IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, April 2025
Impact Factor: 7.67

Crime Data Analysis Against Women and Girls in India

Dheetchanya TB[1] and Dr. S. Chithra Devi[2]
UG Student, Department of Computer Science with Data Analytics[1]
Assistant Professor, Department of Digital and Cyber Forensic Science[2]
Sri Ramakrishna College of Arts & Science, Coimbatore, Tamil Nadu, India

Abstract: *Crime is a critical issue that poses a serious threat to public safety and national development. As crime rates increase across various regions, the need for intelligent systems to detect patterns and predict potential threats becomes essential. This project focuses on the analysis and prediction of crime using Python and machine learning algorithms. The dataset provides comprehensive crime records across different states, genders, and time periods. The project involves multiple stages, including data preprocessing, exploratory data analysis (EDA), visualization, and the application of supervised machine learning techniques. Random Forest and Support Vector Machine (SVM) algorithms are implemented to classify and predict crime occurrences based on selected features. In addition, Python libraries such as Pandas, Seaborn, Matplotlib, and Scikit-learn are utilized for data handling, statistical analysis, and graphical representation. The results provide deep insights into crime trends and help identify hotspots and recurring patterns. The machine learning models enhance the project's practical utility by enabling crime prediction with a high degree of accuracy, thus supporting law enforcement agencies in proactive crime prevention and decision-making.*

Keywords: Crime Data Analysis, Machine Learning, Python, Random Forest, Support Vector Machine (SVM)

I. INTRODUCTION

1.Statement of the problem

Crime rates in India have been steadily increasing, posing a serious challenge for law enforcement agencies to maintain public safety and security. Despite the availability of vast crime data, deriving meaningful insights and patterns remains difficult due to the complexity and volume of the information. Traditional manual methods are time-consuming and inefficient in identifying trends or predicting future crime occurrences. There is a pressing need for an intelligent, automated system that can analyze historical crime data, uncover hidden patterns, and predict potential crime incidents. This project addresses the problem by developing a data-driven analytical model using Python and machine learning algorithms such as Random Forest and SVM, aiming to assist authorities in making informed decisions for crime prevention and resource allocation.

2. Goals

The primary goal of this project is to analyze and interpret the Indian crime dataset using Python and machine learning techniques. It aims to uncover hidden patterns, trends, and correlations in crime data across various states, years, and gender classifications. By applying data visualization tools such as Matplotlib and Seaborn, the project provides a clear and intuitive understanding of crime distribution. Furthermore, the project seeks to develop and implement predictive models using algorithms like Random Forest and Support Vector Machine (SVM) to classify and forecast crime categories. Another key objective is to evaluate the accuracy and efficiency of these models, ultimately helping law enforcement agencies make informed, data-driven decisions for crime prevention and policy planning.

DOI: 10.48175/IJARSCT-24987

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, April 2025
Impact Factor: 7.67

II. SOFTWARE SPECIFICATIONS

The Crime Data Analysis project was developed using Python as the core programming language due to its simplicity, flexibility, and robust ecosystem of data science libraries. Data manipulation and preprocessing were performed using **Pandas**, while **Matplotlib** and **Seaborn**were utilized for generating insightful visualizations. For building and evaluating predictive models, the **Scikit-learn** library was employed, specifically implementing algorithms like **Random Forest** and **Support Vector Machine (SVM)**. The development and execution of the code were carried out in environments such as **Jupyter Notebook** offering an interactive interface and ease of collaboration.The project was compatible with both **Windows** and **Linux** operating systems.

III. DESIGN & FLOWCHART

1.Database Design

A well structured database is essential for efficient storage, retrieval, and analysis of crime data. It consists of tables storing crime type, location, time, victim demographics, and case outcomes. Relationships between tables facilitate seamless querying, while indexing ensures quick access to key data points like crime type and location.

2. Input Design

The input design focuses on gathering and processing crime data from government reports and publicly available datasets, such as those sourced from Kaggle. The raw data undergoes a comprehensive preprocessing stage, which includes handling missing values, removing duplicates, normalizing inconsistent entries, and converting categorical variables into numerical formats using encoding techniques. Feature selection is applied to retain only relevant attributes that contribute to meaningful analysis and model accuracy. Additionally, the data is split into training and testing sets to facilitate machine learning implementation. A user-friendly interface or automated scripts built with Python ensures seamless and efficient data integration, allowing smooth transition from raw inputs to structured data ready for analysis and model training.

