Global Journal of Engineering Innovations & Interdisciplinary Research

Original Article

Crime Prediction and Analysis Using Machine Learning

B Prasad Yadav[1], N Gayathri[2], M Srilekha[2], S Thanmayee[2]

[1]Assistant Professor, Department of Artificial Intelligence and Data Science, Vignan Institute of Technology and Science, Hyderabad, India

[2]UG Student, Department of AI&DS, Vignan Institute of Technology and Science, Hyderabad, India

Correspondence

B. Prasad Yadav

Assistant Professor, Department of Artificial Intelligence and Data Science, Vignan Institute of Technology and Science, Hyderabad, India

- Received Date: 25 May 2025
- Accepted Date: 15 June 2025
- Publication Date: 27 June 2025

Keywords

Crime Analysis, Time-Series Prediction, SARIMA, Prophet, Regional Crime Patterns, Data Visualization, Public Safety, Predictive Analytics, Machine Learning, Seasonal Trends, State-wise Insights.

Abstract

Crime prediction and analysis have been the backbone within which public safety is expected and effective policies can be made. Using time series forecasting techniques such as SARIMA and Prophet, this project describes crime forecasting in Andhra Pradesh, Telangana, and Maharashtra. The overall analysis will be directed towards producing actionable insights for police and policymakers by analysing historical crime data. After thorough data preparation involving cleaning and feature enrichment, Exploratory Data Analysis (EDA) will be conducted to identify important crime patterns and seasonal behaviours. The predictive model of SARIMA captures seasonal variations, while Prophet takes care of the irregularities and missing data with flexibility and accuracy. A comparative assessment of these two models would help to highlight their strength in forecasting crime trends. Crime patterns significantly differ with each region with specific traits attributed to socio-economic and geographical aspects. Some problems include data imbalance, scant historical data, and inconsistencies in records. But the project highlights the application of machine learning to handle real-life problems in crime prevention. It opens up a future direction integrating new, real-time data, socio-economic variables, and geospatial analyses to make predictions more precise and provide actionable insights dynamically. Overall, the project shows the promise of data-driven approaches in improving public safety and reducing crime rates.

Introduction

First and foremost, crime prediction and crime analysis are essential components of contemporary law enforcement strategies aimed at preventing criminal acts, allocating scarce resources prudently, and ensuring public safety. In the past, crime analyses were entirely dependent on compiling historical crime data through statistical methods in order to discern patterns or trends. However, with increasing volume and complexity of data, the traditional methods would fail to handle some of the outliers or the nonlinear interactions that exist within the data, so that the amount of prediction accuracy has not been achieved at its full potential. Subsequently, there has been great interest in integrating advanced computational technologies into crime analysis and prediction. The emergence of machine learning (ML) as a potent tool for the prediction of crime through data for analysis, coupled with advanced algorithms, can reveal hidden patterns in those large datasets and make predictions based on the data. Learning algorithms that have been developed through ML systems include decision trees, random forests, and support vector machines, all of which can learn from previous crime events and make predictions on future crime incidents according to what time, where, and which type of crime occurred. They would also be able to capture complex interrelationships in data located within crime events that cannot be understood through traditional statistical means, therefore equipping individuals in the policing agency with tactical insight about likely places where crime will occur and on what trends. Apart from machine learning, time series forecasting techniques such as Seasonal Auto Regressive Integrated Moving Average (SARIMA) and Prophet are catching pace in developing detailed algorithms for time-based forecasting of crime patterns. SARIMA is a classical time series forecasting model that can model seasonality and trends in data, making it the best candidate for modelling crime events that are cyclical such as spikes during festive seasons and seasonal migration trends. In other words, SARIMA is capable of modelling both the short-term and long-tern relationships that exist as high predictor variables.

Prophet is a new technique for forecasting time series developed by Facebook for handling irregular data with missing values and outliers.

Citation: Yadav MB, Gayathri N, Srilekha M, Thanmayee S. Crime Prediction and Analysis using Machine Learning. GJEIIR. 2025;5(4):077.

It is best applied for sparse and highly inconsistent data like crime records. Due to great flexibility in the nature of the temporal structures being modelled, it can accurately model crime trends, thereby giving reliable predictions regarding seasonal effects and abruptly changing trends. This paper examines how machine learning approaches would be effectively integrated with SARIMA and Prophet for effective crime prediction and analysis. Such strengths when knitted together would be a good opportunity for law enforcement to build a comprehensive approach to crime forecasting and the development of their strategies for preemptive action. This confluence of machine learning, SARIMA, and Prophet is promising to the extent that it will improve crime prediction models and use optimal resource allocations in resource management while leading to effective police strategy.

