# Crime Analyser: A Comprehensive Analysis of Indian Crime Database and Trend Forecasting Technique

Tushar Agarwal
Amity School of Engineering Technology
Amity University
Haryana, India
tushar007720@gmail.com

Arnav Varyani
Department of Computer Science
Thadomal Shahani Engineering College
Mumbai, India
arnavvaryani@gmail.com

**ABSTRACT**

In recent years, India has witnessed a significant and alarming increase in crime, posing a grave threat to the security and well-being of its citizens. Recognizing the gravity of this situation, the chapter aims to comprehensively study these crime patterns and analyze the changes in overall crime based on data obtained from official Indian Government websites. By employing data mining techniques such as eliminating missing values, redundant data, and data transformation, the study converted the raw data into a suitable format for further analysis. To gain meaningful insights, it utilized algorithms such as Linear Regression and Random Forest, to perform crime-type predictions. These predictions cover a span of four years for each state, as well as the entire nation, using data from the years 2001 to 2016. The results of these predictions are presented using simple yet effective visualization charts, allowing for easy interpretation and understanding. Crime analysis and prediction is a new breakthrough in the field of Knowledge discovery from data and Data Science. This study not only comprehensively maps different Indian states with their criminal records in 4 major Streams of crimes: Crimes against women, Crimes against Children, State and local laws, and IPC crimes but also finds insights from this mapping to find fault zones encouraging criminal activities in India. The study further utilizes Machine Learning techniques to find the areas and their associated crime types that need to be worked on in the future to provide efficient legislation.

**Keywords**—Data Science; Machine Lrearning; Deep Learning; Crime Analysis; Trend Analysis;

## I. INTRODUCTION

Today's information age has seen the spread of criminal activities in any and all domain one can imagine. Criminal activities are broadly categorized in 4 major categories: Crime against Women, Crimes against Children, State and local laws and Indian Penal Code Crimes. Data for criminal offences though extensively stored and formulated is m=not ventured enough. Data science techniques have recently taken the edge with their wide range of applications, be it healthcare of agriculture, data tends to be the most expensive commodity that can make or break any business' future. However one such application of data science is analysis crime records and gathering insights to reduce criminal activity. Crime Mapping systems hence are the need of the hour. The chapter aims to develop a Predictive Analytics Approach for Visualizing Crime Patterns in India.

In the scope of this Chapter an innovative website dedicated to researching and analyzing crime patterns in India is generated and walked through. The country has seen a significant increase in crime in recent years, posing a threat to national stability and the wellbeing of its citizens. To address this perilous situation, the chapter intends

to investigate crime patterns using data obtained from official Indian Government websites. Data transformation and preprocessing Techniques are discussed with much detail in the scope of this chapter to facilitate learning their usage for similar yet diverse problem domain. Data mined from these websites after cleaning was trained on machine learning algorithms like linear regression and random forest to obtain trend forecasts for different types of crimes associated with different states of Indian Territory. These forecasts are then presented in simple yet affective visualization charts, providing users with valuable insights into criminal activity trends in India.

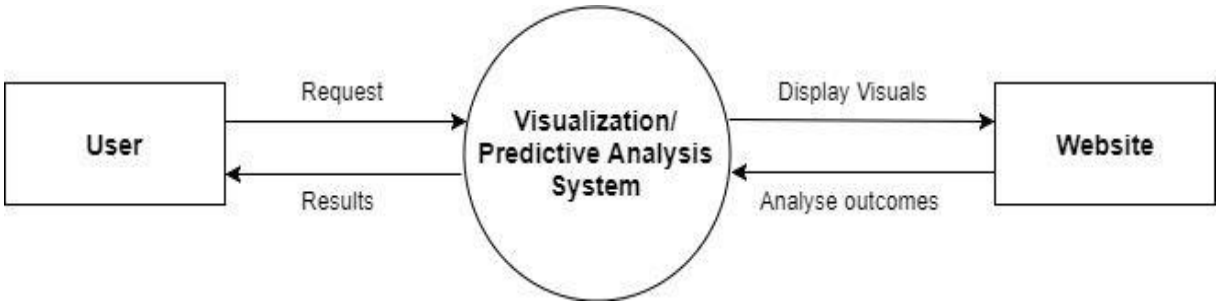


Figure 1. Architectural DFD

The study used Census India demographic data to create visualizations of crime rates per lakh people, categorizing each state based on factors such as literacy and area. These visualizations give forecast analyzers a complete picture of the crime landscape, allowing them to see correlations between crime rates and various demographic factors. The study further offers a user-friendly environment in which all users can benefit fully from analyzing crime trends across all Indian states. By leveraging data-driven insights and Forecasted trends this study gives a comprehensive detailed structure of criminal activities of the past and also predicts their trends for the future, making the model a success if implemented practically.