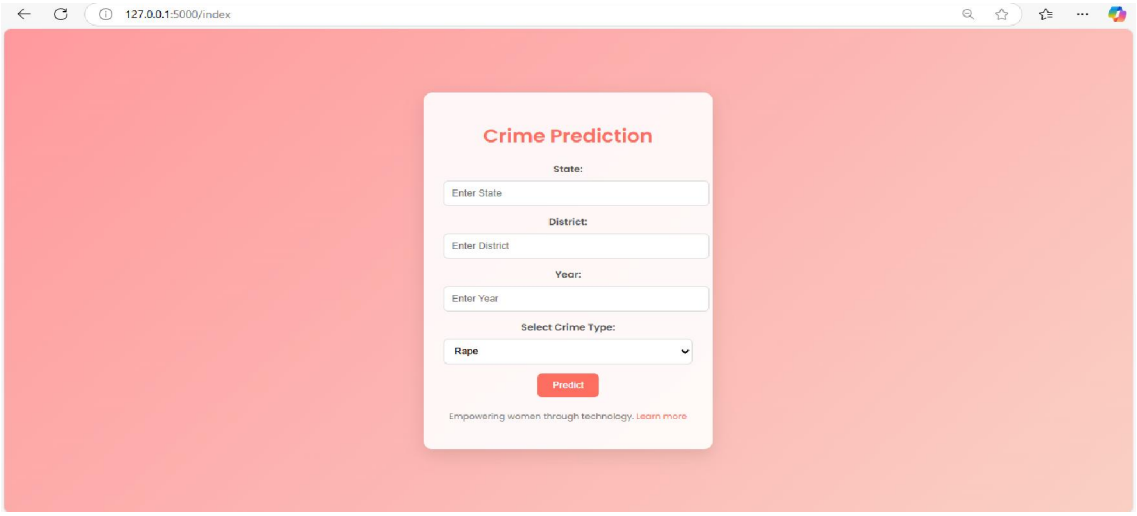

Fig 1: Input Features

3. Output Design

The output design is implemented through an intuitive and interactive web interface that displays crime prediction results based on user-provided input such as state, year, district, and crime type. Once the input is submitted, the system utilizes machine learning models — specifically Support Vector Machine (SVM) and Random Forest — to generate predictions for the likelihood or rate of the selected crime. The results are presented both numerically and graphically, with a bar chart comparing the predictions from both models side-by-side. This visual representation enables easy

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, April 2025
Impact Factor: 7.67

comparison and better understanding of the outcomes. For instance, when analyzing crime in Coimbatore, Tamil Nadu, for the year 2022, the SVM prediction was 24.59 and the Random Forest prediction was 20.62, indicating a relatively low crime rate. Additionally, the interface includes a contextual insight section that interprets the results, providing users with meaningful conclusions, such as safety alerts or preventive suggestions. The platform is further enhanced with user navigation options like "Show SVM Prediction" and "Back to Home" buttons for seamless interaction. Overall, the output design ensures that predictive results are clear, actionable, and supportive of informed decision-making.

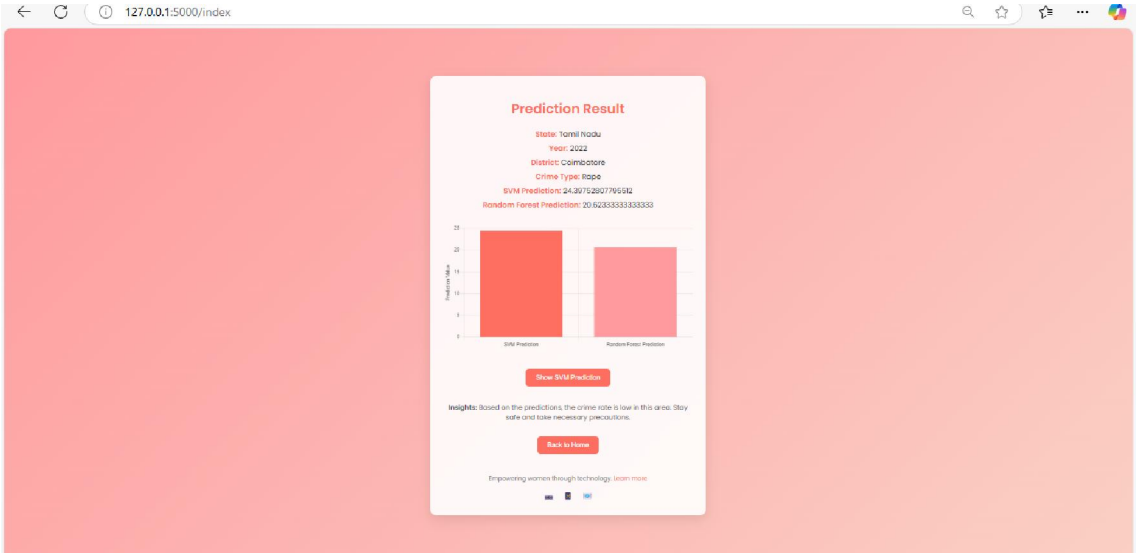

Fig 2: Output display

4. Design Process

The design process of the Crime Data Analysis system follows a client-server architecture powered by a Flask backend. It begins with the **user submitting a request** through the web interface (Step 1). This request is then **sent to the Flask server** (Step 2), which acts as the core processing unit. Upon receiving the request, the server **queries the dataset** (Step 3) and **retrieves the relevant crime data** (Step 4). Next, the server **forwards this data to the machine learning-based Crime Prediction Model** (Step 5), which processes it using algorithms like Random Forest and SVM to generate predictions. The model then **returns the prediction results to the server** (Step 6), which in turn **sends the analyzed results and visual insights back to the user interface**for display (Step 7). This modular and interactive flow ensures real-time processing and enables users to receive

