Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 03 Issue: 12 | Dec-2025
ISSN (O) 3107-6696

Clustering -Based Crime Analysis and Pattern Identification

NEHA SP[1], Mrs Tejaswini A[2]

[1]Student, PES Institute of Technology and Management Studies, Shimoga,
[2]Assistant Professor, PES Institute of Technology and Management Studies, Shimoga

---

Abstract - Crime has increasingly become a major apprehension used for community security, as increasing criminal actions interrupt social stability and general security. The determined incidents instil fear and uncertainty in the minds of citizens. As a result preventing crime is the main responsibility. It needs to be taught in a deliberate and rational way in order to change crime. K means compiles algorithms in this work cluster that are connected to criminal episodes in an attempt to uncover hidden behavioral patterns. In contemporary crime analysis a variety of datasets and open-source tools are used to analyze past events and predict future criminal activity. In terms of the capacity to anticipate the detection of concealed behavior both groups had similar crime rates. This endeavor aims to enhance self-protective decision making by promoting a better understanding of crime and streamline ongoing projects it.

KEYWORDS- *Random Forest K-Means Clustering Machine Learning Crime Analysis Crime Prediction and Crime Patterns.*

Introduction

An introduction. One of the main issues in most cities is crime which causes chaos in both public life and particular safety. Law enforcement finds it challenging to identify a significant pattern due to the volume of crime records that continue to accumulate. Based on historical crime data the Crime Forecast models can predict areas where crime is most likely to occur through data analysis and machine learning algorithms. Both officials and ordinary people can use it thanks to its distinctive visual interfaces. The main technique used by the system to identify crime graphs and categorize related criminal incidents is K-Means clustering.

To improve the predictive capabilities machine learning models like SVM Random Forest and decision trees are used to identify crime categories and forecast future occurrences. After analyzing crime statistics and trends related to locations times frequencies and categories the system produces valuable information that helps law enforcement officials make wise choices. Customers can learn about variations and trends in crime statistics by using visual aids like charts and graphs.

LITERATURE SURVEYS

Many other techniques, including decision tree and K Means clustering, have remained successful in these studies and in crime prediction. Other systems, however, aim illicit activity detection. Regular tuition discusses the increased crime rates through a continuous presentation, each of which provides a range of methods for more precise predictions. The type of data employed, such as demographic or historic crimes data, and the features applied to prediction have a substantial impact on the recall and efficacy of these systems.

[1] "Crime Prediction Through Spatial–Temporal Analysis" Daniel Reyes, Lura Bennett (2016).

Reyes and Bennett stressed the application of spatiotemporal statistical models to find out how criminality varies with time and geography.

Their studies reveal that criminals' activities are non-random, but rather follow a pattern that depends on urban density, population mobility, and neighborhood association. By integrating both magnitudes, they were able to meaningfully improve forecast accuracy based on spatial regression and temporal disintegration techniques.

[2] "Machine Learning Approaches for Urban Crime Forecasting" Priya Sharma, Neeraj Patel (2020).

Sharma and Patel investigated the possible of measured machine learning systems, along with gradient boosting and multilayer perceptron's, and naive Bayes classifiers, in order to forecast the crime rate in urban regions. Their research focuses on feature engineering, showing how different variable star, such as weather patterns, public events, and transit hubs, affect the prediction results. Affording to their finding's different datasets, gradient boosting representations done well than additional models.

[3] "A Data Mining Framework for Crime Hotspot Detection" John Kweka, Samuel Mrosso (2017).

This study suggests a comprehensive approach for identifying crime hotspots using data mining. To identify places with high-rate crimes, the authors employed anomaly detection,

heat map visualization, and clustering techniques. They concluded that real-time patrol deployment and the operational response is probably finished by automated hotspot detection.

[4] "Deep Learning-Based Crime Event Classification" Liyun Chen, Mark Johnson (2021).

Chen and Johnson used written police reports to categorize some types of crimes using neural networks, i.e. recurrent and convolutional models. Through the detection of philological gradations in incident reports and the support of increased exactness in crime classification, their education established that in-depth learning is greater to conventional methods.

