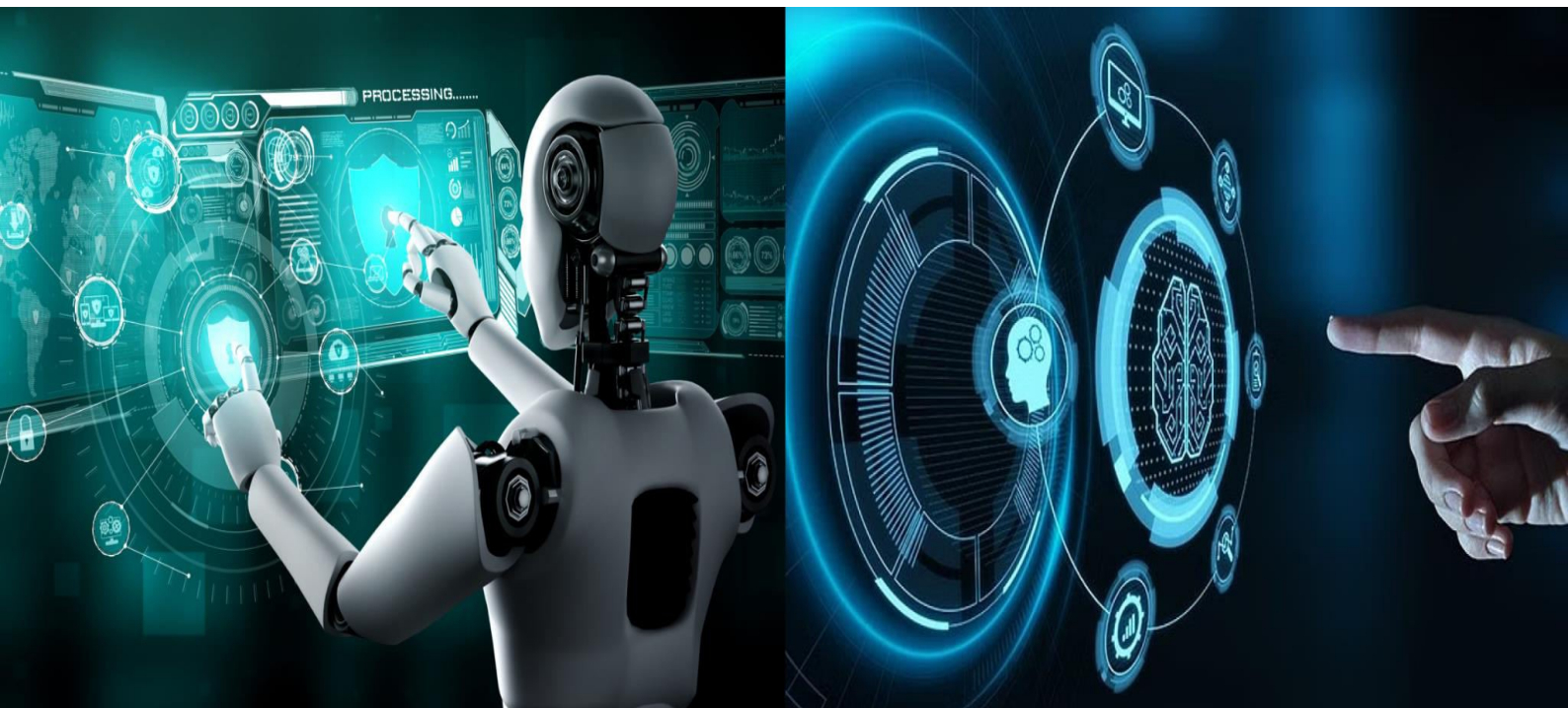


International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)





AI Powered Crime Prediction: Type and Frequency Analysis using Machine Learning

M Sowmiya, Vijay.C, Sivasakthi.V, Tamilarasu.S.S, Thalapati.G

Assistant Professor, Department of Computer Science and Engineering, The Kavary Engineering College, Mecheri, Salem, India

Department of Computer Science and Engineering, The Kavary Engineering College, Mecheri, Salem, India

ABSTRACT: Crime prevention and public safety have become vital concerns in modern society. With the increasing complexity and frequency of crimes, traditional crime-fighting techniques are insufficient. This research paper explores how Artificial Intelligence (AI) can be leveraged for crime rate detection and prediction. We focus on three powerful machine learning algorithms: Decision Tree Classification, Random Forest, and Logistic Regression. These algorithms are applied to historical crime datasets to detect patterns and predict future crime hotspots. The study aims to develop a predictive framework that can assist law enforcement agencies in better resource allocation and proactive policing. Governments and law enforcement organizations are looking for cutting-edge ways to identify, stop, and forecast crimes in light of the escalating rate of criminal activity worldwide. The study on artificial intelligence (AI)-based crime rate detection and prediction is presented in this publication. Decision Tree, Random Forest, and Logistic Regression are machine learning algorithms that are used to identify crime patterns and forecast future trends. To assess the accuracy and efficiency of the suggested models, realworld crime datasets are used for training. According to our findings, AI can greatly support resource allocation, policymaking, and proactive policing. Random Forest exhibits the best prediction accuracy and robustness out of the three models. The development of intelligent systems for improving public safety is aided by this effort.