Objectives:

a) Historical Criminal Database Analysis and Knowledge discovery from data.
b) Drawing efficient visualizations from extracted knowledge to provide structured study resources
c) Using available data to predict future criminal activity associated with different states and their literacy rates.
d) Making a user-friendly interface for fetching aforementioned objectives at the fingertips.
e) Most important of all, discussing and finding aptness of data science techniques in similar problem domains.

The organization of the rest of the article is as follows: Section II inspects the prior studies and analyses their weaknesses and strengths. The methodology of the proposed scheme is illustrated in Section III. The analysis and discussion of the results are described in Section IV. The summary and conclusion of the proposed work are presented in Section V.

## II. RELATED STUDIES AND LITERATURE REVIEW

Tushar Sonawanev et al proposed a study collecting raw criminal records followed by data preparation and characterization based on states/cities and criminal activity profiles. The study used the K-Means clustering technique while associating different crimes together based on Pearson's correlation coefficient [1]. Similarly, Malithi A et al proposed an improved study using data mining algorithms. Missing Data values were handled efficiently by a mix of the K Nearest Neighbor algorithm and learning vector Quantization. Further clustering algorithms included the K Means clustering and DBscan techniques. [2]

Sarpreet Kaurl et al proposed a diverse and refreshing review paper disseminating many different papers and journals published on the topic. Different aspects of the papers like the method used, difficulties, and topics addressed were considered in categorizing research papers for success. [3] However, Neeru Mago et al proposed a different approach emphasizing the use of crime and criminal information systems (CCIS) in the context of

computer-aided crime analysis tools in India. The proposed method includes modules like crime data entry and crime information retrieval. However, this study does not mention any one specific technique used. [4]

The study proposed by Ahmad Shafeek et al deploys data mining techniques to investigate the spatial patterns of crimes in India. Factors like GDP, Literacy Rate, Police Rate and Employment Rate were considered in the making of this model.[5] Similarly, D.Gnana Rajesh and Dr. M. Punithavalli proposed a study examining depth tools like ATAC, Crime Stat, RCAGIS, Crime Connect, and BRAINCEL. While the paper conducts a basic overview of these tools no specific approach is highlighted.[6] S.Yamuna and N.Sudha Bhuvaneswari in the proposed work describe a similar method of crime data analysis using the K-Means clustering and DBscan techniques. However, the paper does not provide any information on trend forecasting and prediction.[7]

A data representation model using Geospatial plots was proposed by P.DHAKSHINAMOORTHY et al, however, the uncertainty of crime patterns renders this approach insufficient [8] similarly Muffeda M proposed a study using Decision trees, Random Forest, and Support vector Machines to perform trend forecasting and Crime Predictions.[9]. Hajela et al in the study propose a classification-based multi-dimensional model of spatial pattern recognition to predict crime patterns.[10]

**Table 1: Related Work**

| Author Citations | Techniques Used | Gap Areas |
|---|---|---|
| [1] | K Means Clustering | No trend Forecasting |
| [2] | K Nearest Neighbor and LVQ | The created tool was not public friendly |
| [3] | Review Paper, No technique Mentioned | No new findings |
| [4] | CCIS, ICIAS | No one specific approach |
| [5] | Used Factors like GDP, Literacy rate | No one specific approach |
| [6] | Review Paper: examines tools | No specific approach |
| [7] | K Means Clustering and DBSCAN | No predictions and forecasting |
| [8] | Visualization based on geospatial Plots | No trend forecasting or predictions |
| [9] | Decision Trees, Random Forest and SVM | Crime Pattern Visualization Missing |
| [10] | GIS | No visualizations |

## III. METHODOLOGY

Crime pattern analysis and prediction is an application of data discovery techniques used to fetch knowledge from existing data sources and then use that knowledge to predict the future.
The first step of any Knowledge discovery process starts with the Acquisition of Data followed by data preprocessing.

a. Data Acquisition: Data acquisition refers to the process of acquiring data through various sources and integrating the same into one data warehouse. For the scope of this chapter, the data was fetched from two different sources: The National Crime Records Bureau (NCRB) and Open Government Data Census India.