PROPOSED METHODOLOGY

The growth of the Crime Awareness scheme shown here includes a quantity of important developments in the use of machine learning algorithms to efficiently estimate illegal behavior. This was proficient by gathering crime statistics from responsible sources, such as places, demographic information, and past crime reports. The data was then exposed, copies were detached, missing values were handled, and formatting was performed to prepare it for machine learning. Datasets are detached into tool and test sets following preprocessing. In spirit, feature selection is utilized to identify related metrics and specific situations for crime prediction.

Proposed Model Diagram

The benefits of the current method are covered in several areas. By combining both approaches to improve the domain results, researchers have created hybrid domain systems that aim to inherit the strengths and avoid the shortcomings of both approaches. Generally, a hybrid system recommender integrates several recommendation techniques to optimize their synergy. Although many useful suggestions may be combined, our activity will primarily focus on the integration of CF and CBF approaches

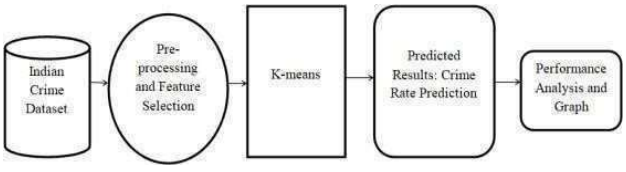


FIGURE 1.1: Proposed Model for Crime Identification

Block Diagram for ML model

Block diagrams are a simple and effective approach to illustrate how a system uses its fundamental graphical component. By showing the input data entering the system, the various operations or processes performed on the data, and the final output produced, it visually depicts information flow within the system. Block diagram are critical gears for system modeling because they simplify the sympathetic of the system structure and behavior. The components of a block diagram usually represent the system's operations and the flow of data between its many components. Block diagrams are used to deconstruct intricate systems into simpler parts that are easier to analyze, create, and explain.

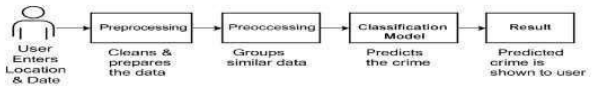


FIGURE 1.2: Block Diagram for Crime Prediction

Datasets and methods for predicting crimes

Fieldwork is the primary source of information for crime databases. Important details, including the year, age, crime type name, crime location, crime date, crime group name, and unit name, were chosen based on this data. The system uses the selected data as input features.

Year	Crime Areas	Crime Type
2010	Chennai-Tamil Nadu	Murder
2010	Chennai-Tamil Nadu	Theft
2010	Madurai-Tamil Nadu	Murder
2011	Nilgiris-Tamil Nadu	Dowry Death
2011	West-Delhi	Theft
2012	East-Delhi	Rape

TABLE 1.1: A Basic Dataset Used for Crime Forecasting

The gathering of data, which may include information, crime figures, and environmental concerns, jolts the process The first step in the process is collecting data such as criminal histories

statistical trends and other relevant information. The data is reprocessed after collection in order to fix missing values eliminate inconsistencies and format it for model development. To assess the models performance objectively the cleaned data will be split up into training data sets.

- **Predicting crimes through machine learning**

Outdated machine learning methods like SVM clustering and K means are still essential in criminal research and forecasting. These algorithms are easy to implement have low processing power requirements and are straightforward to comprehend. K Means for example can help researchers find important trends in criminal episodes based on demographics time periods or geographic locations. Thus anomaly detection methods highlight odd behaviors that might point to fresh dangers.

- **Machine learning techniques for criminal activity**

To help predictors find hidden trends in crime datasets K Means clustering is used to group similar crime episodes. Numerical results are predicted by regression. On the other hand K Means searches for common data clusters. This method is frequently employed by investigators to examine recurring crime categories and ascertain the geographical and temporal distribution of these crimes. K Means may be able to identify hotspots and behavioral patterns by classifying crime episodes based on variables such as locations year frequency and relevant socioeconomic conditions.

Mathematical Formula

The forecasting capability of our sorting system was measured by four major metrics, which include precision, recall, accuracy, and F beta score. crime. This allows us to understand the model's potential from a number of perspectives.