Methodology

In an attempt to overcome the challenges of existing systems, the proposed crime prediction and analysis system will incorporate new technologies and new methodologies into its design. The system intends to give reliable forecasts and actionable insights through user-friendly interfaces tailored to the needs of different stakeholders, including law enforcement, policymakers, and analysts. SARIMA (Seasonal Autoregressive Integrated Moving Average) and Prophet are examples of machine learning models used in time-series forecasting by this system. These can detect trend and seasonal behaviour, anomalies, and long-term patterns that can result in accurate forecasts of the occurrence of crime. For dynamic analysis and immediate updates to ensure timely insights in pro-active decision-making, the system can integrate real-time data streams from various sources, such as police data, socio-economic indicators, and geospatial databases. The holistic view of crime trends is achieved through the integration of socio-economic, temporal, and spatial datasets. Using high-level analysis tools, the users can understand where and at what times they can observe positive correlations as well as areas of high risk and seasonal spikes in crime activities so that interventions can be targeted. The complex data would be made easily interpretable using user-friendly dashboards with interactive charts, heatmaps, and trend lines. These visualizations are available to both technical and non-technical users and enhance accessibility for decisions.

Working model

The Data Ingestion Module has responsibility for the import and validation of raw data from CSV sources. It is a major player converting exact, consistent data going into the system for analysis. Import from local storage, or cloud-based repositories. It supports different formats for flexible data ingestion. Automated validation routines identify missing values, duplicates, inconsistencies to ensure ingested data is complete and accurate. During the import process, all errors are logged into the system for subsequent developer action. Enables bulk data ingestion to scale the system for efficiency with large datasets. These are filters that exclude unwanted columns or rows from being loaded into the system. They are Optional. It allows attaching other data sources without major configuration changes. Enable direct integration with Data Processing Module for seamless workflow. Multi-Threaded Implementation Speeds imports of large volumes of data.

Data Processing Module is specifically made to clean the data coming in its raw form and prepare it for analysis along with the other aspects like making it accurate, consistent, and ready for machine learning model consumption. The prominent feature in it includes filling in missing values through imputation, deleting duplicates, and normalizing all data so it can be modelled. It also converts categorical variables into numerical formats, finds and cleans them from outliers, and also engineers some more features, such as time-based or aggregated metrics. Splits data into training and testing subsets, the modules scaled to large datasets and maintain a log of each processing step to keep it transparent. These processes together help improve the model performance and maintain reproducibility.

The Model Execution Module is the analytical heart of the system as it has machine learning models that can provide all actionable insights and predictive analysis. It comprises the relevant models, that is, SARIMA seasonal patterns and Prophet as trend lines. The module's training phase uses historical data for parameter optimization with respect to the accuracy of the models. Forecast to predict future trends in crime, along with confidence intervals, and reveal seasonal variation with respect to crime rates. Evaluation metrics such as RMSE and MAE are used for an assessment of model performance. The module provides real-time updates, dynamically changing models as there is new information getting available. It also supports batch processing for handling extensive datasets by employing parallel processing techniques. Also, the flexibility to integrate custom models tailored to specific domains and logging of all model parameters, iterations, and outputs for exhaustive documentation and auditing is offered.

The Visualization Module empowers the user to set up dashboards for viewing data in interactive charts and to extract insights easily and actionably. Some of the major functionalities include heatmaps that can show crime hotspots using spatial data and trend analysis to reveal the regularities over time in line and area charts. The module supports the interactive filters through which users may explore data sub-sets based on geography, time, or the type of crime. Personalized dashboards to meet the specific requirements of each stakeholder with added capacity for conditional display. The users can store their visualizations in different types such as PNG, PDF, or Excel files for reporting and subsequent analysis. It also supports real-time visualization; they include dynamic real-time updates as new information is received. Multi-layer maps orchestrate geospatial layers for better context, and the user-friendly interface makes sure commissioning an audience without technical expertise finds it easier to interpret complex data without specialised knowledge.

