

Various Crimes Prediction Against Women

Trishika Tiwari
BTech Computer Science
Dr. Akhilesh Das Gupta Institute
of Professional Studies
Delhi, India
trishikatiwari33@gmail.com

Abstract—Crime against women is a pressing social issue that demands a comprehensive understanding and effective solutions. This research paper presents an in-depth analysis and prediction model of crimes targeting women, leveraging data-driven methodologies. By utilizing a wide range of demographic, socio-economic, and crime related variables, this study aims to shed light on the underlying factors contributing to such crimes and develop a predictive framework to anticipate potential occurrences.

The research begins by gathering and curating an extensive dataset comprising historical crime records and contextual information related to women's safety. Various statistical and machine learning techniques are employed to uncover hidden patterns and correlations within the data. The analysis explores spatial and temporal trends, identifying hotspots and high-risk areas for crimes against women.

Keywords—*Crime Analysis, Flask Architecture, Linear Regression, Random Forest Classifier.*

INTRODUCTION

In India, according to the National Commission for Women (NCW), 23,722 complaints of crimes against women were received in 2020, the highest in the last six years. It is a big threat to humanity and in some parts of India, women are still treated as prisoners. Violence against women is perhaps as old as mankind. Not only in rural areas of India but also in urban areas women face a lot of problems like dowry, trafficking, acid attacks, miscarriage, kidnapping, and abduction of women. Crime against women is majorly happening because of the inefficient legal justice system, weak rules of law, crime prediction, and criminal identification. There is a need to analyse and predict the crime rate and the necessary steps to be taken further by the government officials to avoid the increase in threats and crime against women. Women are adored in India. They are granted the status of being good. But the reality is entirely different. The safety of women has become increasingly problematic as time goes on.

The frequency of violations is rapidly increasing over time. It is now recognized as a global problem, and many countries are making adjustments to coincide with declining crime rates. There are numerous forms of violence against women, and it can happen everywhere, including at home, in public, and at work.

PROBLEM STATEMENT

The purpose of this study is to forecast a variety of crimes against women, evaluate these crimes in the years to come, and comprehend the efficacy of different predictive machine learning algorithms. Based on this knowledge, the government might assume responsibility and work to reduce crime.

METHODOLOGY

This section would demonstrate the methodology and the principle in an elaborating way. The objective of the paper is to analyze and predict the crimes against women in India. Visualization of analysis and prediction of crime against women in India with the data gathered from National Criminal Records Bureau(NCRB) which was used in order to predict the rate of crime against women and its severity in the forthcoming years, in different areas of the country based on the previous year criminal records of India. Data Preprocessing plays a vital role before implementing the machine learning algorithms. From the Dataset acquired from the government portal there were undefined and unnamed features which needed to be dropped as they were not related and had no meaning to adding these features to the algorithms. The system uses linear regression Algorithm in major section i.e. Predicting the Crime Patterns with minimum accuracy rates as a factor. Linear Regression is the suitable Algorithm to predict the pattern of this data compared to other algorithms used for testing purposes such as Random Forest Algorithm, and found linear regression as the best fit for this proposed system with maximum accuracy in the results. The system is scalable as the dataset increases, the variation in the visualization of analysis and the accuracy of the prediction increases. As the increase in the system's

workload, the system would be able to process and so the system is scalable to an apex.

DataCollection: Collect allavailable datasets on crimes against women, including incident reports, demographic information, geographic locations, and time stamps. Police reports, public documents, and databases are among possible sources.

Data Preprocessing: Remove errors, outliers, and missing values from the acquired data. Make sure the data is compatible with machine learning algorithms by normalizing and standardizing it. Develop useful predictors by extracting features and engineering them.

Conducting EDA (Exploratory Data Analysis): Apply EDA to comprehend the data distribution, spot important patterns, and see trends graphically. This stage is useful for choosing the right characteristics.

Algorithm Selection and Implementation: For model creation, choose machine learning algorithms like Support Vector Machines (SVM), Decision Trees, Naïve Bayes, K-Nearest Neighbors (KNN), Linear Regression, CART, and Naïve Bayes. Put these algorithms into action by selecting an appropriate programming language and ML framework.

Model Training and Validation: Separate the dataset and test subsets. Utilize techniques such as k-fold cross-validation to train the models and verify their evaluate performance. Adjust the model's parameters to your liking and determine its accuracy, precision, F1 score, and recall.

Analysis and Prediction: Make use of the trained models to pinpoint areas prone to crime and pinpoint populations who are particularly at risk. Use the results of the analysis to inform decisions about allocating resources and preventing crime.

