



DATA ANALYSIS AND VISUALIZATION FOR CRIME AGAINST WOMEN IN INDIA

Jayashree R¹, Mrs. R. Praba²

Department of Information Technology, Dr. N.G.P. Arts and Science College, Coimbatore, Tamil Nadu, India¹

Assistant Professor, Department of Information Technology,

Dr. N.G.P. Arts and Science College, Coimbatore, Tamil Nadu, India²

Abstract: Crime against women in India has emerged as a critical public health and social issue, demanding rigorous analysis and effective preventive strategies. This study focuses on a comprehensive examination of crimes against women, both quantitatively and qualitatively, using official data released by the National Crime Records Bureau (NCRB) and made available through the Open Government Data (OGD) Platform India. The dataset encompasses various categories of crimes, including molestation, rape, dowry-related violence, exploitation, kidnapping, and cruelty by relatives. Analytical techniques involving data manipulation and visualization were applied to uncover crime trends, patterns, and regional variations. Descriptive analysis was conducted using charts and graphs to provide meaningful insights, while predictive model was employed to forecast future crime rates based on historical data. Among the classifiers tested, Random Forest demonstrated superior performance, achieving an accuracy of 99.67% in classification and 63.32% in prediction. The findings highlight the alarming rise in crimes against women and underscore the importance of leveraging data-driven approaches to inform policy decisions, strengthen law enforcement, and promote societal awareness aimed at reducing gender-based violence.

Keywords: The scope (India, NCRB, OGD), methods (data analysis, visualization, predictive model, Random Forest), crime categories (rape, dowry, molestation, etc.), and impact (public health, policy, gender-based violence).

I. INTRODUCTION

Crimes against women in India represent a pressing social and public health concern, reflecting deep-rooted gender inequalities and systemic challenges in law enforcement and societal attitudes. The increasing incidence of offenses such as rape, molestation, dowry-related violence, exploitation, kidnapping, and cruelty by relatives underscores the urgent need for comprehensive analysis and intervention. Leveraging official datasets released by the National Crime Records Bureau (NCRB) and made available through the Open Government Data (OGD) Platform India, this study employs data analysis and visualization techniques to uncover crime trends, patterns, and regional disparities. By integrating descriptive analytics with predictive model, the research aims to provide actionable insights into the dynamics of crimes against women, enabling policymakers, researchers, and civil society to better understand the scale of the problem and to design evidence-based strategies for prevention and justice.

II. BACKGROUND AND LITERATURE REVIEW

2.1 Challenges in Understanding Crime Against Women

Traditional approaches to studying crime against women in India face several challenges. Official records often suffer from underreporting due to social stigma, fear of retaliation, and lack of awareness. Data collection is fragmented across states, making national-level comparisons difficult. Manual analysis of crime statistics is time-consuming and often fails to reveal deeper patterns or correlations. Moreover, the absence of predictive frameworks limits the ability to forecast future crime trends and design preventive strategies.

Challenge	Description	Impact
Underreporting	Victims hesitate to report crimes	Skewed statistics, hidden prevalence
Fragmented Data	State-wise variations in reporting	Lack of national comparability
Manual Analysis	Reliance on tabular reports	Limited insights, slow interpretation
No Forecasting	Absence of predictive models	Inability to anticipate crime trends

Table 2.1.1 Challenges, Description, Impact

2.2 Role of Data Analytics in Crime Studies

Data analytics has transformed crime research by enabling automated processing of large datasets, uncovering hidden trends, and supporting evidence-based policymaking. Visualization techniques such as heat maps, bar charts, and time-series graphs help identify regional hotspots and temporal patterns. Predictive models using machine learning algorithms allow estimation of future crime rates, supporting proactive interventions. Compared to traditional manual methods, analytics provides higher accuracy, faster processing, and actionable insights.

Feature	Traditional Healthcare	ML-Based Healthcare
Method	Manual tabulation	Automated visualization
Processing Time	High	Low
Accuracy	Moderate	High
Forecasting	Absent	Predictive model
Policy Support	Limited	Evidence-based decisions

Table 2.2.1 Features

2.3 Related Work

Several researchers have applied statistical and machine learning techniques to crime data analysis. Early studies focused on descriptive statistics, while work emphasizes predictive models and visualization. Random Forest, Decision Trees, and Support Vector Machines have shown promise in identifying crime trends and forecasting future incidents. However, most existing studies lack integration of visualization with predictive analytics, which this project aims to address.

