IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Analysis and Comparison of Machine Learning Methods for Crime Hotspot Detection

[1]Dr Samit Kumar Singh, [2]Dr P Suresh Kumar, [3]Dr B Raja Narender, [4]Dr M Sandhya Rani

[1] Associate Professor, Department of H&S - Mechanical Engineering

[2] Associate Professor, Department of Computer Science & Engineering,

[3] Associate Professor, Department of H&S - Mechanical Engineering,

[4] Associate Professor & HOD, Department of Information Technology,

[1,2,3,4] Bhoj Reddy Engineering College for Women, Hyderabad, Telangana, India

Abstract—Crime analysis studies criminal patterns to support law-enforcement decisions. As digitalization generates massive crime datasets, manual analysis becomes inefficient, increasing the need for automated, data-driven methods. This study proposes a machine-learning framework for Crime Hotspot Detection using the Vancouver Crime Dataset (2003–2017). Four models—Random Forest, KNN, Decision Tree, and Naïve Bayes—were applied to spatial and temporal features to identify high-risk zones. Random Forest performed best, showing machine learning's usefulness for proactive policing and public-safety improvement.

Keywords— Crime Analysis, Machine Learning, Prediction, Data Mining.

I. INTRODUCTION

Criminality activity is increasing and becoming harder for authorities to manage, as offenders now exhibit both random and organized behaviors. While studying victims is challenging, crime scenes and digital records provide valuable clues. Law enforcement agencies must therefore rely on advanced analytical methods to detect crime patterns and predict future offenses from large, complex datasets. Predictive policing, powered by data mining and machine learning, enables early identification of high-risk areas and supports efficient resource allocation.

With rapid urban growth and evolving criminal methods, crime patterns have become more dynamic and unpredictable. Prior studies such as Zhang et al. [1] and Mugdha et al. [2] emphasize the growing importance of data-driven policing tools. This study examines multiple machine-learning approaches for hotspot detection and outlines related work, analytical methodologies, results, and future directions aimed at enhancing crime prevention and public safety.

II. LITERATURE SURVEY

To remedy classification errors and pick attributions with more values, ID3 revisions employ importance-attribute significance on attributes with fewer values but more importance [1]. This study [2] used cluster analysis to classify analogous noise road occurrences for road characterization. A dataset of Milan, Italy, noise events confirmed the method. distinct huge cities have distinct population, traffic, and event densities [5]. This means crime rates vary widely [4], [5]. When searching crime data for real hotspots, this is crucial. Recently published research [3], [4], [5] shows that multi-density clustering is superior than traditional

methods for locating hotspots in cities with varied density levels, especially when crime event density changes often. In [3], multi-density clustering tactics outperform standard methods for urban hotspot identification.

III. METHODOLOGY

A. Prediction Model:

This study predicts crime utilizing complex algorithms like Random Forest, KNN Classification, Decision Trees, and Bayesian Methods. Machine learning algorithms create prediction models with loads of data. More data helps machine learning forecast.

A methodology for cluster recognition using k-means and crime forecasting using neural-fuzzy networks is provided in [6]. The UCI crimes dataset "Communities within the United States" contains 43,420 crime events. The proposed forecasting methods outperformed Regression Tree and Gaussian Process Regression in terms of RMSE and MAE.

