

Analyzing Crimes in India using R Programming

Himanshu Aggarwal^{1, #}, Aditya Rawat¹, Ms. Narinder Kaur¹

¹*Dept. of Information Technology, Maharaja Agrasen Institute of Technology, Delhi, India*

[#]Corresponding Author, Email: *himanshuea@gmail.com*

Abstract— There has been a rapid increase in the crime in the recent past years. Crime prevention has become a furore task. The cops in their role to catch offenders are required to remain effectively ahead in the eternal race between villains and heroes. One of the key affairs of the police is how to improve investigation's effectiveness. There is need for user interactive interfaces based on current technologies to give them the much needed edge and fulfil the new emerging responsibilities of the police. These sad scenarios compel us to analyse the crime data with an aim of acknowledging the nature and depth of the problem and predicting the future crime rates. The paper highlights the predicted crime rates for near future which can be used by Indian police in curbing increasing crimes and using their resources more efficiently. Paper aims at identifying crime prone spots with help of data analysis technique. The technique used is linear regression.

Keywords— Crimes in India, Data Analysis, Linear Regression, R programming

1. INTRODUCTION

Crime is an act of breaking the given set of rules and regulations set by the legal authorities. A crime is indictable on the basis of its extent and severity. Instead crime is subjectively defined as anything that violates the proscriptions on behaviour established by the power elite.

For our aimed research, the case involved for crime analysis is India. It is the 7th largest nation in the world in terms of area. Detailed crime analysis helps to predict the crime rates for upcoming future. Crime in India is surveyed by the National Crime Records Bureau (NCRB) [1]. NCRB

is responsible to collect and evaluate the information and on the basis of which, comparison is made subsequently. Crimes in India is accounted in various forms chiefly; Robbery, Rape, Dowry, Extortion, Kidnapping, Sex Violations. The list is uncountable. There are horrific cases of Organised Crimes as well, like Arms Trafficking and Illegal Drug Trade.

As technology is improving the world in a long run where it rotates around a large, immense data, if this data is analysed in a right direction, then we can eliminate the existing problems. Data Analysis helps in processing of large amounts of data and discovering hidden information. Our System's objective is to process huge amount of criminal records for different crimes and for every state so that prediction can be made for future use. The data for the proposed research is obtained from OGD (Open government data) platform (<http://data.gov.in>).

2. METHODOLOGY

Data Analysis is the act of trying to learn something from a dataset. It is the method that can be used to analyse data and the process of analysing it. There are many different forms of data, but people usually think of quantitative data first. These are data such as census data or survey data. There is also scientific data, such as data physicists collect about the cosmos. These days, most people are interested in analysing the vast amounts of data

collected through transactions with a variety of companies and/or websites [2]. Often, though, people forget about qualitative data, which is data that has not yet had any form imposed on it by humans. There are several phases that can be distinguished, as shown in Fig. 1.

3.2 Data Collection

Enormous amount of crime data is collected at the end of year at police records. This data is made available by OGD platform. This data is in the form of number of cases recorded all over the nation throughout the years. 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 Pre-processing.

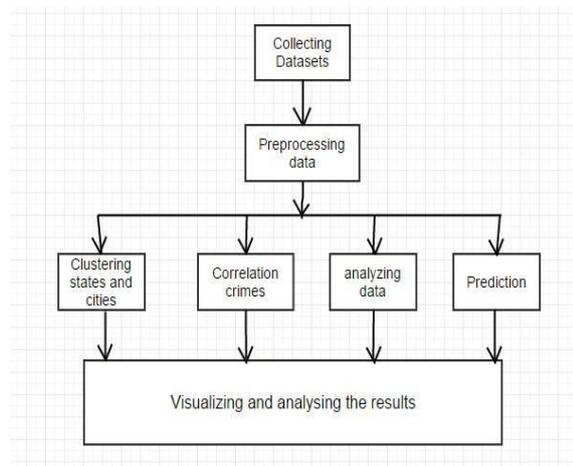


Fig. 1: Methodology for crime pattern analysis

2.2 Data Classification

We classify the data set into various groups based on certain characteristics of the data object here we group crimes according to states & union territories. Classification of the crime is done on the basis of different types of crime.