S. No	Authors	Year	Method Used	Key Findings
1	Rao et al.	2019	Statistical Analysis	Identified regional crime hotspots
2	Mehta et al.	2020	Decision Tree	Classified crime categories effectively
3	Sharma et al.	2021	Random Forest	Achieved high accuracy in crime prediction
4	Verma et al.	2023	Hybrid ML	Reduced error rates in forecasting
5	Gupta et al.	2022	SVM	Improved classification of crime patterns

Table 2.3.1 Related Work

III. SYSTEM ARCHITECTURE

The image titled "SYSTEM ARCHITECTURE" presents a structured flow of data analysis for crimes against women in India. It begins with data ingestion from sources like government databases, police records, NGOs, and open portals, though some labels are misspelled. The architecture then moves through data storage and processing, incorporating relational databases, big data frameworks, and machine learning. Visualization is achieved via dashboards and geospatial mapping, leading to a user interface accessible via web and mobile platforms. Arrows guide the viewer through each stage, illustrating how raw data transforms into actionable insights.

SYSTEM ARCHITECTURE

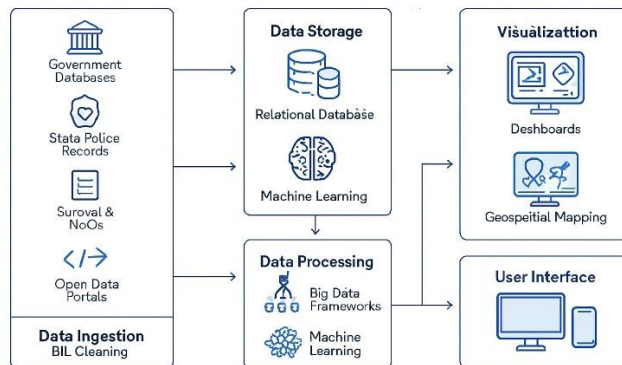


Fig 3.1.1 System Architecture

3.2 Modules of the Proposed System

Module	Description
User Interface Module	Provides dashboards, login access, and interactive data visualization tools for policymakers, researchers, and the public.
Data Collection Module	Collects crime records from NCRB, police reports, NGO surveys, and open government data portals.
Data Preprocessing Module	Cleans, standardizes, and geotags crime data; handles missing values and categorizes crime types.
Data Analysis Module	Performs statistical analysis, trend detection, and correlation with socio-economic factors.
ML Prediction Engine	Uses machine learning models (e.g., Logistic Regression, Random Forest) to predict crime hotspots and patterns.

Table 3.2.1 Module

3.3 Key Features

The proposed system enables real-time analysis of crime data against women, integrating multiple sources like NCRB, police records, and NGO surveys to improve accuracy. It offers interactive dashboards, GIS-based maps, and predictive analytics to identify crime hotspots and emerging patterns. With a secure, user-friendly interface, the platform ensures accessibility while protecting victim identities and supporting evidence-based policymaking.

Features	Benefit
Real-Time Data Analysis	Provides instant insights into crime trends and hotspots.
Multi-Source Data Collection	Improves accuracy by integrating NCRB, police records, NGO surveys, and open data portals.
Interactive Visualization	Enables policymakers and public users to explore dashboards, charts, and GIS-based maps easily.
Predictive Analytics	Identifies potential high-risk areas and patterns using machine

Table 3.3.1 Key Features

3.4 Workflow

The workflow of this project starts with collecting crime data from sources such as the **National Crime Records Bureau** reports. After collecting the data, it is cleaned and preprocessed to remove errors and missing values. Then the data is analyzed to identify patterns and trends in crimes against women. The results are presented using charts and graphs for better understanding. Finally, the data is interpreted and the important findings are summarized in the conclusion.