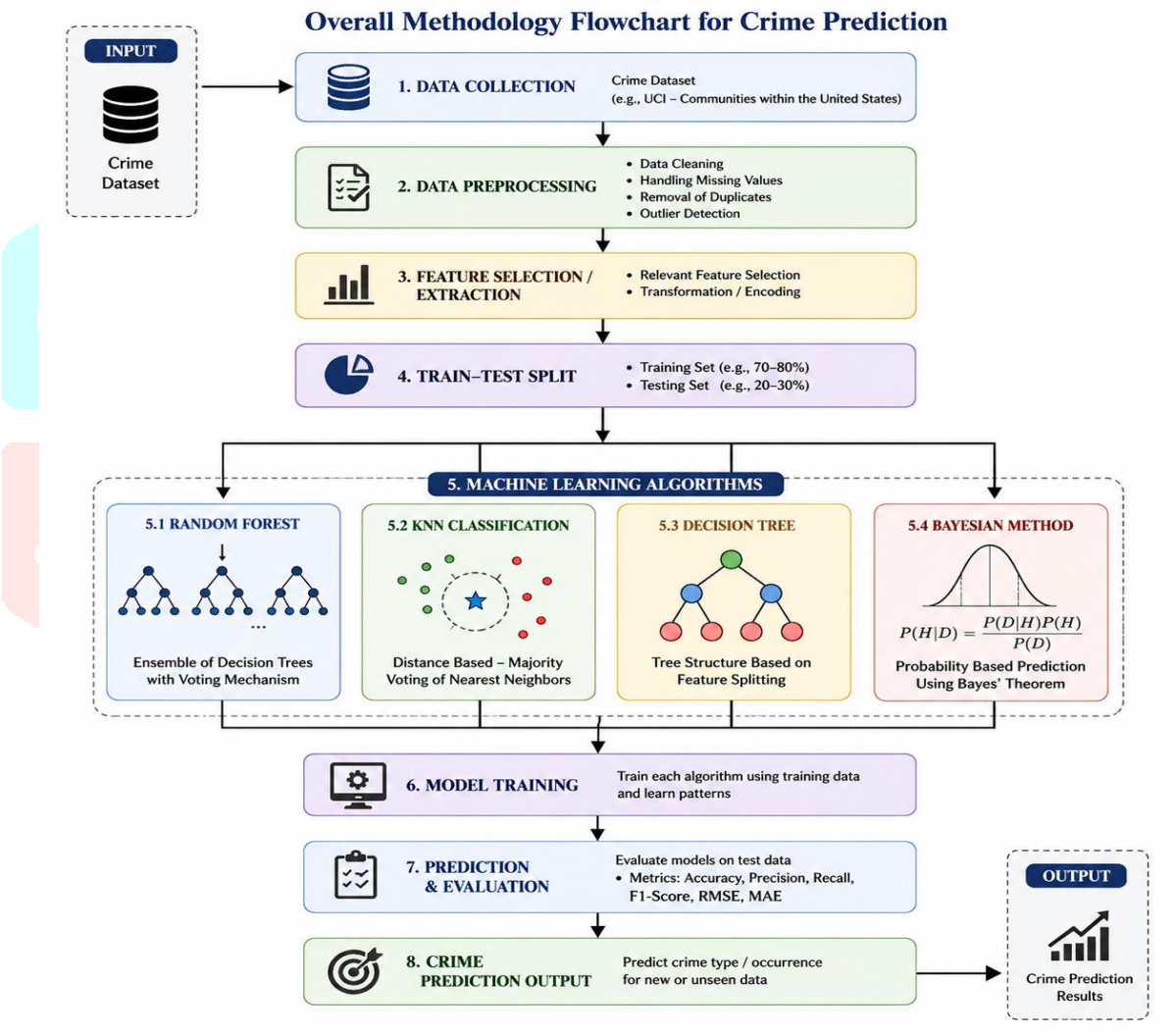

Fig. 1: Proposed Machine Learning Based Crime Prediction Framework

The described works have different purposes. Some approaches focus primarily on crime prediction [8], some on hotspot discovery [11], while others contribute region discretization to event forecasting workflows [4], [7]. Crime hotspot finding is the emphasis of [9], which uses Random Forest and Kernel Density Estimation. [11] uses Graph Convolutional Networks. Tensor representation and decomposition are used in [8].

1. Random Forest

Random forests are composed of tree classifiers {h(x, αk),k = 1...}, where x is the input vector, αk is a distributed independent random vector, and output is chosen by voting The meta classifier in CART is an uncut regression tree (h(x, αk)). Like random forest, bagging isolates training sample attributes. We employ M attributes, assign F < M to internal nodes, randomly select F attributes, and find the best split mode for training f attributes. Divide nodes. MDT random forest classifier-voted.

2. KNN Classification

Nuanced KNN finds K nearest categories using instance feature vector. Calculate training-set-to-new data feature value distance carefully. Data is classified by k=1 nearest neighbor. Distance-based weighting and majority voting improve KNN classification. Weighted voting considers neighbors. The nuanced method classifies closest neighbors better since they share qualities. Class majority votes count in k neighbors. Inputs are mostly classified. KNNs classify using k neighbors. Local training case consensus classifies input. KNN is adaptable for pattern detection and predictive modelling due to complex feature vectors, distance computations, and decision-making.