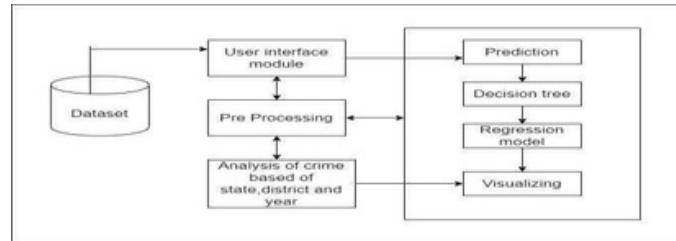
EXISTING SYSTEM

The general public is receiving assistance from current specialised advancements in sophisticated measurement analysis and portrayal in a variety of ways to evaluate the information with social importance. Data about mis behaviour in many component areas is one of these socially significant workouts. The evaluation of the wrongdoing facts will help to improve mental cycle relationships and demand responsible actions to change the rate over segment areas. Advancements in the field of determining age, along with publicly available records and donations, in some manner make it easier for criminals to acquire their wrongdoings and include them in larger, more egregious crimes than before. As a result, in both wealthy and less developed countries, prices are rising at an extremely unjustified rate. upheld the minor misconduct from the previous year in the Indian states.

WORKFLOW OF THE SYSTEM

In recent times, crime and violence against women have become a blatant topic of discussion in India. The steady increase in crime rates has caused our administration and the media to particularly concentrate on this subject. Various Indian provinces are the scene of numerous wrongdoings. This project illustrates how AI can be used to identify infractions committed by various states and offer them with a clear justification. This supports the objective investigation of material pertaining to misbehaviour concerns involving women and may

serve as a persuasive manual for the public authority in formulating strategies for preventive measures



TOOLS AND TECHNOLOGIES USED

3.1 Python Implementation: Scripts and Programs

Machine learning analysis and prediction of crimes against women may be accomplished well using Python.

When it comes to automating activities and executing sophisticated algorithms without human input, scripts are crucial in Python. For effective model testing and improvement, these scripts may be run repeatedly and include Python code. Improving crime prediction models relies heavily on this iterative method.

3.2 Key Python Libraries for Crime Prediction

NumPy

Pandas

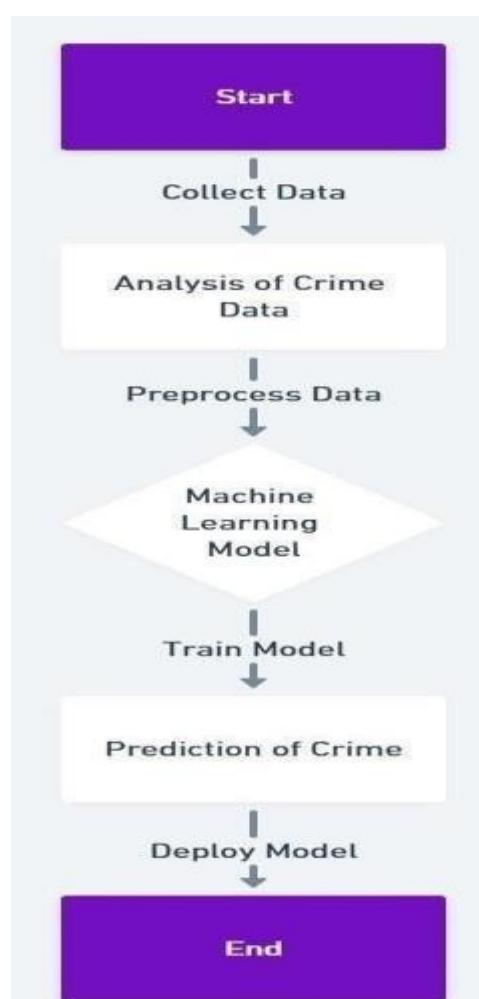
Scikit-Learn

Tensor Flow and Keas

Matplotlib and Seaborn

Stats models.

Sequence diagram



IMPLEMENTATION

4.1 Development Environment

This criminal prediction system's development environment comprises: [1] Programming Languages: Python's strong support for data analysis and machine learning packages led to its selection as the main programming language. [2] Software Development Environments (IDEs): PyCharm and Visual Studio Code are among the IDEs used by the development team. Git is used for version control, while GitHub is used to host repositories. The team communicates and keeps tabs on projects using collaboration platforms like Slack and JIRA.