3. ANALYSIS USING R PROGRAMMING

In this research we take a look at the gory crime scene across India to determine

which states are the hot spots in crimes. To get the answers to these questions we performed analysis of the state-wise crime data with the data taken from (OGD). OGD is an online platform which stands for Open Government data. It provides dataset on various topics. The information present in OGD is available for crimes in different states under different ‘crime heads’ like Murder, rape, theft, dowry deaths, dacoity, burglary, kidnapping & abduction etc. The data is available for years from 2001 to 2016. This data is plotted as a scatter plot and a linear regression line is then fit on the available data.

Based on this linear model, the projected incidence of crimes likes rapes, dowry deaths, abduction & kidnapping is performed for each of the states. This is then used to build a table of different crime heads for all the states predicting the number of crimes till the year 2020. Fortunately, R crunches through the data sets quite easily.

3.1 Linear Regression

Regression analysis is a statistical method that is used to analyse the relationship between a dependent variable and one or more independent variables (this also be extended in many different ways) [3]. The independent variable is the predictor variable whose value is congregated through experiments. The dependent variables are called response variables whose value is obtained from the predictor variable.

Linear Regression is a mathematical approach to compute the value of a particular variable based on other variable(s) as shown on Fig. 2 [4]. The general mathematical equation $Y = mX + c$, which is the equation for a straight line is used in linear regression, where:

- Y is the Dependent Variable
- X is the independent Variable
- m and c are constants which are called the coefficients.

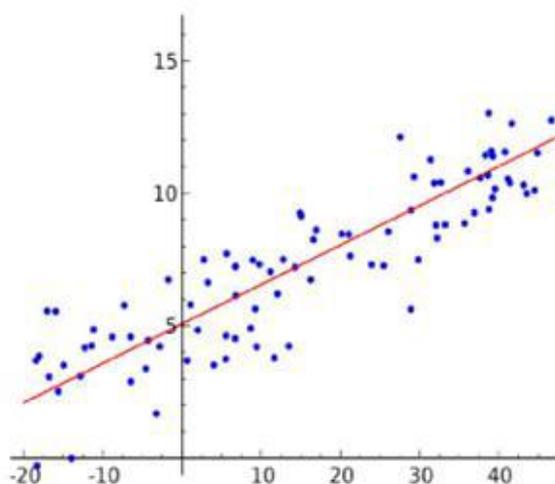


Fig. 2: Linear Regression

3.2 Data Analysis Concepts And Techniques Pre-processing

Data set was made available from OGD Platform [6]. The number of years available and used was between 2001 and 2016.

Data Attributes: The following attributes were used and presented in the data set for the city crime statistics.

- Property
 - Robbery
 - Burglary
 - Theft
 - Dacoity
- Violent
 - Attempt To Commit Murder
 - Murder
 - Culpable Homicide Not Amounting To Murder
- Crime against Women
 - Rape
 - Dowry Death
 - Assault On Women With Intent To Outrage Her Modesty
 - Cruelty by Husband and her relatives.
 - Insult To Modesty Of Women
- Others
 - Abduction & Kidnapping of others
 - Criminal Breach of Trust
 - Arson
 - Cheating
 - Counterfeiting
 - Others IPC crimes

4. EXECUTION IN R STUDIO

The methodology of regression is predicting crime rates when years are known [5]. To do this we need to have the relationship between crime rates and corresponding years.

Following steps are used in order to establish the relationship:

- I. Experiment of gathering a sample of observed values of years and crime values is carried out.
- II. `lm()` function in R is then used to model the relationship.
- III. Find the coefficients from the model created and create the mathematical equation using these.
- IV. Coefficients are then used to predict the future crime rates for subsequent years.

5. PREDICTION IN CRIME TRENDS

The quest is the prediction of future crime rates for every state of India. This means that we analysed the crime values changing from one year to the next and used data analysis to propose those changes into the future. The basic methodology used here is linear regression model which predicts future crime rate and thus identify hot spots in crime. The state with highest crime rate is the hotspot. The predicted values are then plotted on a scatter plot. Since Uttar Pradesh is one of the leading states when it comes to crimes, the plots under different crime heads for Uttar Pradesh are shown.