3. Decision Tree

Decision Regression and classification trees are supervised machine learning. Subgrouping input space by features optimizes Gini impurity or information gain. Alternatives branch at nodes. Feature thresholds determine best data pattern tree splits. The tree predicts and explains financial, healthcare, and marketing decision criteria using input-based branches.

4. Bayesian Methods

Bayesian machine learning updates model parameters using Bayes' theorem and new data. Bayesian inference treats model parameters as probability distributions, unlike frequentist techniques. Bayes' theorem converts parameter evaluations to posterior distributions from likelihood function data. It models complex interactions, evaluates uncertainty, and forecasts probabilistically. Due to data accumulation, Bayesian projections work with less data. Flexible and robust, they manage regression, classification, and model selection uncertainty.

B. Crime Analysis using ML Algorithms

1. Define the objective of the Problem Statement

Expectations matter. Forecast rain using weather data. Note facts that aid problem-solving or action.

2. Data Gathering

Acquire recognized data. Web scraping and manual data acquisition are choices. Beginners don't need data for machine learning. Collect data from lots of websites. Crime hotspot detection requires data on crime type, time, location, population density, and environmental factors. Proper data collection and storage are essential for accurate analysis and reliable hotspot prediction. Data collection and preservation are needed for analysis.

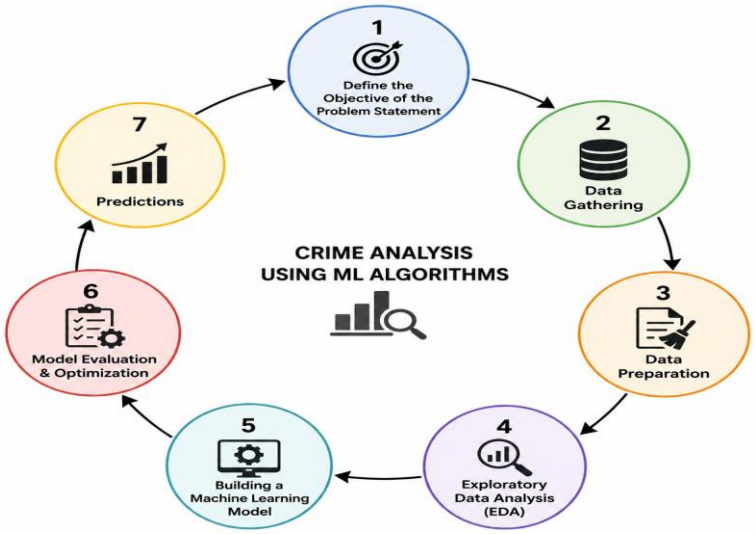

Fig 2: Workflow of Crime Prediction using ML Algorithms

3. Data Preparation

Most data are mis formatted. Data is multivariable, redundant, and missing. Remove these discrepancies to avoid inaccurate computations and forecasts. Fix data discrepancies quickly.

4. Exploratory Data Analysis

Intensive data analysis reveals concealed crimes. Machine learning brainstorming begins with Exploratory Data Analysis (EDA), which helps find trends and anomalies in criminal data. This phase shows how location, time, crime type, and arrest rates relate. Understanding that late-night hours, locations, and seasons increase crime risk is crucial for crime prediction. Correlation detection and thorough documenting of insights using EDA constitute the basis for accurate predictive models

5. Building a Machine Learning Model

Machine Learning Models use all Data Exploration trends. Start with training and testing data. The model will be built and tested using training data. Machine-learning model reasoning. Data, complexity, and problem type determine solutions. The machine learning fixes various difficulties.

6. Model Evaluation & Optimization

Training data-based model testing follows. We test the model's accuracy and efficacy. Checking improves models. Cross-validation and parameter tweaking improve model performance.

7. Predictions

Refined models predict crime categorically.

IV. EXPERIMENTAL AREA AND DATA VISUALIZATION ANALYSIS

A. Experimental Area

For This study uses free Kaggle Vancouver crime dataset which comprises 530,652. Regional crime patterns exist in the dataset. The complicated crime scene in Vancouver is shown by crime category, year, month, day, hour, location, latitude, longitude, and more [1]. Annual, monthly, or hourly analysis may reveal trends. Crime distribution and concentration by latitude, longitude, and place. Full criminal event and pattern analysis is possible with the dataset's broad attributes. This massive, well-curated dataset is used to research Vancouver's crime dynamics using algorithms and machine learning. Understanding regional crime's many features will improve crime prevention and law enforcement forecasts.