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | STATE/UT | CRIME HEAD | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 |
| 2 | Andhra Pradesh | INFANTICIDE | 1 | 1 | 3 | 1 | 1 | 0 | 1 | 0 | 9 | 6 | 0 | 8 | 7 | 8 | 6 | 16 |
| 3 | Arunachal Pradesh | INFANTICIDE | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | Assam | INFANTICIDE | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 5 | Bihar | INFANTICIDE | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 2 | 3 | 1 | 0 | 1 | 0 | 0 |
| 6 | Chhattisgarh | INFANTICIDE | 6 | 6 | 9 | 12 | 9 | 8 | 8 | 7 | 3 | 1 | 8 | 8 | 15 | 5 | 4 | 1 |
| 7 | Goa | INFANTICIDE | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 8 | Gujarat | INFANTICIDE | 1 | 4 | 0 | 5 | 2 | 0 | 7 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| 9 | Haryana | INFANTICIDE | 0 | 6 | 5 | 2 | 0 | 10 | 1 | 4 | 0 | 7 | 0 | 0 | 0 | 3 | 5 | 1 |
| 10 | Himachal Pradesh | INFANTICIDE | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 2 | 0 |
| 11 | Jammu & Kashmir | INFANTICIDE | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |
| 12 | Jharkhand | INFANTICIDE | 3 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 3 | 1 | 0 | 1 | 4 | 0 | 0 |
| 13 | Karnataka | INFANTICIDE | 17 | 3 | 0 | 8 | 5 | 9 | 3 | 13 | 4 | 2 | 8 | 3 | 4 | 11 | 0 | 2 |
| 14 | Kerala | INFANTICIDE | 1 | 1 | 0 | 2 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 4 | 0 |
| 15 | Madhya Pradesh | INFANTICIDE | 31 | 18 | 24 | 23 | 28 | 43 | 29 | 22 | 12 | 20 | 13 | 17 | 8 | 14 | 25 | 14 |
| 16 | Maharashtra | INFANTICIDE | 26 | 24 | 11 | 6 | 3 | 4 | 2 | 3 | 1 | 3 | 3 | 11 | 6 | 12 | 7 | 7 |
| 17 | Manipur | INFANTICIDE | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 18 | Meghalaya | INFANTICIDE | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| 19 | Mizoram | INFANTICIDE | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 20 | Nagaland | INFANTICIDE | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 21 | Odisha | INFANTICIDE | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 1 |
| 22 | Punjab | INFANTICIDE | 3 | 0 | 8 | 7 | 9 | 2 | 3 | 2 | 6 | 8 | 1 | 6 | 7 | 8 | 3 | 8 |
| 23 | Rajasthan | INFANTICIDE | 3 | 2 | 9 | 1 | 1 | 6 | 1 | 4 | 2 | 7 | 3 | 3 | 8 | 33 | 18 | 14 |
| 24 | Sikkim | INFANTICIDE | 0 | 0 | 0 | 1 | 1 | 3 | 0 | 2 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 |
| 25 | Tamil Nadu | INFANTICIDE | 1 | 16 | 14 | 17 | 1 | 1 | 10 | 20 | 9 | 7 | 7 | 6 | 13 | 0 | 2 | 2 |

Figure 2: Data Snippet

b. Data Preprocessing: Data preprocessing techniques refer to the process of converting raw data into usable data formats. For the scope of this chapter Data preprocessing steps include: Eliminating Redundant records, Handling NA values, and Data Transformation.[11]

Exploratory Data Analysis refers to the process of analysis and visualizing data points for further knowledge mining from data.[12] The scope of this chapter Includes performing exploratory data Analysis in 3 steps:

a. Future Crime Rate Prediction: The module provides future predictions for crime patterns and trend forecasts for the next 4 years based on historical data.
b. Comparative Analysis: This module would present the comparative observations for all 4 major streams of crimes aforementioned in the chapter.
c. Visualizations: This module is for representation purposes. This deals with generation of visual graphics to make analysis easier.