Each scatter plot below depicts crime rates for years ranging from 2001 to 2020. The circular points on the graph shows the data available on different crime from year 2001 to 2016 and triangular shaped point depicts corresponding predicted crime values from 2017 to 2020. Similarly future crimes rates are predicted for 29 states and 7 union territories under different crimes heads. Predictive analysis shows the following trends for different crimes:

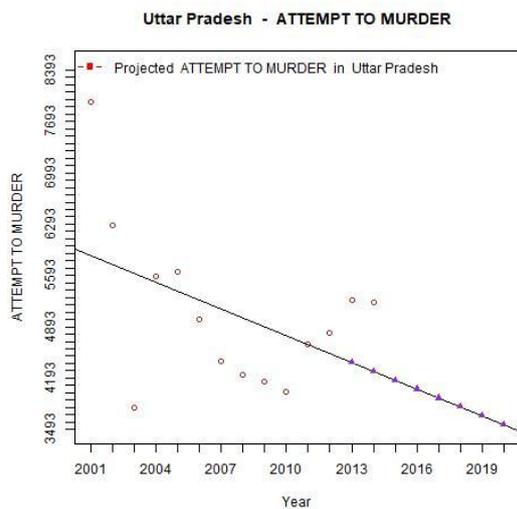


Fig. 3: Attempt to murder rates in UP

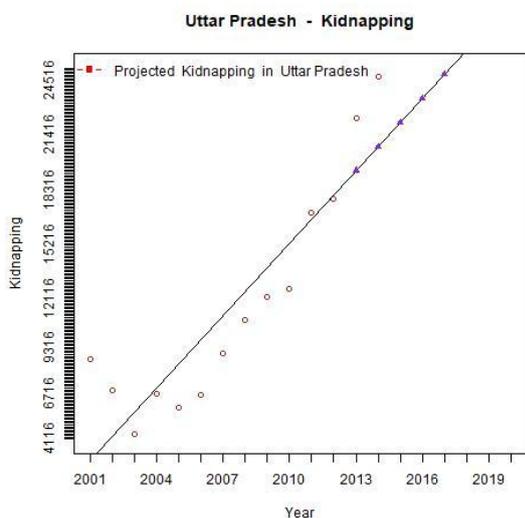


Fig. 4: Kidnapping rates in UP

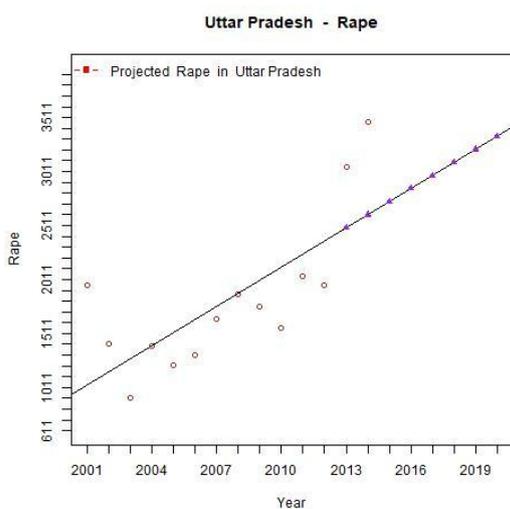


Fig. 5: Rape crime rates in UP

- **Arson** (the criminal act of deliberately setting fire to

property)- is prominently higher in the regions of Maharashtra and Assam with over thousand such cases.

- **Attempt to Murder** – Uttar Pradesh and Bihar have relatively higher possible rates than other states.
- **Rape** – Again Madhya Pradesh, Rajasthan, Uttar Pradesh, West Bengal and Assam are considered as hot spots for Rape crime with crime rates touching over 3000 cases in a year. Whereas Goa, Chandigarh, Daman & Diu, Lakshadweep Sikkim, Dadar & Nagar Haveli are considered very safe place for women with Lakshadweep having only one registered rape case.
- **Burglary** (trespassing into someone else’s property with an intent to commit a crime)- Maharashtra is again the most prominent state with over 17 thousand such cases. It is followed by Madhya Pradesh with cases upto 11 thousand.
- **Causing Death by Negligence** – Tamil Nadu is predicted to have the most number of cases under this crime head. Predicted value for Tamil is 15476, followed by Uttar Pradesh with cases upto 14644. Maharashtra and Andhra Pradesh have predicted crimes rates over 13 thousand. Thus require awareness and education among people.
- **Dowry Deaths**- Almost every state has shown decline in Dowry Death cases except Uttar Pradesh and Bihar. Prediction for Uttar Pradesh shows almost 2500 cases for subsequent years whereas Bihar has predicted crime rate value of over 1500. Hence require strict policy.
- **Counterfeit** (made in exact imitation of something valuable with the intention to deceive or defraud)- This crime head as low