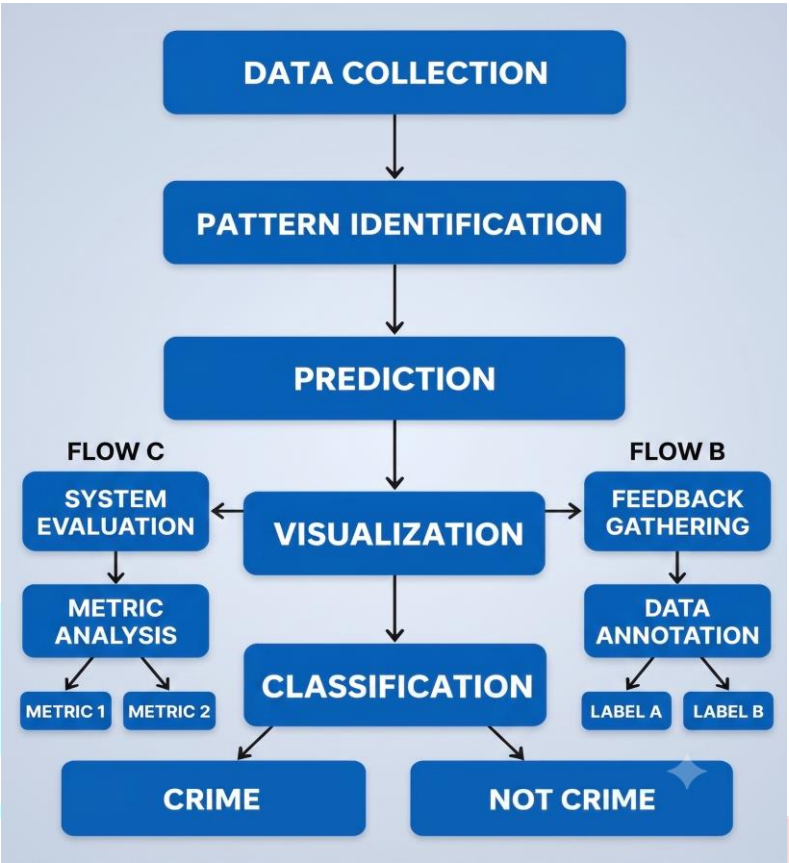

Fig 3: Experimental Steps

B. Selection of Type of Crime

Burglary is commonly considered "crime of property in public places". Snatching, robbery, and embezzlement that steal property fall under this category. Finding hotspots in this town's public property crime statistics is feasible. From idle enforcement to proactive prevention and management with precise crime predictions, police can improve community safety.

C. Data Visualization Analysis

The experimental district's P-GIS database shows 2015–2018 crime statistics. Following street range location, the study area map provides case point data and database text coordinates. To imitate police activity, crime hot spot prediction tests should be small. In Griffith et al.'s gridding processing method, 150 m x 150 m grids are used to disperse data and investigate police activity. Larger 150-m grids reduce hotspots and concentrate case points [12]. This section improves crime hotspot predictions and incidence distribution. One officer can patrol 150 meters each unit under current police protocols. Predictions aid crime prevention and management.

Fig 4: Data Visualization [1]

Fig 5: Year-wise Crime [1]

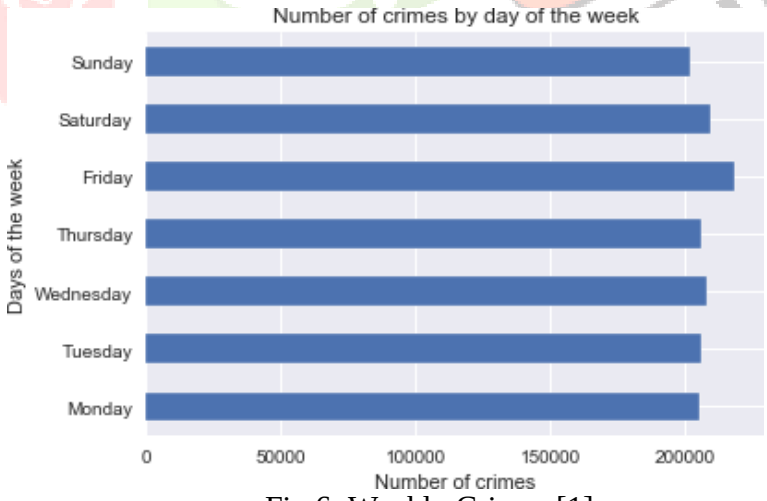

Fig 6: Weekly Crimes [1]