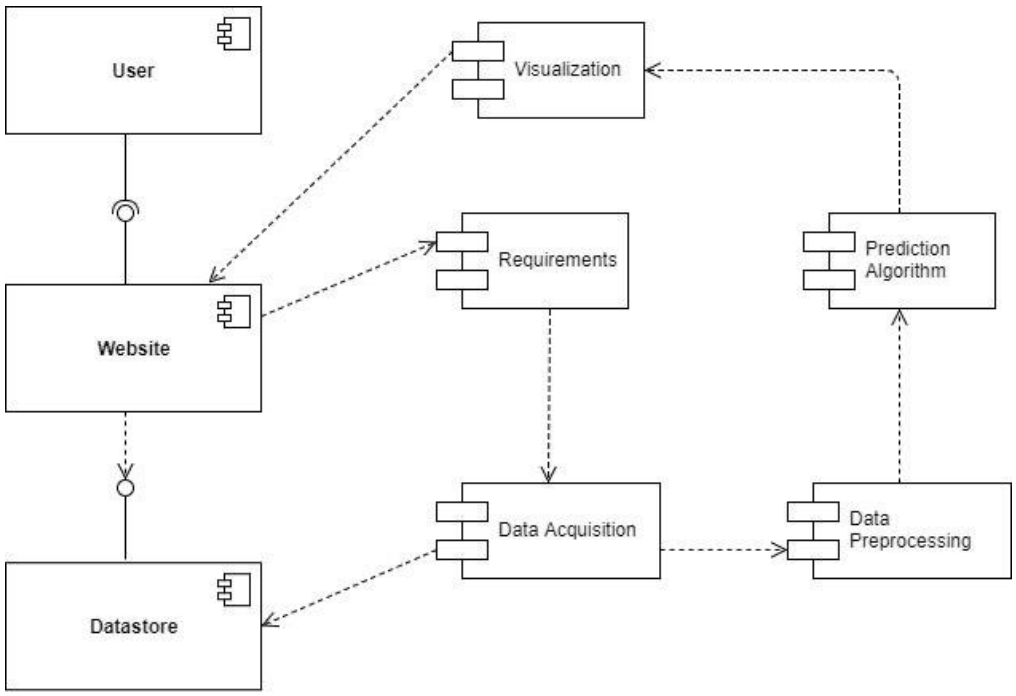


**Figure 3: Component Diagram**

The EDA process consisted of a predictions module, this module was trained on historical data to provide predictions for crime trends for the next four years. The prediction threshold was set to 65% ie any and all predictions having an accuracy lower than 65% were discarded and no visualization graphs were made for them. This mathematical algorithm present in the script tags of HTML files helped the study to maintain efficient results. This process was accomplished with the help of 2 states of the art machine learning algorithms:

a. Linear Regression: This technique is used to calculate the values of a variable based on other variables available.
b. Random Forest: Random Forest is an agglomeration of multiple decision trees together to facilitate more accurate results.

Further Processes include the creation of a FLASK API to provide all of the aforementioned services to end users. The API rendered with different HTML files and Script codes provided not only easy access to the algorithm but also user susceptibility to the algorithmic environment. The API also made it easier for the end user to navigate from one functionality to other, bringing important yet subtle tasks of prediction, visualization and trend forecasting to their fingertips.