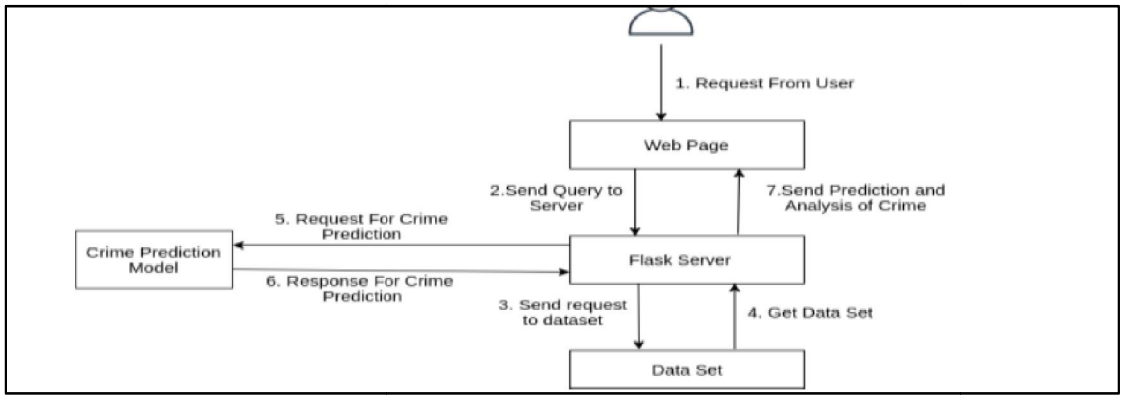

Fig 5: Design Process

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, April 2025
Impact Factor: 7.67

IV. RESULTS AND DISCUSSION

The Crime Data Analysis system effectively utilizes machine learning algorithms to predict and interpret crime trends based on historical data. After preprocessing and analyzing the dataset, two models — Support Vector Machine (SVM) and Random Forest — were implemented to classify and predict crime rates across different states, years, and crime categories. The models were trained and evaluated using real-world crime data sourced from Kaggle. The performance of both models was compared in terms of accuracy, interpretability, and consistency.

The results indicate that both models successfully predicted the crime rates, with Random Forest demonstrating slightly better performance in terms of accuracy and stability due to its ensemble nature. The SVM model, however, offered strong classification capability in cases where the data had clear margins. A sample output shows that for the district of Coimbatore, Tamil Nadu in 2022, the SVM predicted a crime rate of 24.59, while Random Forest predicted 20.62 for the crime type "Rape." This dual-model output allows users to interpret crime levels with more reliability.

Visualization tools such as bar charts and heat maps further supported the model's findings, offering intuitive representations of crime trends and high-risk zones. The system also includes an insightful web interface that enables users to interact with the model and instantly retrieve predictions based on selected parameters. These outputs can be highly valuable to law enforcement agencies, policymakers, and researchers, offering actionable insights for resource allocation and crime prevention strategies.

V. CONCLUSION

The Crime Data Analysis project demonstrates the potential of machine learning in understanding, predicting, and preventing criminal activities through data-driven approaches. By utilizing classification algorithms like Random Forest and Support Vector Machine (SVM), the system successfully analyzed crime trends and provided predictive insights based on real-time user input. The integration of a web-based interface with visualization tools made the analysis more interactive and user-friendly, enabling users to interpret complex crime data effortlessly. The results revealed meaningful patterns and regional variations in crime rates, contributing valuable support to law enforcement planning and public safety initiatives. Overall, the project not only emphasizes the importance of technology in crime analytics but also promotes its application for building safer communities through informed decision-making.

VI. ACKNOWLEDGMENT

The authors extend sincere gratitude to the Department of Digital and Cyber Forensic Science,Sri Ramakrishna College of Arts & Science, for providing the computational resources and infrastructure essential for this research. Special thanks to Dr. S. Chithra Devi, Department of Digital and Cyber Forensic Science, for her unwavering support and encouragement. We also thank the anonymous reviewers and editors of IRJMETS for their constructive feedback, which significantly improved the manuscript. Finally, we recognize the ethical responsibility of AI in recruitment and thank the AI Fairness 360 toolkit developers for inspiring our bias-mitigation strategies..

REFERENCES

[1].Scikit-learn: Machine Learning in Python – Pedregosa, F. et al., *Journal of Machine Learning Research,* 2011.
[2].Cristianini, N., &Shawe-Taylor, J. (2000*). An Introduction to Support Vector Machines and Other Kernel-based Learning Methods.* Cambridge University Press.
[3].Breiman, L. (2001). Random Forests*. Machine Learning*, 45(1), 5–32.
[4].A Study on Crime Rate Prediction using Machine Learning Techniques", IJARSCT, Volume 2, Issue 5, May 2022.
[5].India Crime Statistics – National Crime Records Bureau (NCRB). Ministry of Home Affairs, Government of India.