D. Statistics of cases

Only 2012 had more cases, and 2015 had the fewest of the five. Two-week cases fluctuated across six years. Most two-week intervals had 40–80 cases, averaging 58. Similar case volume curve change trend over six years is shown in figure. Thus, two weeks before and after the holidays, case volume lowers and climbs. Every January and February, cases fall. Every year, spring break had the fewest cases.

Fig 7: Rolling sum of all crimes from 2008-2018

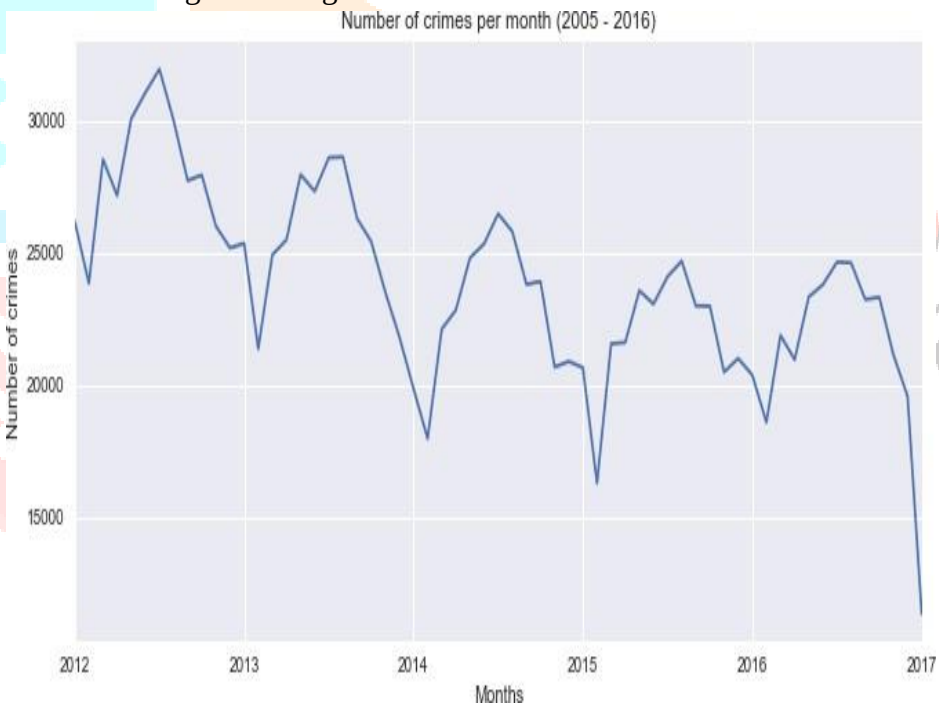