Architectural Structure:
The diagram below represents the entire system architecture clearly representing all and each module of the setup. The structure is however as follows:

a. Invention of Problem
b. Data Acquisition

c. Data Preprocessing
d. Data Analysis
e. Crime Ratio Prediction
f. Comparative Analysis
g. Visualizations

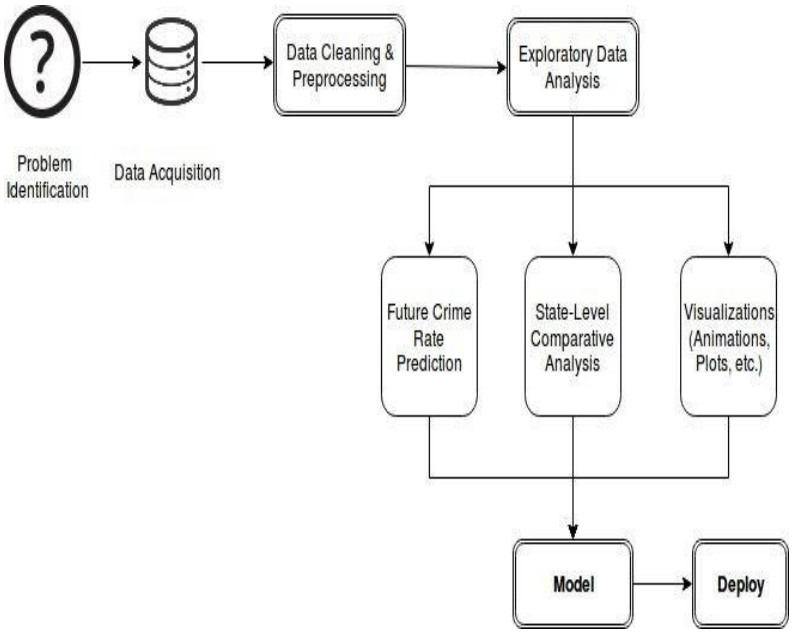


**Figure 4: Architectural Structure**

## IV. RESULTS AND DISCUSSIONS

The Creation of the Flask API led to the formation of a responsive web API, This API facilitated Crime Data Analysis, Crime Rate Prediction, and Crime visualizations. All of the aforementioned functionalities were stored as separate Python/HTML/Javascript code files depending on the use case and were rendered into the API.
A threshold value of 65% for predictions made it easier for the model to refrain from garbage output generation.[13]

a. Look and Feel of the Website: A dynamic responsive web API creation was needed to enhance user experience and make the model available for practical implementation.

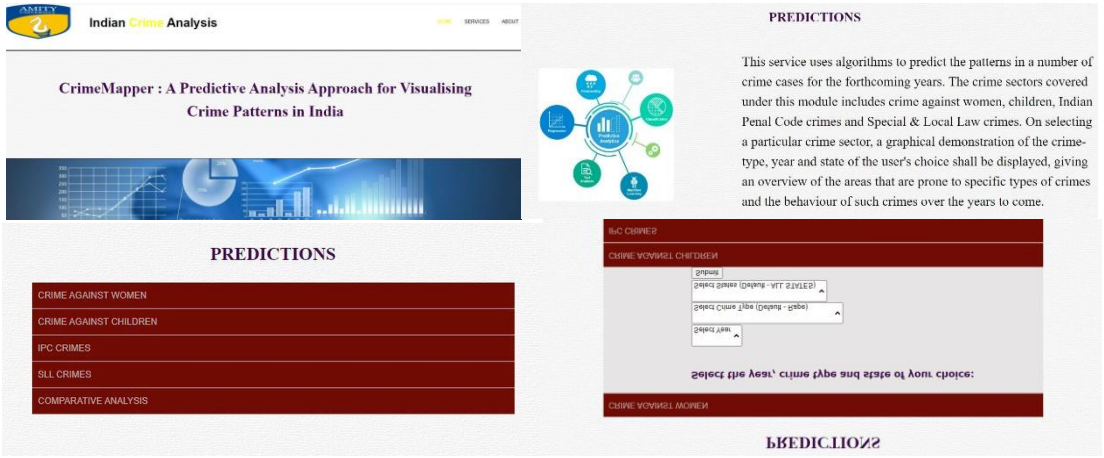


**Figure 5: Look and Feel of the web API**

b. Data Prediction Graphs: For any and all predictions exceeding 65% accuracy, Prediction graphs were generated. However, for predictions lacking required accuracy, a message of low accuracy constraint was flashed on the screen.