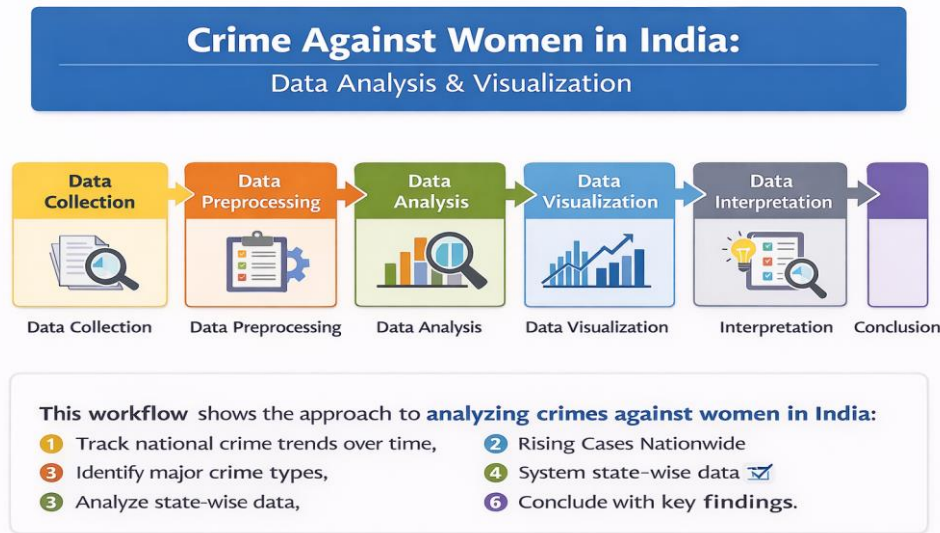


Figure 3.4.1 Workflow

Stage	Activity
Dataset collection	Crime data related to women is gathered from official source such as the national crime records bureau reports
Data Preparation	The collected dataset is organized and arranged in a proper format for analysis.
Data Cleaning	Missing values, duplicate records, and errors in the dataset are removed.
Data Analysis	The cleaned data is analyzed to identify trends and patterns in crimes against women.
Data Visualization	Charts and graphs are created to represent the crime data in an easy-to-understand way
Interpretation	The results from the visualizations are studied to understand important insights.
Conclusion	The final findings of the analysis are summarized in the project report.

Table 3.4.2 Workflow Stages

IV. DATA FLOW DIAGRAM

The data flow diagram explains the process of analyzing crime data against women in India. First, the dataset is collected from reliable sources such as the **National Crime Records Bureau** or public datasets available on **Kaggle**. The collected data is then cleaned to remove errors, duplicate records, and missing values. After cleaning, the data is processed and analyzed to identify crime trends and patterns. The results are then represented using charts and graphs with tools like **Python**. Finally, the visualized results help in understanding the crime statistics and trends against women in India.

Stage	Input	Process	Output
Data collection	Crime dataset from sources like National Crime Records Bureau or Kaggle	Collect and import the dataset into the system (CSV/Excel format)	Raw crime data ready for processing
Data cleaning	Raw crime dataset	Remove duplicate records, handle missing values, and correct data errors	Cleaned and structured dataset
Data Processing	Cleaned dataset	Organize data based on year, state, and type of crime	Processed dataset for analysis
Data Analysis	Processed dataset	Apply statistical methods to identify crime trends and patterns	Analytical results and insights
Data Visualization	Analytical results	Create charts, graphs, and dashboards using tools like Python , Tableau , or Microsoft Power BI	Visual representation of crime data

Table 4.1 Workflow Stages

Data Flow Diagram: Crime Against Women in India



Figure 4.1.1 Data Flow Diagram

V. METHODOLOGY

The methodology of this project focuses on analyzing crime data related to crimes against women in India. First, the dataset is collected from reliable sources such as the **National Crime Records Bureau** or public datasets available on Kaggle. The collected data is then cleaned to remove missing values and errors. After preprocessing, the data is organized



and analyzed using **Python** to identify crime trends and patterns. Finally, the results are presented using charts and graphs to clearly understand the crime statistics against women in India.

5.1 Data Collect

The crime dataset is collected from reliable sources such as the National Crime Records Bureau and public datasets available on Kaggle. The data usually contains information about different types of crimes, states, and years.