Fig 8: No of crimes/month from 2005-2016

E. Time Series Analysis

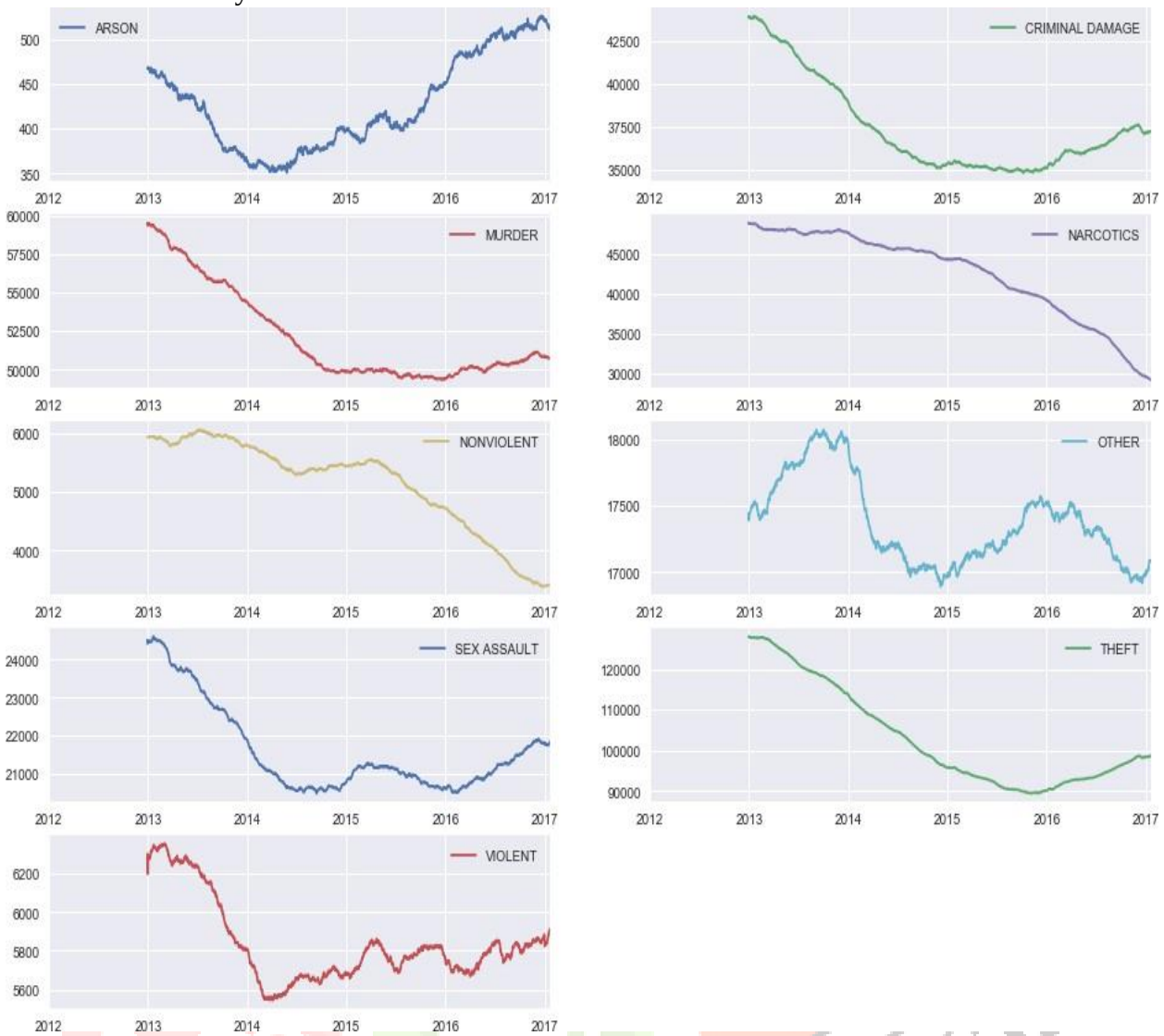

Fig 9: Type of Crime: Year vs No. of crime [1]

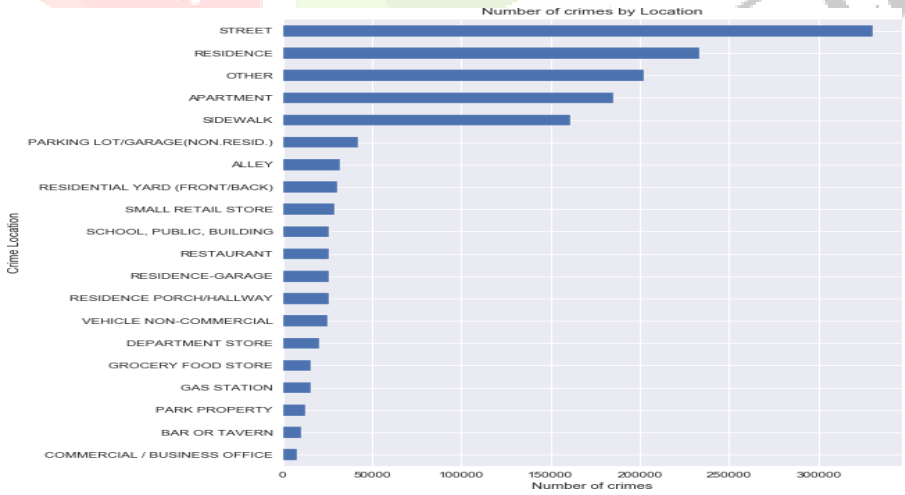

Fig 10: Number of Crimes by Location [1]

The Seasonality, trends, and the difficult six-year crime progression are presented in an additive time series decomposition graph. The upper graphic shows trends and variations from a daily time series. The bottom graphic illustrates monthly time series with longer-term trends [1].