The Alert and Visualization Module is instrumental in automating notifications to the relevant stakeholders at high-risk time zones or places so that they can be alerted in a timely manner during critical situations. The feature of the module consists of having thresholds that are configurable, wherein a user can set up the actual conditions that may trigger alert generation, and notifications would be sent out in real-time within the affected areas of concern, using emails and SMS messaging thus keeping relevant stakeholders in the loop. Very likely, it brings out the concept of scheduled alerts for everyday announcements. The module has a section of priority levels based on urgency. The alerts are incorporated directly into dashboards providing ease of access and monitoring in real time. Historical logs maintain a record of all the alerts generated for future audits and study purposes. In addition, the module has custom alerts, allowing users to set up rule-based notifications as they desire. It was built for scalability that is, notifications would not slow down even when sent out to several stakeholders at once.

The Security Module is built to defend the sensitive data

types and also secured compliance with privacy standards and security standards. This includes some features like encryption which makes use of AES-256 to facilitate safe storage and transfer of data so that all information stays put while being transported in different media and even when it is at rest. Access control is done with the use of role-based permissions where unauthorized users are role-restricted from certain sensitive data and resources. The system is equipped with audit trails: all activities within the system are monitored concerning security and accountability. Data anonymization, on the other hand, is significant as it takes personally identifiable information to secure users. It has automated regular backups for data recovery during system failure. Firewall integration secures the system from unauthorized outside access. Modules ensure that compliance is with the GDPR and any other essential data protection regulation. In addition, the conduct of penetration testing is performed on timely

bases to find and remediate any potential weaknesses which would help keep the system insulated from the damage that might come from the new, evolving threat vector.

The Reporting Module provides reports capturing past as well as projected crime patterns. The module enables stakeholders to generate reports on crime trend history as well as projections. The major features of this module are as follows: automated report generation, which produces reports at regular intervals without human involvement, and visual summaries, which incorporate visuals, graphs, and key statistical information for easy interpretation. Users can create their own custom report templates for specific requirements and so have flexibility regarding report design. The module also offers multiple forms of export formats, including PDF, Excel, and CSV, for ease of dissemination and analysis. Real-time reports present up-to-date details based on the latest data, while users can use comparative analysis to see how actual historical results compare to what they predicted. These users should expect distribution options, which enable a report to automatically send to their stakeholders by email, and archiving. This ensures that all these reports generated are accessible at a later period. The module is thus robustly functional, designed for separate feature development and testing, with integration testing giving attention to the interaction between modules, which creates a facility to adapt as the system needs change.

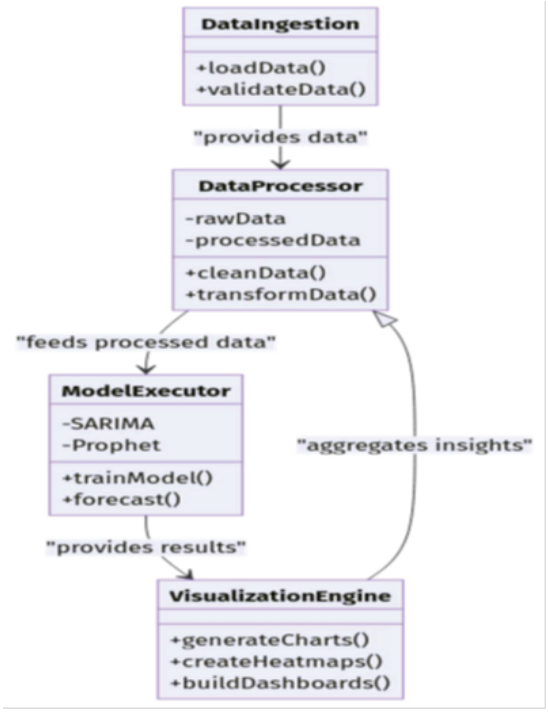

Figure 1. *Representation of Structure and relationships between different modules*