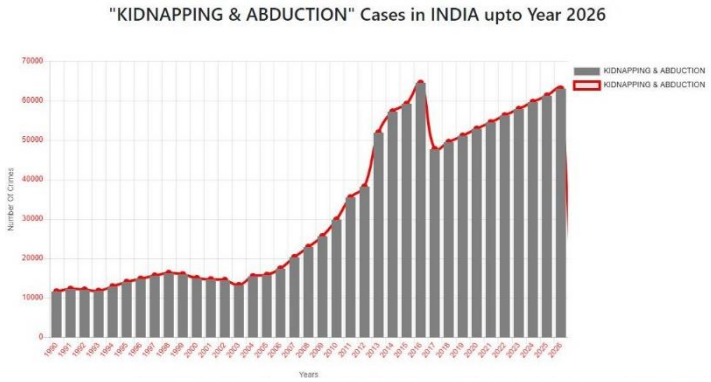


**Figure 6: Prediction graph with accuracy>65%**

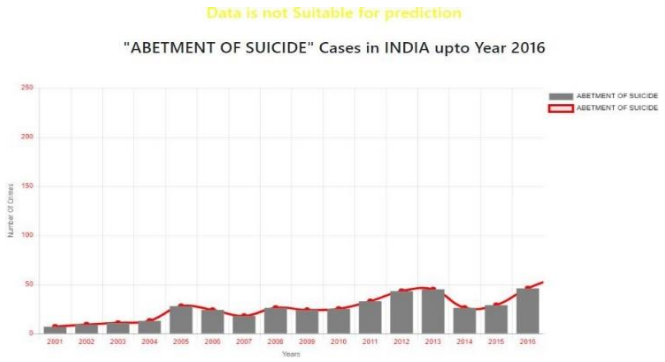


**Figure 6: Prediction graph with accuracy<65%**

c. State-wise comparison of graphs: Different graphs plotted for different states with the same crime were plotted against each other to check the difference in crime patterns of individual states.

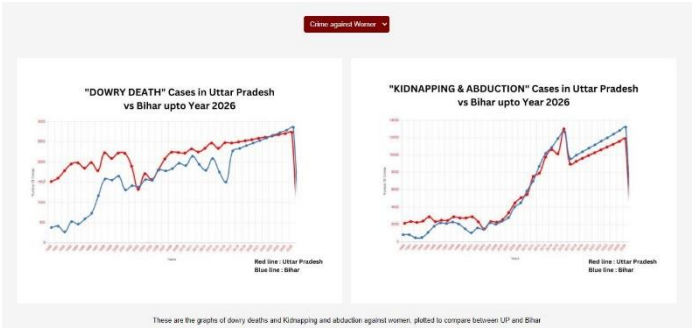


**Figure 7: Comparative graphs for the same crime**

## V. CONCLUSION AND FUTURE SCOPE

Conclusion: The proposed study is a novel technique for understanding and visualizing crime patterns in India. It effectively turned raw data from official Indian government websites into an analysis-ready format by employing predictive analytics and advanced data mining techniques. Through simple visualization charts, users obtain useful insights into crime patterns, with crime type estimates reinforced by demographic characteristics such as literacy rates and region size. The website emphasizes user friendliness, ensuring that people with varied degrees of data literacy may readily explore and profit from the analytical tools provided. Overall, the proposed model is an asset in the fight against crime due to its innovative use of predictive analytics, data mining techniques, and user-

friendly visualizations. This project paves the way for informed decision-making, enhanced safety measures, and a collaborative effort towards a more secure and prosperous India.

Future Scope: Even though the proposed model provides extensive use case scenario and optimum results, future work can increase its capabilities ten folds. Some such options are as follows:

a. Real-time Data Integration: The project currently uses data that dates from 2001 to 2016. Real-time information on crime patterns could be obtained by incorporating data from official sources. Users would be able to access the most recent crime patterns and make timely decisions if mechanisms to integrate and process real-time data were developed.

b. Geographic Visualization: Users could investigate crime patterns on a spatial level by enhancing the visualization capabilities to include geographic mapping. Users would be able to view crime hotspots, identify high-risk areas, and analyze the spatial distribution of crime incidents if geographic information systems (GIS) were integrated into the platform.[14]

c. Enhanced Predictive Models: The accuracy and dependability of crime type estimates would be improved by continuously enhancing the prediction algorithms used in the project. Investigating more sophisticated machine learning methods, like ensemble models or deep learning algorithms, may enhance the predictive abilities and offer more detailed perceptions of crime patterns.