The accompanying visualizations tracked crimes over six years. Arson, theft, non-violent offences, sexual assaults, murder, criminal damage, and violent crimes are covered. This tighter segmentation allows crime

category analysis and comparison, revealing chronological linkages or divergences.

This visual tour uses advanced analytics to analyze crime types' history and complexity. Police make choices and prevent crime by analyzing crime patterns.

Below are bar graphs showing the number of crimes over six years by region, type of area (streets, dwellings, flats, parking spaces, lanes, etc.), and offence.

The bar graphs show crime by street, residence, flat, parking lot, alley, and more. This large study detects criminal episodes in various settings and explains their geographical dynamics and contexts.

The bar graphs show theft, assault, burglary, vandalism, and other crimes over six years. This comprehensive strategy explores patterns, correlations, and trends to understand crime dynamics and influence crime prevention and public safety measures.

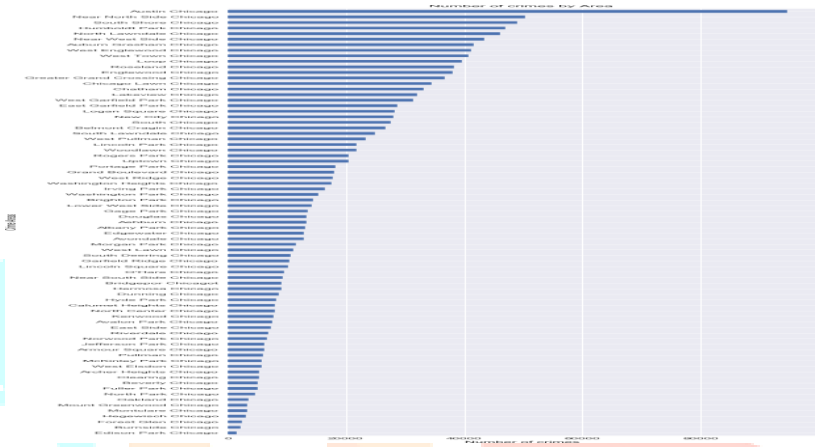

Fig 11: Number of Crimes by Area [1]

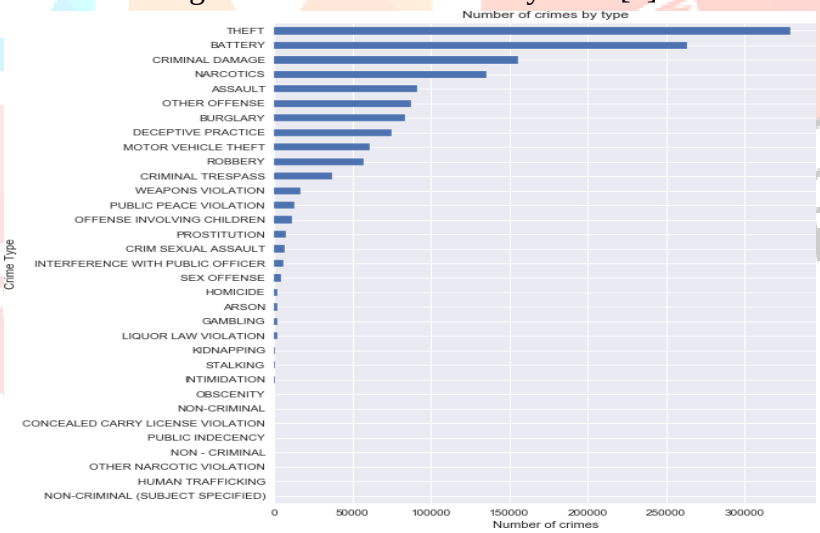

Fig 12: Number of Crimes by Type [1]

V. RESULT ANALYSIS

This study focuses on data mining–based crime analysis with the objective of reducing crime using advanced technologies and enabling proactive law enforcement. Crime rates are analyzed using five years of data on cases filed and arrests, revealing a high number of reported cases despite modest arrest figures, emphasizing prevention over arrests. Table X compares four classifiers: RF, KNN, DT, and NB. Random Forest outperforms others with the highest accuracy (96.3%), balanced precision and recall, and low computation time. KNN and DT show competitive results, while Naïve Bayes performs least effectively. Overall, ensemble methods prove most suitable for crime prediction.

