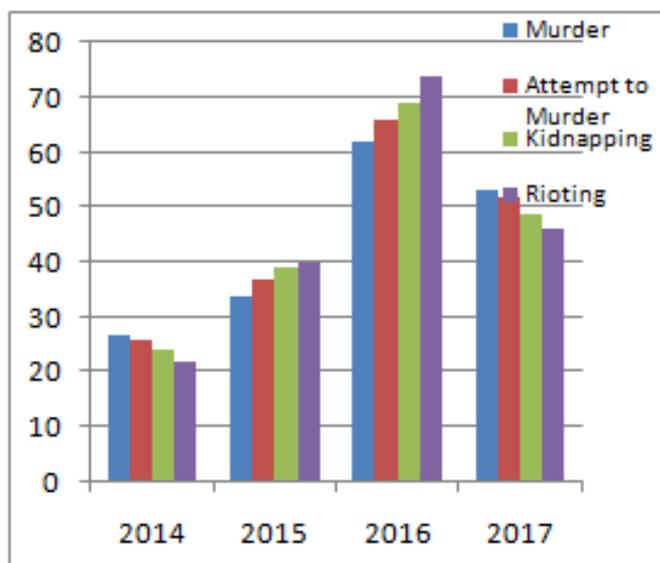


Figure 2: Comparison of Crimes

Crime Dataset

Table 1: Tamilnadu Crime Dataset

| Case No. | Date of Occurrence | A-Name | A-Age | A-POB | A-Langu age | A-Hair color | A-Skin color | A-Special features |
|----------|--------------------|------------|-------|------------|-------------|--------------|--------------|---------------------|
| CASE 1 | 25.09.04 | Raju | 38 | chennai | T | B | B | 6 fingers |
| CASE 2 | 15.08.05 | Suresh | 45 | chennai | TE | BR | F | N |
| CASE 3 | 29.01.06 | Victo | 42 | vellore | H | G | BR | left hand amputated |
| CASE 4 | 18.04.07 | Saroja | 37 | andhra | T | BD | PW | N |
| CASE 5 | 23.05.08 | Gopu | 39 | chennai | T | G | W | one eye |
| CASE 6 | 12.12.09 | Vinodhini | 42 | Salem | T | A | MBR | N |
| CASE 7 | 07.11.10 | Gopal | 46 | Karur | TE | BR | F | Bald head |
| CASE 8 | 12.05.11 | Ramu | 40 | Chennai | MA | C | MOBR | Curling hair |
| CASE 9 | 14.09.12 | Rani | 44 | Kovilpatti | TE | G | DBR | long nails |
| CASE 10 | 10.04.13 | Shanmug am | 39 | Madurai | T | R | F | N |

8. Experimental Study and Analysis

```
> dataVar <- read.csv(file.choose())
> table(dataVar$A.Name)

Geetha  Gopal  Gopi  Gopu Manohari  Papu  Raju  Ramu  Rani
      2      3      2      1      1      1      4      1      1

SarojaShanmugam  Srija  Sudha  Suresh  Victo  VijayaVinodhini
      2      1      2      1      4      1      2      1

>
> dataVar <- read.csv(file.choose())
> table(dataVar$A.Name)

Geetha  Gopal  Gopu  Raju  Ramu  Rani  SarojaShanmugam  Srija  Suresh  Victo  Vinodhini
      2      1      1      4      1      1      2      1      1      3      1      1

>
```

Figure 3: Generation of frequent data

The crime dataset in excel format is converted to csv file and given as input in the above output. The same was executed using R tool and the output has been retrieved. From the list of suspects, the suspect named Raju has committed the offense 4 times that is higher than number of times committed by others. So, from the above output the suspect who did the crime maximum number of times is the near suspect and through which the investigation team can able to find out the criminals easily.

9. Conclusion

Big Data analysis plays a significant role in the identification of crimes and supports the investigation team. Through the use of R tool, the crime datasets are executed and the suspects with maximum count becomes the near suspect for the police department. So, the above technology based on the supporting attributes and modus operandi will surely help the investigating team to identify the criminals at the short span of time.

References

- [1] S. Sathyadevan, M. Devan, and S. Surya Gangadharan, —Crime analysis and prediction using data mining, in Networks Soft Computing(ICNSC), 2014 First International Conference on, Aug 2014, pp. 406–412.
- [2] Rajanikanth Aluvalu, Tirthraj Chauhan, 2016. “Using Big Data Analytics for developing Crime Predictive Model”, First International Conference on Research & Entrepreneurship (ICRE 2016), Vol.1.
- [3] Lenin Mookiah, William Eberle, Ambareen Siraj, Survey of Crime Analysis and Prediction, Proceedings of the twenty-Eighth International Florida Artificial Intelligence Research Society Conference, 2015.

- [4] Mugdha Jain, Chakradhar Varma, Adapting K-means for Clustering in Big Data, International Journal of Computer Application, Volume 101-No.1, September 2014.
- [5] Keshav Sanse, Meena Sharma, Clustering methods for Big data analysis, IJARCT, Volume 4, Issue 3, March 2015.
- [6] Nivranshu Hans, Sana Mahajan, SN Omkar, Big Data Clustering Using Genetic Algorithm on Hadoop Mapreduce, International Journal of Scientific & Technology Research Volume 4, Issue 4, April 2015.
- [7] Tahani Almanie, Rsha Mirza and Elizabeth Lor, 2015. “Crime Prediction Based on Crime Types and Using Spatial and Temporal Criminal Hotspots”, International Journal of Data Mining & Knowledge Management Process (IJDKP), Vol.5, Issue.No.4.
- [8] Shalini Jain, Satendra Sonare, Big Data Analysis Using HDFS, C-MEANS and Map reduce, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 4, 2015.
- [9] Setu Kumar Chaturvedi, Nikhil Dubey, A Survey Paper on Crime Prediction Technique Using Data Mining, Int. Journal of Engineering Research and Applications, Vol. 4, Issue 3(version 1), March 2014.
- [10] Santhi Baskaran L. Bhuvaneshwari, “Customized Travel Recommendation System using Big Data Analytics”, Customized Travel Recommendation System using Big Data Analytics, Vol.2.
- [11] Aarathi Srinivas Nadathur, Gayathri Narayanan, Indraj Ravichandran, Srividhya.S, Kayalvizhi.J. 2018. “Crime Analysis and Prediction using Big Data”, International Journal of Pure and Applied Mathematics, Vol.119, Issue.12, Pp.207-211.
- [12] Sindhu K. K., Dr. B. B. Meshram, A Digital Forensic Tool for Cyber-Crime Data Mining, An International Journal (ESTIJ), Vol.2, No.1, 2012.