- Precision (q): Precision indicates the accuracy of the projected affirmative cases.

$$\text{Precision (q)} = \frac{tv}{tv+fv}$$

Where fv = False Positive (crimes incorrectly predicted) and tv = True Positive (crimes successfully predicted)

- Recall(rc): Recall quantifies the quantity of models that precisely recognize true positive case, or actual crimes.

$$\text{Recall (rc)} = \frac{tv}{tv+fnv}$$

Where, fnv stands for False Negatives (missed crimes).

- Accuracy: Accuracy is the proportion of overall predictions—both criminal and non-criminal—that were accurate.

$$\text{Accuracy} = \frac{tv+tnv}{tv+fv+fnv+tnv}$$

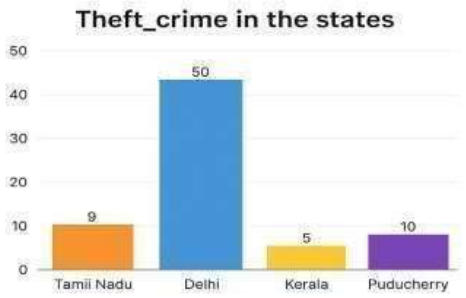
Where tnv stands for True Negative, or accurately anticipated non-crimes.

- F-beta Score: By integrating recall and precision, the F-beta Score provides a extra inclusive picture. They are more dependable than accurate, and the data is unbalanced.

$$\text{F-beta Score} = \frac{2*(rc*q)}{rc+q}$$

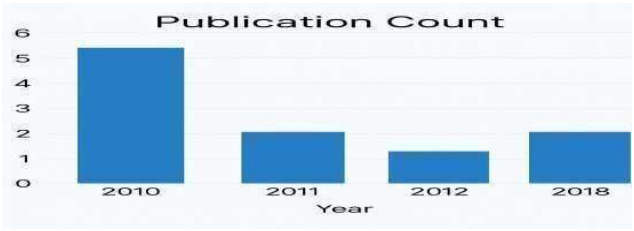
This rating evaluates how successfully the model strikes a compromise between preventing False warnings (precision) and accurately recognizing crimes (i.e., recall).

V. GRAPHS



Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 03 Issue: 12 | Dec-2025
ISSN (O) 3107-6696

GRAPH 1.1: Distribution of Crime Analysis



GRAPH 1.2: Distribution of research publications for trends in crime prediction by year

Using historical crime statistics, the graphs show how crime incidents have changed. Graph 1.1 narrows down the hours with the highest incidence by illustrating how incidents change with each hour of the day. The quantity of studies on crime forecast that were available among 2018 and 2021 is compiled in Graph 1.2. It is evident that research activity increased steadily up to 2021, after which it slightly decreased in 2022. The COVID-19 lockdown provided sufficient opportunity for academics to delve deeply into crime-related studies, which is likely why there was a big surge in 2020 and 2021. In conclusion, these graphs reveal the timing of the crimes and demonstrate the growing interest in crime-prediction research.

VI. EXPERIMENTAL RESULT

A historical crime dataset with information's including year, locations, and incident, types was exploited to judge the Crime Insight systems.

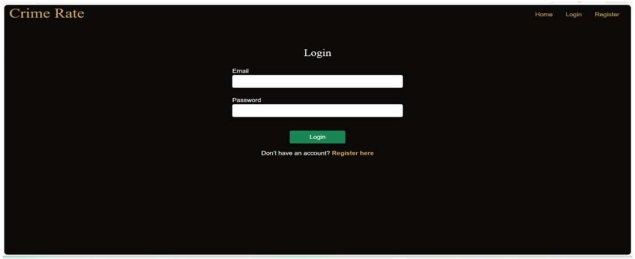
Before the data was cast-off to trains and test several machine learning models, it was cleaned and pre-processed. Reproduce and algorithms are including Random Forest, K-Means, and SVM, were associated assess their predictions performance. Among Random Forest deliveries the highest accuracy in identifies crime types and location. The systems were able to detect consistent patterns and highlight areas with higher risk founded on trained models. Model presentation was unhurried using classification intelligences, confusion matrices, and accuracy metrics.