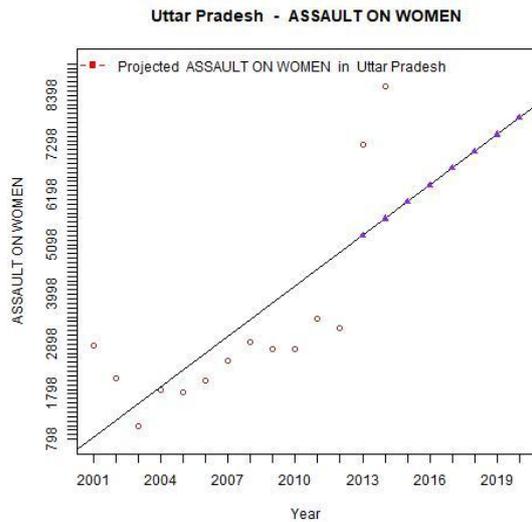


Fig. 6: Assault on women in UP

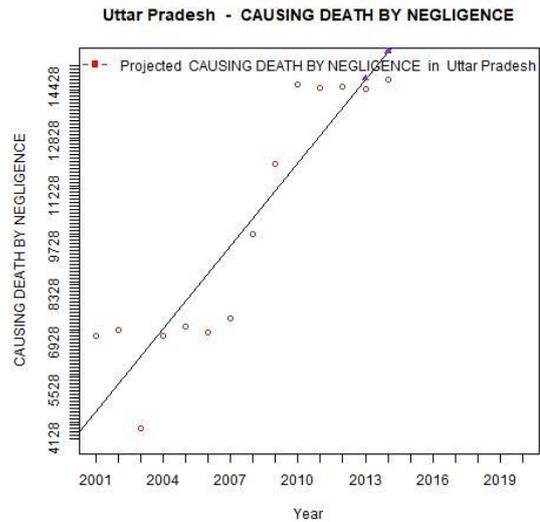


Fig. 8 Causing death by negligence in UP

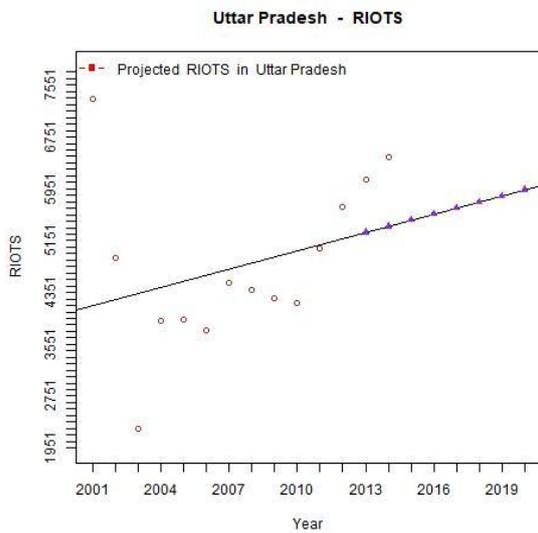


Fig. 7: Riots crime rates in UP

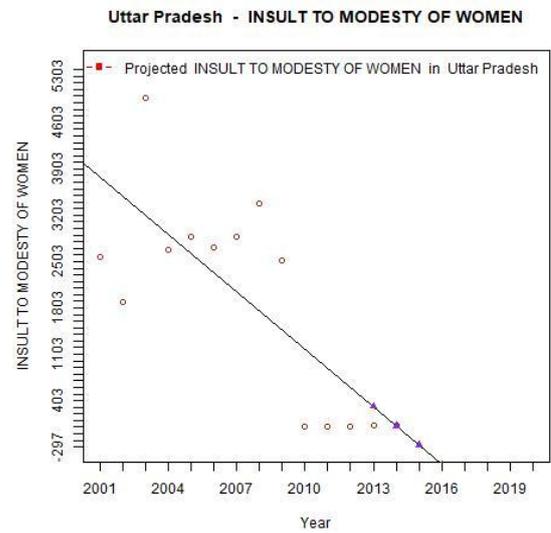


Fig. 9: Insult To Modesty To Women in UP

rates as compared to other crime heads. With upto 329 cases, Tamil Nadu has the highest rate.