d. Comparative Analysis: Introducing features that enable users to compare crime patterns across different time periods, states, or demographic factors would enhance the analytical capabilities of the platform. Allowing users to select specific regions or timeframes and generate comparative visualizations would facilitate a deeper understanding of crime trends and their variations.

e. Crime Prevention Strategies: Beyond visualization and analysis, incorporating recommendations or suggestions for crime prevention strategies based on the identified patterns and trends could significantly contribute to the platform's practical utility. This could include sharing insights into effective policing methods, community engagement initiatives, or targeted interventions based on risk factors identified.

## REFERENCES


1. Tushar Sonawanev, Shirin Shaikh, Shaista Shaikh, Rahul Shinde, Asif Sayyad, “Crime Pattern Analysis, Visualization and Prediction using Data Mining”, IJARIIE International Journal of Advanced Research and Innovative Ideas in Education, Vol-1 Issue-4 2015, IJARIIE-ISSN(O)-2395-4396.
2. Malathi. A , Dr. S. Santhosh Baboo and Anbarasi. A, “Enhanced Algorithm to Predict a future crime using Data Mining”, International Journal of Computer Applications (0975 – 8887), Volume 21– No.1, May 2011.
3. Sarpreet Kaur1 and Dr. Williamjeet Singh, “Systematic Review of Crime Data Mining”, IJARCS International Journal of Advanced Research in Computer Science, Volume 8, No. 5, May-June 2017, ISSN No. 0976-5697.
4. Neeru Mago, “Design and Implementation of Intelligent Crime Information and Analysis System (ICIAS) based on Crime Data Mining”, IJARCSSE International Journal of Advanced Research in Computer Science and Software Engineering, Volume 6, Issue 1, January 2016 ISSN: 2277 128X.
5. Ahamed Shafeeq B M, Dr. Binu V S, “Spatial Patterns of Crimes in India using Data Mining Techniques”, Certified International Journal of Engineering and Innovative Technology (IJEIT) Volume 3, Issue 11, ISSN: 2277-3754 ISO 9001:2008, May 2014.
6. D.Gnana Rajesh and Dr. M. Punithavalli, “A Study on the Existing Computer Aided Crime Analysis Tools to Handle Indian State of Affairs”, IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 2, No 3, March 2012 ISSN (Online): 1694-0814.
7. S.Yamuna, N.Sudha Bhuvaneswari , “Data mining Techniques to Analyse and Predict Crimes”, The International Journal of Engineering And Science (IJES), Volume 1, Issue 2, Pages 243-247 2012, ISSN: 2319 – 1813, ISBN: 2319 – 1805.
8. P.DHAKSHINAMOORTHY, T.KALAISELVAN, “CRIME PATTERN DETECTION USING DATA MINING”, International Journal of Advanced Research in Computer Science and Applications Vol.1, Issue 1, July 2013, ISSN 2321 -872X.
9. Mufeeda M. (2021). “Crime Data Analysis, Visualization and Prediction” international journal for research in applied science and engineering technology, 9(3), 71–78.
10. Hajela Gaurav, Chawla Meenu, & Rasool Akhtar. (2021). “A multi-dimensional crime spatial pattern analysis and prediction model based on classification” Etri Journal, 43(2), 272–287.
11. Agarwal, Vivek. (2015). Research on Data Preprocessing and Categorization Technique for Smartphone Review Analysis. International Journal of Computer Applications. 131. 30-36. 10.5120/ijca2015907309.
12. Komorowski, Matthieu & Marshall, Dominic & Salciccioli, Justin & Crutain, Yves. (2016). Exploratory Data Analysis. 10.1007/978-3-319-43742-2_15.
13. Relan, Kunal. (2019). Building REST APIs with Flask: Create Python Web Services with MySQL. 10.1007/978-1-4842-5022-8.
14. Wang, Sizhe & Li, Wenwen & Gu, Zhining. (2023). GeoGraphViz: Geographically Constrained 3D Force-Directed Graph for Knowledge Graph Visualization.