Result and discussion

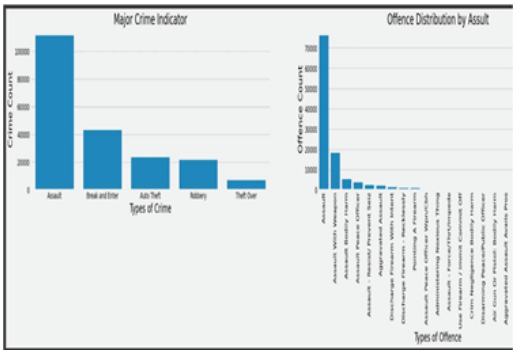

Figure 1. *Bar charts comparing crime categories like Assault and Auto Theft.*

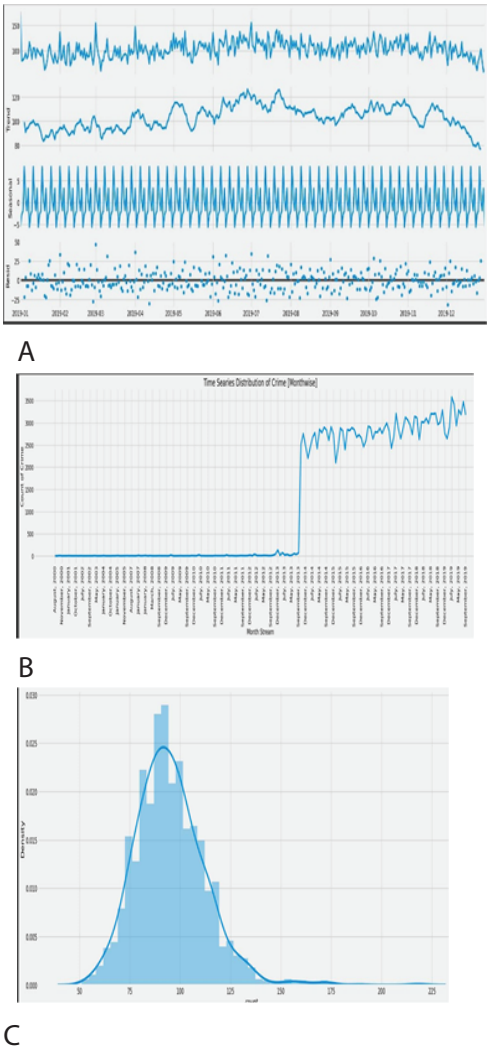

Figure 3. *Data transformation showcasing preprocessing steps.*

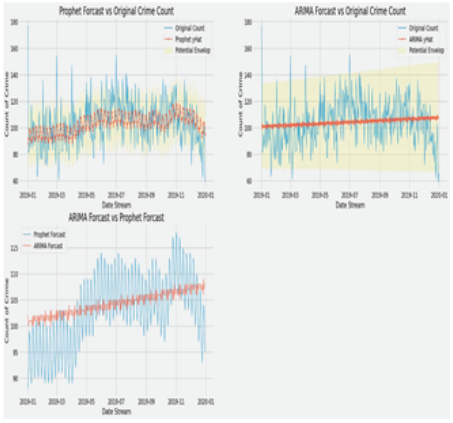

Figure 4. *Graph comparing actual vs. predicted values for SARIMA and Prophet models.*