- **Cruelty on Women by Husband or his Relatives-** West Bengal is predicted to have 24230 case in year 2018, followed by Rajasthan with 15631 and Andhra Pradesh with 13775 cases. States such as Madhya Pradesh, Maharashtra and Kerala have an average of 4000 such cases.

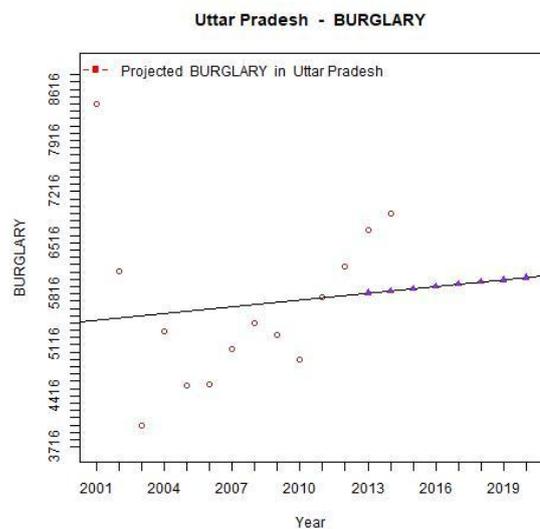


Fig. 9: Burglary in UP

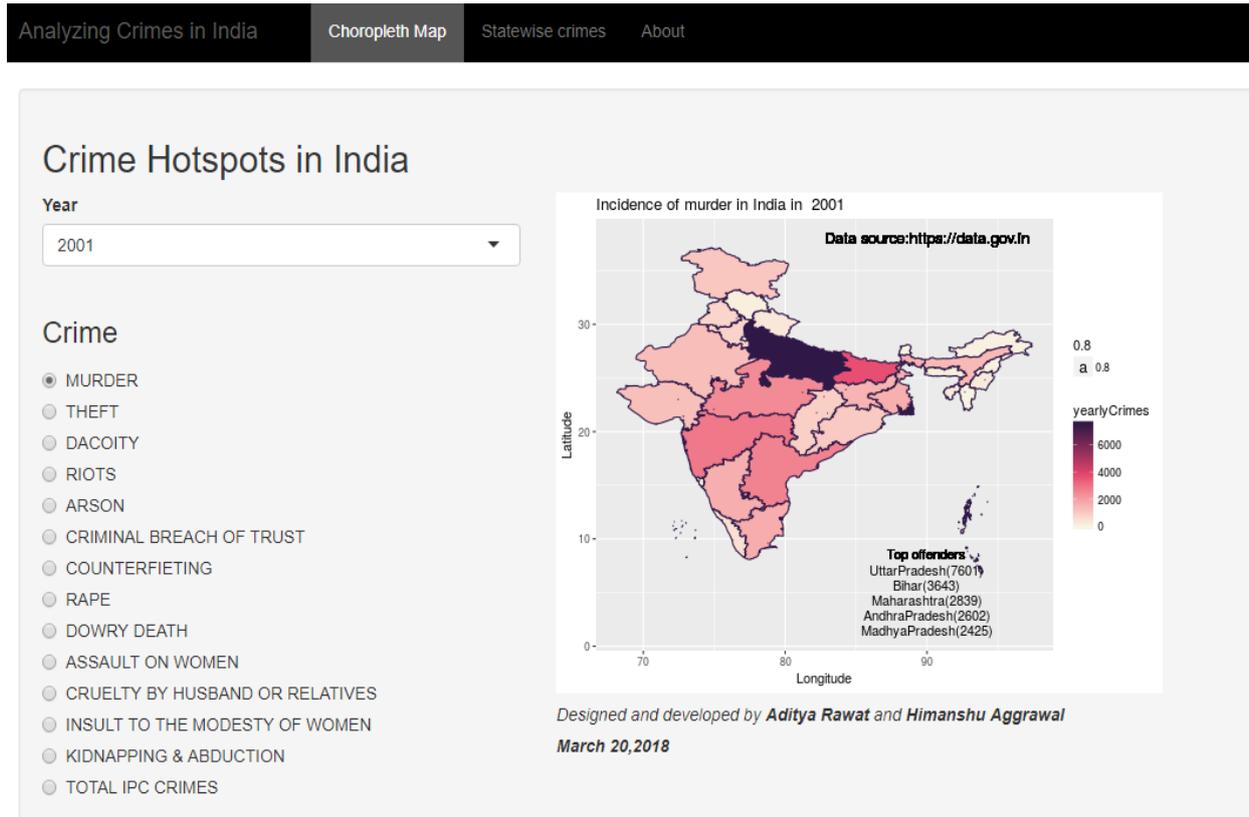


Fig. 10: Crime Hotspots in India

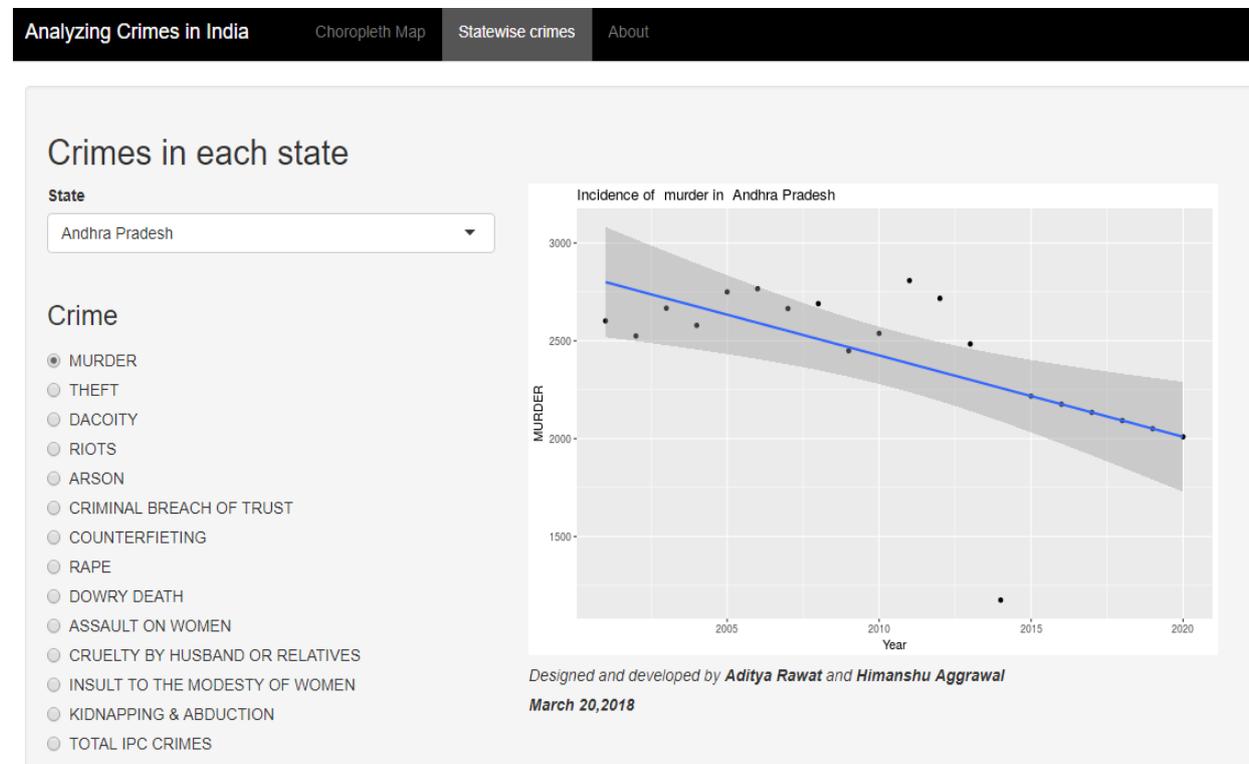


Fig. 11: Crimes in Each State

- **Theft-** It is one of the highest committed crimes in every state. Maharashtra has the highest predicted crime rate with a whopping value of 57215. It's being followed by Uttar Pradesh with 43050 cases. States such as Andhra Pradesh, Madhya Pradesh, Gujarat, Rajasthan and Harayana have more than 25000 predicted cases for year 2018.
- **Dacoity**(an act of violent robbery committed by an armed gang)- Dacoity is predicted to be highest in Maharashtra with value of 938. It being followed by Bihar and Karnataka with an average of 300 cases for year 2018.
- **Culpable Homicide Not Amounting to Murder-** This crime head is predicted to be highest in Uttar Pradesh with value of 1448 for year 2018. It's followed by West Bengal with crime rate of 488.
- **Kidnapping & Abduction-** Uttar Pradesh is yet again the major hot spot for kidnapping with predicted value of upto 23484 cases. West Bengal, Assam, Bihar and Rajasthan are predicted to have more than 11000 cases.
- **Insult to Modesty of Women-** Andhra Pradesh wins here with up to 4422 cases for year 2018.
- **Murder-** Uttar Pradesh is again the leader here with 4051 cases. Followed by Bihar with 3174 cases.

6. CRIME ANALYSIS TOOL

A crime analyzing tool is developed and deployed at <https://addie11.shinyapps.io/AnalyzingCrimesInIndia/> The tool provides two main tabs for different purpose as described below:

- The first tab plots choropleth Map of India for different crimes for year 2001 to 2020. The crime predictions made for year 2017 to 2020 is used

to plot the choropleth map for future crimes and hence identify the hotspots. It allows us to instantly visualize the states having the highest criminal cases under IPC (Indian Penal Code) for different crimes and outputs the top five offenders for each crime. The far north, north east and west of the country seem to be relatively safer as compared to the centre and north side as shown on Fig. 11.

- Another tab (State wise crimes) is used for plotting crime graph for each state of India. The plotted graph shows the variation of crime trend over the year for every state. Hence it may help government to deploy more police force in states which show increasing crime rates as shown in Fig. 12.

7. CONCLUSION

To alleviate crime related problem data analysis has been applied in the context of law enforcement and intelligence. To anticipate crime trends we use linear regression model in this paper. To analyse the city crime, data from Open Government Data (OGD) is taken and the predictive analysis technique is used. To lessen and even prevent crime in forth coming years the results of this data analysis could potentially be used. From the encouraging results, we believe that crime data analysis has a promising future for increasing the effectiveness and efficiency of criminal and intelligence analysis.

Many future directions can be explored in this still young field. Visual and intuitive criminal and intelligence investigation techniques could be developed for crime pattern.

ACKNOWLEDGMENT