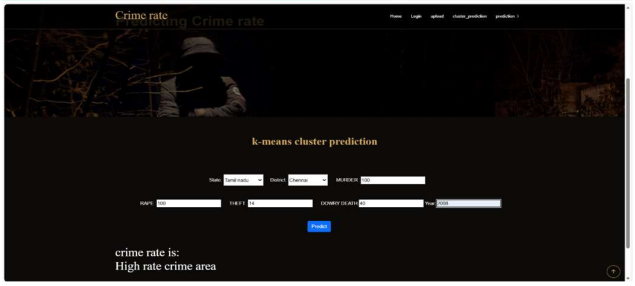
Overall, the consequences show that the proposed approach can generate reliable predictions and support law-enforcement decision-making.

Model	Accuracy	Precision	Recall	F1 score	Training time
Random Forest	89.7%	89.3%	88.5%	88.9%	2.4
Support Vector Machine	86.7%	85.2%	84.9%	85.0%	3.2
K-Nearest Neighbours	81.5%	78.0%	79.1%	79.5%	0.9

TABLE 2.1: Crime Performance Comparison Using Machine Learning Techniques

VII. RESULTS





VIII. CONCLUSION

Crime has become more advanced with technical growth, creating new tasks for law-enforcement activities. Machine learning is increasingly used to analyse criminal behaviour and forecast future occurrences. Researchers study crime systems and patterns to improve prediction accuracy. The Crime Awareness system supports this process by investigative ancient data to identify areas and times with higher crime risk. These perceptions help police plan protective actions and respond more efficiently. The system also allows the public to

recognize crime trends easily. Future enhancements may include real-time data, mobile features, and stronger algorithms. Such upgrades can further enhance public safety. Using technology in this way helps build harmless and more secure societies. This research highlights the value of ML procedures in crime prediction. It also offers useful datasets and instructions for future trainings.

By enabling better planning and quicker response, the system supports overall public safety. In the coming, Crime Insight can be enhanced with features such as mobile access, live data combination, and more advanced algorithms to improve forecast accuracy. Using technology in this way is an effective approach to creating safer communities. This study also donates to the field by importance the effect of various machine learning methods and present useful datasets and directions for investigators who continue to explore crime prediction.

IX. FUTURE ENHANCEMENT

Crime Insight can be improved in a quantity of behaviors in the coming for sensible application. For better projections and real-time information, it can be connected to real-time crime statistics from police and news websites. The technology uses sophisticated algorithms to increase accuracy. It is possible to incorporate a map function that shows crime-ridden areas as heatmaps. It is possible to create a mobile application that both users and law enforcement can readily access at any time. People may also be intelligent to usage the system to report crimes directly. It is possible to search social media for early signs of problems or criminal activity. The system has a voice assistant to help users and is multilingual. People in dangerous places may receive warnings. Real-time monitoring and prompt response can be facilitated by adding extra security, feedback systems, and integration with CCTV or sensors. These enhancements will increase Crime Insight's efficacy and benefits both the general public and law enforcement.

REFERENCES

[1] A hybrid learning framework for forecasting urban crime patterns," by R. J. Kumar, A. Banerjee, and S. Verma, in IEEE Int. Conf. Data Sci.

Adv. Analytics (DSAA), 2019, pp. 112–119.

[2] Spatial clustering and predictive mapping of criminal incident using GIS-integrated machine learning," International Journal Geoinformatics, vol. 16, no. 2, pp. 45–58, 2020, H. Liu and C. Park.

[3] Deep spatio-temporal networks for crime trend prediction in metropolitan regions," by M. Rodriguez, T. Silva, and J. Gomes, in IEEE Int,

Joint Conf. Neural Netw. (IJCNN), 2021.

[4] Crime risk assessment using boosted trees and socio-economic indicators," D. Choi and R. Malik, Expert Syst. Appl., vol. 159, Mar. 2020.

[5] Predicting burglary hotspots using random forest and density-based clustering, by L. Santos and H. Ribeiro, Proceedings of the
Global Conference on Big Data (BigData), 2021.

[6] Text-driven crime category prediction from police narratives using NLP and CNN," ACM Trans. Inf. Syst., vol. 40, no. 4, pp. 1–23, 2022,
G. Ahmed, P. Chauhan, and S. Naaz.