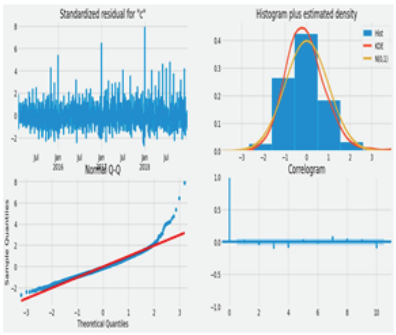

Figure 5. *Diagnostic plots for SARIMA, detailing residual patterns and fit.*

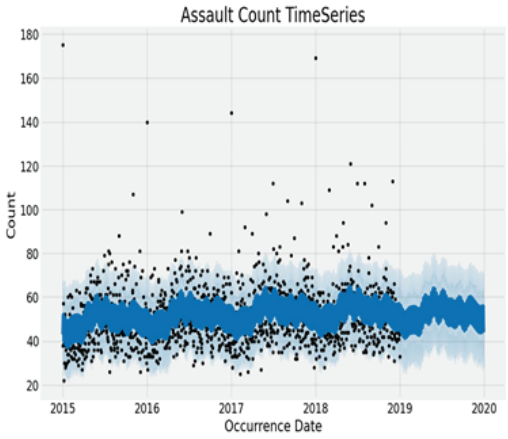

Figure 6. *Line graph visualizing crime trends derived from time-series data.*

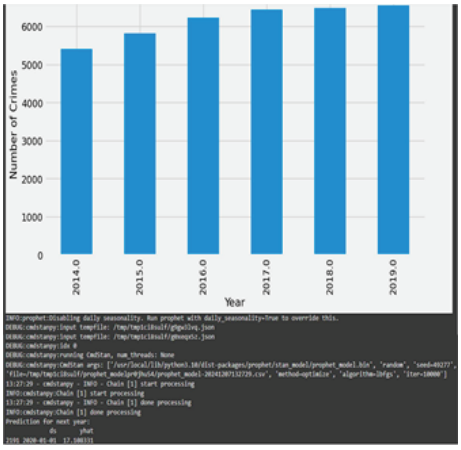

Figure 7. *Snapshot of a generated report, highlighting key metrics.*

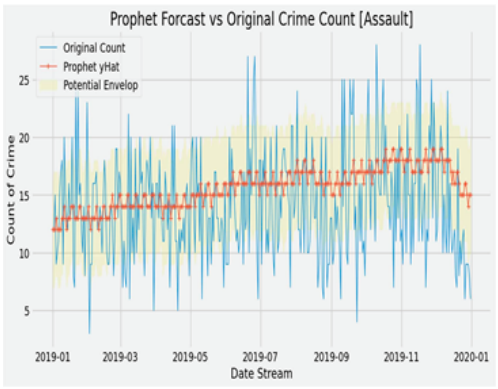

Figure 8. *Chart comparing historical data to forecasted crime rates.*

References

1. R. Bhallamudi et al., "Deep Learning Model for Resolution Enhancement of Biomedical Images for Biometrics," in Generative Artificial Intelligence for Biomedical and Smart Health Informatics, Wiley Online Library, pp. 321–341, 2025.
2. R. Bhallamudi et al., "Artificial Intelligence Probabilities Scheme for Disease Prevention Data Set Construction in Intelligent Smart Healthcare Scenario," SLAS Technology, vol. 29, pp. 2472–6303, 2024, Elsevier.
3. R. Bhallamudi, "Improved Selection Method for Evolutionary Artificial Neural Network Design," Pakistan Heart Journal, vol. 56, pp. 985–992, 2023.
4. R. Bhallamudi et al., "Time and Statistical Complexity of Proposed Evolutionary Algorithm in Artificial Neural Networks," Pakistan Heart Journal, vol. 56, pp. 1014–1019, 2023.
5. R. Krishna et al., "Smart Governance in Public Agencies Using Big Data," The International Journal of Analytical and Experimental Modal Analysis (IJAEMA), vol. 7, pp. 1082–1095, 2020.
6. N. M. Krishna, "Object Detection and Tracking Using YOLO," in 3rd International Conference on Inventive Research in Computing Applications (ICIRCA-2021), IEEE, Sept. 2021, ISBN: 978-0-7381-4627-0.
7. N. M. Krishna, "Deep Learning Convolutional Neural Network (CNN) with Gaussian Mixture Model for Predicting Pancreatic Cancer," Springer US, vol. 1380-7501, pp. 1–15, Feb. 2019.
8. N. M. Krishna, "Emotion Recognition Using Skew Gaussian Mixture Model for Brain–Computer Interaction," in SCDA-2018, Textbook Chapter, ISBN: 978-981-13-0514, pp. 297–305, Springer, 2018.
9. N. M. Krishna, "A Novel Approach for Effective Emotion Recognition Using Double Truncated Gaussian Mixture Model and EEG," I.J. Intelligent Systems and Applications, vol. 6, pp. 33–42, 2017.

10. N. M. Krishna, "Object Detection and Tracking Using YOLO," in 3rd International Conference on Inventive Research in Computing Applications (ICIRCA-2021), IEEE, Sept. 2021, ISBN: 978-0-7381-4627-0.
11. T. S. L. Prasad, K. B. Manikandan, and J. Vinoj, "Shielding NLP Systems: An In-depth Survey on Advanced AI Techniques for Adversarial Attack Detection in Cyber Security," in 2024 3rd International Conference on Automation, Computing and Renewable Systems (ICACRS), IEEE, 2024.
12. S. Sowjanya et al., "Bioacoustics Signal Authentication for E-Medical Records Using Blockchain," in 2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS), vol. 1, IEEE, 2024.
13. N. V. N. Sowjanya, G. Swetha, and T. S. L. Prasad, "AI Based Improved Vehicle Detection and Classification in Patterns Using Deep Learning," in Disruptive Technologies in Computing and Communication Systems: Proceedings of the 1st International Conference on Disruptive Technologies in Computing and Communication Systems, CRC Press, 2024.
14. C. V. P. Krishna and T. S. L. Prasad, "Weapon Detection Using Deep Learning," Journal of Optoelectronics Laser, vol. 41, no. 7, pp. 557–567, 2022.
15. T. S. L. Prasad et al., "Deep Learning Based Crowd Counting Using Image and Video," 2024.