4.2 System Components

Key components of the system are implemented as follows: [1] Module for Preprocessing Data: In this section, we clean, normalize, and extract features from the data. Prior to SVM modeling, data is prepared by implementing algorithms that deal with missing values, scale numerical features, and encode categorical variables. [2] Machine Learning Model Module: SVM algorithms form the backbone of the system. To categorize and forecast crime trends, the SVM model is built and trained with the use of crime data from the past. To get the best possible results, we validate our model and fine-tune its hyperparameters. [3] User Interface: Analysts and law enforcement will find the user interface easy to utilize. It provides a user-friendly interface with reports and visualizations that let users enter crime-related data and examine prediction results.

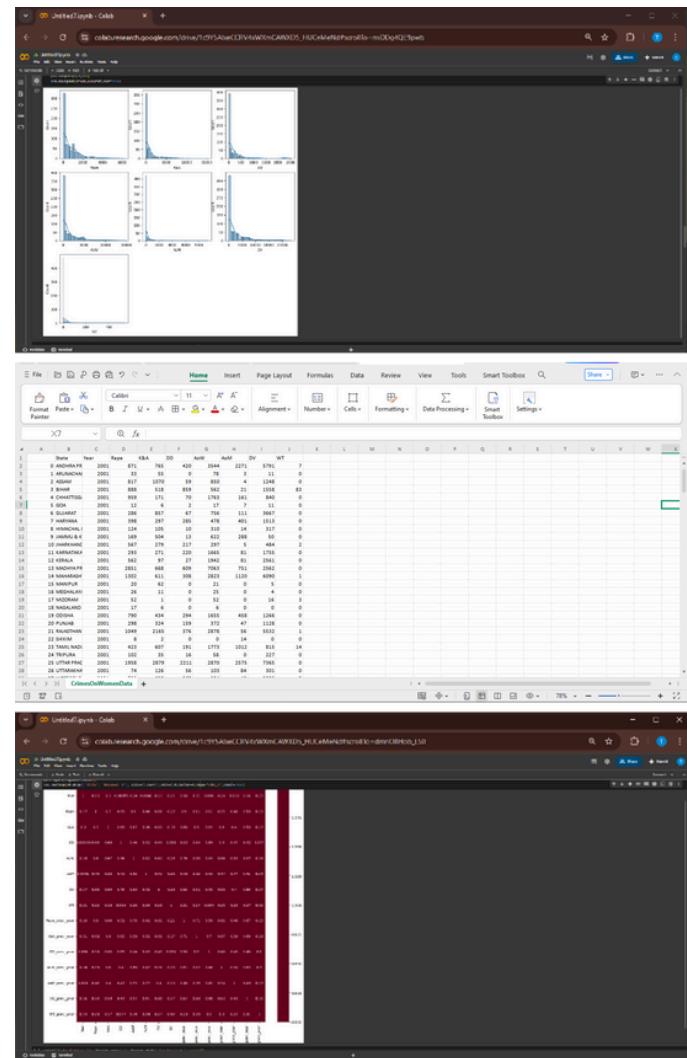
4.3 Integration

Integration efforts include: [1] API Integration: Securely transmit data to sources like social media platforms and public criminal databases via the use of APIs. The secure and uninterrupted transfer of data into the system is guaranteed by these APIs. [2] Database Integration: In order to store preprocessed data, trained SVM models, and prediction outputs, the system integrates with databases such as SQL or NoSQL. Efficient data management and retrieval are made possible by this connection.

4.4 Deployment

Deployment details include: [1] Environment of Deployment: Based on the organization's requirements and infrastructure, it may be implemented either on-premises or in the cloud, such as AWS or Azure. [2] Efficient Configuration Management: All system settings, dependencies, and environment variables are carefully monitored to guarantee consistent deployment in different environments. This promotes seamless operation and scalability.

SCREENSHOT



ACKNOWLEDGMENT

We would like to thank our guide **Dr Ritu** for their valuable support and guidance throughout this research on predicting crimes against women using machine learning and Python. Their suggestions and encouragement helped me a lot in completing this work. We are also thankful to my friends and family for their constant motivation and support during this journey. Finally, we sincerely appreciate the open-source community for providing the tools and resources that made this research possible.

REFERENCES

- [1] National Crime Records Bureau (NCRB). *Crime in India 2022* (Book 1). Government of India. ncrb.gov.in
- [2] UN Women. *Report of the Secretary-General: Intensifying efforts to end violence against women and girls* (A/79/500), 8 Oct 2024. [UN Women](https://www.un.org/womenwatch/daw/crimeagainstwomen/)
- [3] Kaggle. *Crimes Against Women in India (2001–2021)* (state-wise annual counts). [Kaggle](https://www.kaggle.com/datasets/akashdeep99/crimes-against-women-in-india-state-wise-annual-counts)
- [4] Kaggle. *Crime Against Women 2001–2014 (India)* (state & district level).