

VISUALIZATION AND CRIME CATEGORIZATION USING CRIME DATASET : AN BIG DATA APPROACH

1. SELVA LAKSHMAN.M

2.SIDDHARTH.M

3.SUBASH.S

4.UDAYAKUMAR.P.P

Department of Electronics and Instrumentation Engineering, 2014-2018
RMK ENGINEERING COLLEGE, Affiliated to ANNA UNIVERSITY

ABSTRACT

Crime frequency have varied over time, with a sharp rise after 1963, reaching a broad peak between the 1960s and early 1990s. A huge amount of data set is generated every year on the basis of reporting of crime. This data can be proven very useful in analyzing and predicting crime and help us prevent the crime to some extent. Crime analysis is an area of vital importance in the crime investigation department. Study of crime data can help us analyze crime pattern, inter-related clues& important hidden relations between the crimes. This is the reason why data mining can be great support to analyze, visualize and predict crime using crime data set. Categorization and inter relation of data set makes it easy to study the similarities & dissimilarities amongst the data objects.[8][3]

Keywords: *crime investigation; categorization; prediction; visualization; data mining; big data analytics;*

I. INTRODUCTION

Crimes are a social nuisance and it has a direct and indirect effect on the society. Governments spend huge sums of money through law enforcement agencies to prevent and curtail crimes from taking place. Today, many law enforcement bodies have large volumes of data related to crimes, which need to be processed to get useful information.

So the main objectives of this project are,

- To aid law enforcement agencies with cost effective data analytics techniques
- For effective management of resources and extraction of information.

Two ways we can look into crime data:

- Descriptive analytics emphasis on identifying spatial and temporal relationships with crime data.
- Predictive analytics methods are mainly used for predicting category of crime which can be occurred somewhere at a given time.

Predictive analytics include,

- Crime Mapping
- Types of Crime Analysis
- The patterns of crimes over time
- Cross comparison and analysis
- Association Analysis: Crime Date Frequency Analysis

II. TECHNOLOGY BEHIND THE PROJECT

For the implementation of the proposed system we require various technologies such as R Programming Language, R Studio, OpenStreetMap, Data Analytics and Big data.[7]

2.1 R Studio

RStudio is a free and open-source integrated development environment for R, a programming language for statistical computing and graphics. R Studio was founded by JJ Allaire, creator of the programming language ColdFusion.

2.2 R

R is a programming language and free software environment for statistical computing and graphics that is supported by the R Foundation for Statistical Computing.

2.3 OpenStreetMap

OpenStreetMap (OSM) is a collaborative project to create a free editable map of the world. The creation and growth of OSM has been motivated by restrictions on use or availability of map information across much of the world, and the advent of inexpensive portable satellite navigation devices. OSM is considered a prominent example of volunteered geographic information. This crowd sourced data is then made available under the Open Database License. The site is supported by the OpenStreetMap Foundation, a non-profit organization registered in England and Wales. Rather than the map itself, the data generated by the OpenStreetMap project is considered its primary output. The data is then available for use in both traditional applications, like its usage by Craigslist, OsmAnd, Geo caching, MapQuest Open, JMP statistical software, and Foursquare to replace Google Maps, and more unusual roles like replacing the default data included with GPS receivers.

2.4 Data Mining

Data mining is the process of discovering patterns in large data sets involving methods at the intersection of machine learning, statistics, and database systems. It is an interdisciplinary subfield of computer science. Aside from the raw analysis step, it involves database and data management aspects, data pre-processing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating.



Fig.1 Stages of Data Mining

2.5 Big data

The large volume of data – both structured and unstructured – that inundates a business on a day-to-day basis are collectively known as Big Data.

2.5 Data analytics:

Data Analytics is the science of analyzing data to convert information to useful knowledge. This knowledge could help us understand our world better, and in many contexts enable us to make better decisions.



Fig.2 Advantages of big data analytics

IV. EXISTING SYSTEM

- The current crime categorization and visualization software focuses only on the location of the crime that had occurred.
- It categories the factors based on the frequency or intensity of the crime.
- This data will not be much sufficient to prevent the rate of crimes that are happening and do not give the insight of the crime pattern.[6]



Fig.3 shows how the existing system appears.

V. PROPOSED SYSTEM

- This project will be including the age factors of people doing crime and people who are being affected by crime.

- This also has the functionality of comparing different states to find which would be the safest one for the people who move into it.
- It also gives information about top ten safest cities and violent cities in each state.
- The year wise crime dataset along with the number of occurrence of each crime type is listed in an OpenStreetMap.
- This project also analyzes the cyclic nature of each crime.