We express our deep gratitude to Mr. Nitesh Kumar, Project Coordinator, Assistant Professor, Department of

Information & Technology Engineering for his valuable guidance and suggestion throughout my project work. We are thankful to him for his valuable guidance. We would like to extend my sincere thanks to Head of the Department, Dr. M.L. Sharma for his time to time suggestions to complete our project work. We are also thankful to Prof. (Dr.) M.L. Goyal, Director for providing us the facilities to carry out our project work.

REFERENCES

- [1] K.S.Arthisree, M.E, A.Jaganraj, Arulmigu Meenakshi, "Identify Crime Detection Using Data Analysis Techniques", International Journal of Advanced Research in Computer Science and Software Engineering, Amman College Of Engineering, Thiruvannamalai district, Near Kanchipuram, India, volume 3, Issue 8, August 2013.
- [2] B.Chandra and M. P.Gupta, "Crime Data Mining for Indian Police Information System Manish Gupta", International journal of Applied Research in Computer Science.
- [3] D.E. Brown, "The regional crime analysis program (RECAP): A frame work to catch criminals", In Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, Vol. 3, pp. 2848-2853, 1998.
- [4] Ram Singla, Deepak Dembla, Yogesh Chaba and Kevika Singla CITM, "An Optimal KD Model for Crime Pattern Detection Based on Semantic Link Analysis", Journal of Japee Institute Of Technology, volume IC3, Issue 2008.
- [5] Malathi, Dr. S. Santhosh Baboo Reader, "An Enhanced Algorithm to Predict a Future Crime using Data Mining" ,International Journal of Computer Applications (09758887), Volume 21 No.1, May 2011.
- [6] TongWang, Cynthia Rudin, Daniel Wagner, and Rich Sevieri, "Learning to detect patterns of crime", Journal of Environmental science, computer science and engineering & technology. 1:124131, 2012.