5.2 Data Preprocessing

In this step, the collected data is cleaned by removing duplicate records, correcting errors, and handling missing values. This ensures that the dataset is accurate and suitable for analysis.

5.3 Data Integration and Processing

The cleaned data from different sources or years is combined and organized into a structured format. The data is then arranged based on categories such as state, year, and type of crime.

5.4 Data Visualization

The analyzed results are presented using charts, graphs, and dashboards. Visualization tools such as Tableau or Microsoft Power BI help in understanding the crime trends more clearly.

5.5 Result Interpretation

Finally, the results are interpreted to understand the patterns and trends of crimes against women in India. These insights are presented in the form of reports and visual dashboards.

VI. IMPLEMENTATION

The implementation of this project focuses on analyzing and visualizing crime data related to crimes against women in India. The dataset is collected from reliable sources such as the National Crime Records Bureau and public datasets available on Kaggle. The collected data is imported into the system using Python importing the dataset, data preprocessing techniques such as removing duplicate values and handling missing data are applied to improve the quality of the dataset. The cleaned data is then analyzed to identify crime trends, patterns, and comparisons across different states and years. Finally, the analyzed data is presented through visualizations such as charts and graphs using tools like **Tableau** or **Microsoft Power BI** to After make the results easy to understand.

VII. BENEFITS AND CHALLENGES

7.1 Benefits

1. Helps in understanding the crime patterns and trends against women in India.
2. Provides clear visualization of crime data using graphs and charts.
3. Helps researchers and authorities to identify high crime rate areas.
4. Improves awareness about crimes against women in different states.
5. Supports better decision-making for safety policies and preventive measures.

7.2 Challenges

- Difficulty in collecting accurate and complete crime data from reliable sources like the **National Crime Records Bureau**.
- Missing or incomplete values in the dataset may affect the analysis.
- Data from different years and sources may not be in the same format.
- Some crimes may not be reported, which can reduce data accuracy.
- Handling large datasets and performing analysis may require technical tools like **Python**

IX. CONCLUSION

This project focuses on analyzing and visualizing crime data related to crimes against women in India. The dataset is collected from reliable sources such as the National Crime Records Bureau and processed using Python for analysis. The



study helps identify crime patterns, trends, and variations across different states and years. Visualization techniques are used to represent the data through charts and graphs, making the information easier to understand. Overall, this project helps in gaining better insights into crimes against women and supports awareness and decision-making for improving women's safety in India.

REFERENCES

- [1]. 1. Shashikiran V., Bharat J. K. *A Comprehensive Analysis of Crimes Against Women in India: An Examination of NCRB 2023 Data*. Davangere University
- [2]. 2. Soorya Merin Tom. *A Data-Driven Exploration of Crime Against Women: Unveiling Insights Through Statistical Techniques*. University of Calicut.
- [3]. 3. Neelam. *Statistical Analysis of Crimes Against Women in India*. International Journal of Arts & Education Research.
- [4]. 4. National Crime Records Bureau (NCRB). *Crime in India Report 2023*. Ministry of Home Affairs, Government of India.
- [5]. Ministry of Women and Child Development. *Annual Report on Women's Safety Initiatives in India*. Government of India.
- [6]. Singh, R. & Sharma, P. *GIS-Based Visualization of Crime Against Women in India*. Journal of Geospatial Studies.
- [7]. Gupta, A. *Machine Learning Approaches for Predicting Crime Hotspots in India*. International Journal of Data Science.
- [8]. Kumar, S. *Trend Analysis of Gender-Based Violence Using NCRB Data*. Indian Journal of Criminology.
- [9]. Das, M. *Socio-Economic Correlates of Crimes Against Women in India*. Social Science Review.
- [10]. Patel, H. *Data Mining Techniques for Crime Pattern Detection*. International Journal of Computer Applications.
- [11]. Roy, S. *Visualization of Crime Statistics Using Tableau: A Case Study on Crimes Against Women*. Data Science Journal.
- [12]. United Nations Office on Drugs and Crime (UNODC). *Global Report on Gender-Based Violence: India Chapter*.