VI. METHODOLOGY

6.1. Data collection:

Enormous amount of crime data is gathered at the end of year at crime bureau records. This data is made available by National Crime Bureau of Records. This data is in the form of number of cases recorded all over the nation round the year. The data is in raw form and also contains some errors as well as missing values. Hence pre- processing of data becomes very essential in order to bring the data in proper and clean form. Pre-processing of data includes data cleansing and Preprocessing.[10]

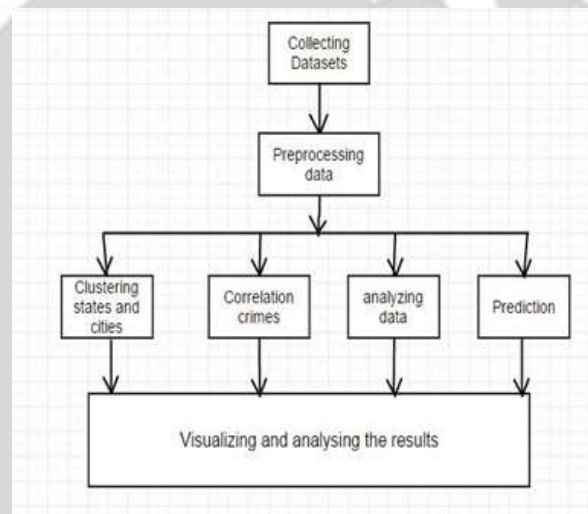


Fig.4 Methodology of crime categorization

6.2. Prediction:

Crimes are predicted using linear regression. Prediction about various types of crimes and most probable places of occurrences of crime will be predicted linear regression is an approach for modeling the relationship between a scalar dependent variable y and one or more explanatory variables (or independent variable denoted X . The case of one explanatory variable is called simple linear regression .In simple linear regression, we predict value of one variable from the value of a second variable. The variable we are predicting is called the criterion variable and is referred to as Y . The variable we are basing our predictions on is called the predictor variable and is referred to as X . When there is only one predictor variable, the prediction method is called simple regression. In simple linear regression, the predictions of Y when plotted as a function of X form a straight line. Linear regression consists of finding the best-fitting straight line through the points. The best fitting line is called a regression line. The formula for a regression line is,

$$Y = b + (a * X);$$

where , Y is the predicted score, b is the slope of the line, and a is the intercept[5][10]

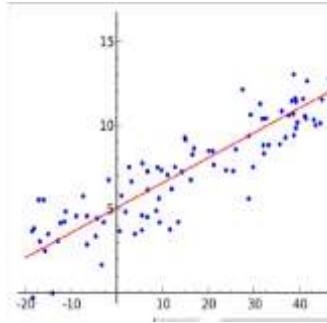


Fig.5 Linear Regression

VII. CONCLUSION

The outcome of this research will be to analyze, correlate and predict the crimes from huge data set available. Results will be in the form of correlation between various violation and location of crime i.e. state/city and the category of violation. Crime can also be correlated on the basis of age group, location of crime & type of crime. Prediction of the crime will be displayed using various diagrams like bar charts, OpenStreetMap, and graph.

VIII. REFERENCE

1. Dr. S. Santhosh Baboo. Article enhanced algorithm to predict a future crime using data mining. International Journal of Computer Applications, 21(1):1–6, May 2011. Published by Foundation of Computer Science.
2. Yujun Bao and Xiaoyan Jiang, “Design of electric energy meter for long-distance data information transfers which based upon GPRS”,ISA 2009, International workshop on intelligent systems and applications, 2009.
3. <http://www.legalmatch.com/lawlibrary/article/what-are-the-different-typesof-crimes.html>
4. <http://www.kdd.org/curriculum/index.html>
5. Data Clustering: Algorithms and Applications by Charu C. Aggarwal, CRC Press Publications
6. Soma ShekaraSreenadh Reddy Depuru,LingefengEang,VijayDevabhaktuni, “Electricity Theft: Overview,issues, prevention and a smart meter based approach to control theft”,Energy policy, vol.39,no.2,pp.10071015,Feb.2011
7. Simson L. Garfinkel, “Digital Forensics: Modern Crime leaves an Electronic trail. Finding and preserving that evidence requires careful methods as well as technical skill, September-October 2013
8. Shyam Varan Nath,”Crime Pattern Detection Using Data Mining”,2006 IEEE/ WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology
9. <http://nlp.stanford.edu/software/jenny-ner-2007.pdf>.
10. www.oracle.com/technetwork/topics/entarch/oracle-wp-big-data-refarch-2019930.pdf
11. <https://www.ethz.ch/content/dam/ethz/special-interest/bse/.../dmc-references.pdf>