Table 1: Results

Model	Accuracy %	Precision %	Recall %	F1 Score %	Time/Sec
RF	96.3	95.01	96.00	95.92	3.2
KNN	94.2	94.90	94.7	95.03	3.9
DT	93.5	91.32	98.26	94.62	4.1
NB	88.7	88.1	87.6	88.2	3.7

VI. CONCLUSION

Close inspection and analysis reveal that the Random Forest categorization approach increases complex experimental data processing logical coherence and usability. This technique reveals co-offenders' complex networks' hidden links, including criminal partners and networks beyond the network's context. RF classification is over 96% accurate. The full forensic kit generates and analyses comprehensive victim system data during assault. Despite technology advances, the Criminal Investigation Analysis (CIA) program must be addressed to facilitate complicated violent crime investigations. Low CIA program accuracy concerns real-world inquiry efficacy. Better forensics can influence criminal investigations.

REFERENCES

1. X. Zhang, L. Liu, L. Xiao, and J. Ji, "Comparison of machine learning algorithms for predicting crime hotspots," *IEEE*, 2020.
2. [2] U. M. Butt, S. Letchmunan, F. H. Hassan, M. Ali, A. Baqir, T. W. Koh, and H. H. R. Sherazi, "Spatio-temporal crime predictions by leveraging artificial intelligence for citizens security in smart cities," *IEEE Access*, vol. 9, pp. 47516–47529, 2021.
3. E. Cesario, P. Lindia, and A. Vinci, "Detecting multi-density urban hotspots in a smart city: Approaches, challenges and applications," *Big Data and Cognitive Computing*, vol. 7, no. 1, p. 29, Feb. 2023.
4. E. Cesario, P. Lindia, and A. Vinci, "Multi-density crime predictor: An approach to forecast criminal activities in multi-density crime hotspots," *Journal of Big Data*, vol. 11, no. 1, p. 75, May 2024.
5. E. Cesario, P. I. Uchubilo, A. Vinci, and X. Zhu, "Multi-density urban hotspots detection in smart cities: A data-driven approach and experiments," *Pervasive and Mobile Computing*, vol. 86, Art. no. 101687, Oct. 2022.
6. M. Jaber, R. Sheibani, and H. Shakeri, "A model for predicting crimes using big data and neural-fuzzy networks," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 17, p. 6985, Aug. 2022.
7. Z. Li, C. Huang, L. Xia, Y. Xu, and J. Pei, "Spatial–temporal hypergraph self-supervised learning for crime prediction," in *Proc. IEEE 38th Int. Conf. Data Engineering (ICDE)*, pp. 2984–2996, 2022.
8. W. Liang, Z. Wu, Z. Li, and Y. Ge, "Crime tensor: Fine-scale crime prediction via tensor learning with spatiotemporal consistency," *ACM Transactions on Intelligent Systems and Technology*, vol. 13, no. 2, pp. 1–24, 2022.

9. R. Ahmad, A. Nawaz, G. Mustafa, T. Ali, M. Tlija, M. A. El-Meligy, and Z. Ahmed, "CHART: Intelligent crime hotspot detection and real-time tracking using machine learning," *Computers, Materials & Continua*, vol. 81, no. 3, pp. 4171–4194, 2024.
10. T. Zubair, S. K. Fatima, N. Ahmed, and A. Khan, "Crime hotspot prediction using deep graph convolutional networks," *arXiv preprint* arXiv:2506.13116, 2025.
11. Q. Zhu, F. Zhang, S. Liu, L. Wang, and S. Wang, "Static or dynamic? Characterize and forecast the evolution of urban crime distribution," *Expert Systems with Applications*, vol. 190, Art. no. 116115, Mar. 2022.
12. U. Thongsatapornwatana, "A survey of data mining techniques for analyzing crime patterns," in *Proc. 2nd Asian Conf. Defence Technology (ACDT)*, pp. 123–128,2016.