

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Special Issue 2, March 2018

Big Data Approach for Crime Classification and Visualization Using Crime Dataset

Sanjana.R, Himaja Sai.S, Sruthi.S

Department of Information Technology, R.M.D. Engineering College, Kavaraipettai, Tamil Nadu

ABSTRACT: Crime rates have varied over time, with a sharp rise after 1963, reaching a broad peak between the 1970s and early 1990s. A huge amount of data set is generated every year on the basis of reporting of crime. This data can prove very useful in analysing and predicting crime and help us prevent the crime to some extent. Crime analysis is an area of vital importance in police department. Study of crime data can help us analyse crime pattern, inter-related clues & important hidden relations between the crimes. That is why data mining can be great aid to analyse, visualize and predict crime using crime data set. Classification and correlation of data set makes it easy to understand similarities & dissimilarities amongst the data objects.

KEYWORDS: crime; classification; prediction; visualization; data mining; big data analytics;

I. INTRODUCTION

Crimes are a social nuisance and it has a direct effect on a society. Governments spend lots of money through law enforcement agencies to try and stop 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 focuses 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
- What are the patterns of crimes over time? In other words, is crime cyclic?
- Cross comparison
- 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.

2.1R Studio

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

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Special Issue 2, March 2018

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

The OpenStreetMap Foundation is a company limited by guarantee, registered in England and Wales on 22 August 2006. It is a non-profit foundation whose aim is to support and enable the development of freely-reusable geospatial data.

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.

2.5 Big data

Big data is a term that describes the large volume of data – both structured and unstructured – that inundates a business on a day-to-day basis.

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.1 Advantages of big data analytics

IV. EXISTING SYSTEM

- The current crime classification 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.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Special Issue 2, March 2018

Fig.2 shows how the existing system looks

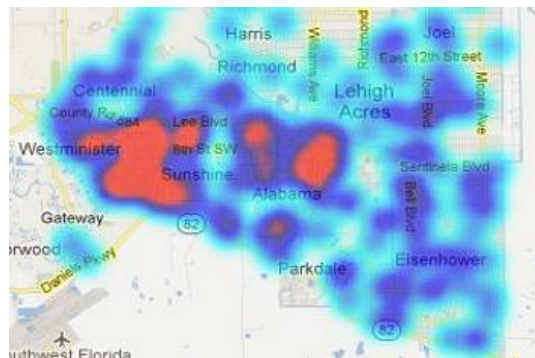


Fig.1 Existing system for crime classification

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 analyse the cyclic nature of each crime.

VI. METHODOLOGY

6.1. Data collection:

Enormous amount of crime data is collected at the end of year at police 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 throughout the year. The data is in raw form and also contains some wrong as well as missing values. Hence pre- processing of data becomes very necessary in order to bring the data in proper and clean form. Pre-processing of data includes data cleansing and Preprocessing.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Special Issue 2, March 2018

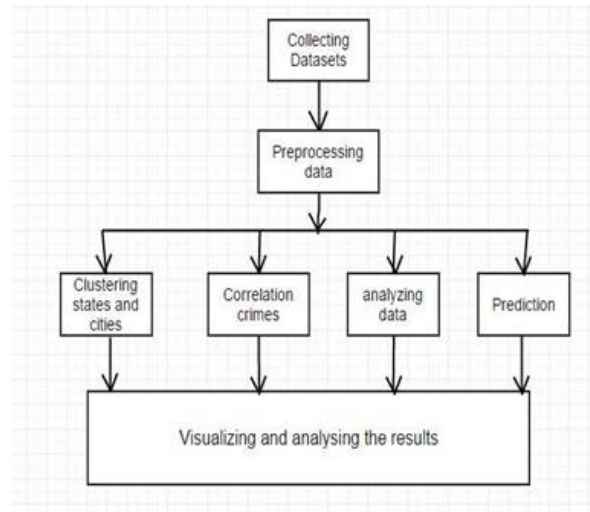


Fig.3 Methodology of crime classification

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. Line a 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

$$Y = aX + b$$

where Y is the predicted score, b is the slope of the line, and a is the intercept

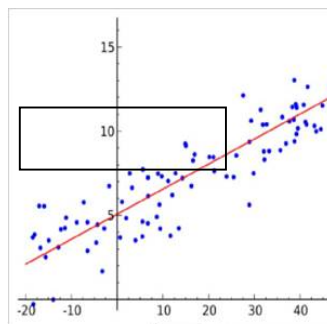


Fig.4 Linear Regression

VII. CONCLUSION

Result of this research will be to analyse, correlate and predict the crimes from huge data set available. Results will be in the form of correlation between various crime and location of crime i.e. state/city. Crime can also be correlated on

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Special Issue 2, March 2018

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.

REFERENCES

1. Malathi. A and 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. Eibe Frank and Remco R. Bouckaert. Naive bayes for text classification with unbalanced classes. In Proceedings of the 10th European Conference on Principle and Practice of Knowledge Discovery in Databases, PKDD'06, pages 503–510, Berlin, Heidelberg, 2006. Springer-Verlag.
3. Ralf Hartmut Güting. Graphdb: Modeling and querying graphs in databases. In Proceedings of the 20th International Conference on Very Large Data Bases, VLDB '94, pages 297–308, San Francisco, CA, USA, 1994. Morgan Kaufmann Publishers Inc
4. Lior Rokach and Oded Maimon. Decision trees. In Oded Maimon and Lior Rokach, editors, The Data Mining and Knowledge Discovery Handbook, pages 165–192. Springer, 2005.
5. <http://www.legalmatch.com/lawlibrary/article/what-are-the-different-typesof-crimes.html>
6. <http://www.kdd.org/curriculum/index.html>
7. <http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/palace/datamining.html>
8. Data Clustering: Algorithms and Applications by Charu C. Aggarwal, CRC Press Publications
9. Data Mining: Concepts and Techniques by Jiawei Han, Micheline Kamber, Jian Pei, Morgan Kaufmann Publishers
10. Tong Wang, Cynthia Rudin, Daniel Wagner, and Rich Sevieri, "Learning to Detect Patterns of Crime", available at: <http://web.mit.edu/rudin/www/docs/.pdf>
11. Arthisree K.S and Jaganraj A, "Identify Crime Detection Using Data Mining Techniques", ISSN: 2277 128X, © 2013 IJARCSSE
12. Shyam Varan Nath, "Crime Pattern Detection Using Data Mining", 2006 IEEE/ WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technolo
13. <http://nlp.stanford.edu/software/jenny-ner-2007.pdf>.