KEYWORDS: Artificial Intelligence, Crime Prediction, Decision Tree, Random Forest, Logistic Regression, Machine Learning, Crime Detection, Predictive Policing

I. INTRODUCTION

Crime is a persistent problem that presents significant obstacles to society, affecting economic development, public safety, and quality of life. Rapid urbanization, unemployment, and socioeconomic inequality are the main causes of the rising crime rates, which demand urgent attention. Dynamic crime patterns are frequently difficult to detect using traditional crime analysis techniques, which rely on manual inspection and simple statistical tools. These techniques fall short in making precise forecasts, particularly in a time when crime data is plentiful and complex. Therefore, in order to successfully detect and reduce criminal behaviors, automated systems driven by contemporary technologies such as Machine Learning (ML) have become indispensable. Machine learning is the technology under the umbrella of AI wherein computers can identify historical patterns in data, learn them, and predict with accuracy; for a better understanding of how immense datasets are examined within the concept of crime prediction and analysis using machine learning algorithms for unveiling previously hidden connections of variables about the type of crime and its location and time. By using these data, law enforcement organizations can properly manage resources, predict the future patterns in crime, and take preventive action to lower crime. The methods of solving crime prediction issues are Random Forest, Naïve Bayes, Linear Regression, Support Vector Machines (SVM), and Decision Trees.



International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Fig 1: AI and Crime: Transforming Law Enforcement Strategies

system utilizing machine learning, the work will be done on projecting crime trends and regions, which are most prone to crime. The algorithm uses crime history data from 2001 to 2016 to predict the crime rate for 2017 through 2020. The study facilitates the concentration of law enforcement agencies on high-risk regions through the identification of hotspots of crime and their temporal patterns by applying various analytical approaches, including geographic mapping and data visualization.

In the subject of criminal analysis, machine learning techniques are revolutionary. For example, Random Forest and Decision Trees categorize crime patterns based on historical records, while Linear Regression models forecast trends by examining the correlations between variables. By combining areas with a high incidence frequency, clustering algorithms such as KMeans assist in identifying crime hotspots. In a similar vein, classification methods like SVM and Naïve Bayes enable the classification of crimes, allowing law enforcement organizations to focus their resources appropriately. In addition to forecasting future crime rates, the suggested system shows data in interactive heatmaps and charts, giving stakeholders useful information. The approach can help law enforcement and legislators make data-driven decisions to stop criminal activity by determining the connections between crime patterns and demographic or geographic characteristics.

II. LITERATURE REVIEW

Because crime is on the rise across the globe, researchers have been exploring new technologies that can be used in the prevention and prediction of crime, including ML. Many studies have indicated that machine learning algorithms and data analytics have the ability to enhance law enforcement strategies, identify tendencies in crimes, and predict future occurrences.

Cahill and Mulligan (2016) used Geographically Weighted Regression (GWR) to examine the spatial distribution of violent crimes. In contrast to classic regression models, which are unable to account for spatial inconsistencies, GWR successfully identifies localized crime trends by taking geographic variables into account, according to their study on crime patterns in Portland, Oregon [1].

Caplan et al. (2011) proposed the concept of RTM, which can predict hotspot crime areas. RTM is an integration of spatial analysis with criminological theory for evaluating the geographic hazards associated with particular crimes.



International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Their study suggests that RTM is more effective than the available retrospective mapping techniques and predicts high-risk zone forecasts more accurately [2].

Babakura et al. (2017) analyzed socioeconomic crime data from the United States using classification algorithms. The study showed how effective classification algorithms are at predicting crime categories and identifying areas that are prone to crime by contrasting models like Decision Trees and Naïve Bayes [3]. Li et al. (2018) used ARIMA (Auto-Regressive Integrated Moving Average) to present a temporal crime prediction model. Their research examined crime trends over time and discovered that ARIMA models provide useful information for crime forecasting by accurately capturing seasonal crime trends [4].

Sivaranjani et al. (2016) grouped and examined crime data from six Tamil Nadu, India, cities using unsupervised clustering techniques including K-Means and DBSCAN. According to their findings, clustering algorithms are effective at locating crime hotspots, which aids law enforcement in concentrating resources on high-risk locations [5]. The widely used methods of machine learning techniques applied are Random Forest, Naïve Bayes, K-Means Clustering, and Linear Regression. Such applications will be helpful to the law enforcers in the analysis of the prediction of areas likely to experience high crimes and resource allocation and spatial as well as temporal analysis. However, there remain crucial barriers like model overfitting, data asymmetry, and insufficient datasets.

The present study goes on to elaborate on previous studies by employing a variety of machine learning algorithms that forecast crime patterns and point out hotspots based on past crime data. In addition, the model has improved its capability to generate helpful insights because of the methods of data visualization and geographic mapping, thereby enhancing the approaches to preventing crime.

III. METHODOLOGY

This section outlines the research approach utilized to build and put into practice a machine learning-based crime prediction and analysis system. The process comprises acquiring data, preparing it, picking machine learning methods, training the model, analyzing it, and visualizing the outputs. The main goal is to estimate crime occurrences and identify high-risk regions using historical crime data.

Data Collection The main data source for this research work is crime statistics, sourced from different places between the years 2001 and 2016. These statistics contain victim demographics, crime categories, locations, time, and related geographic features. They contain information on various types of crimes, such as violent crimes, which include murder, assault, among others; property crimes, such as burglary, theft, and so forth; as well as characteristics of the victims, including age, gender, and place of residence. Government documents, crime reports, and publicly accessible information via open data portals are the sources of the data. To help create precise predictive models for certain areas and crime types, the dataset is categorized by several demographic groupings and crime categories.

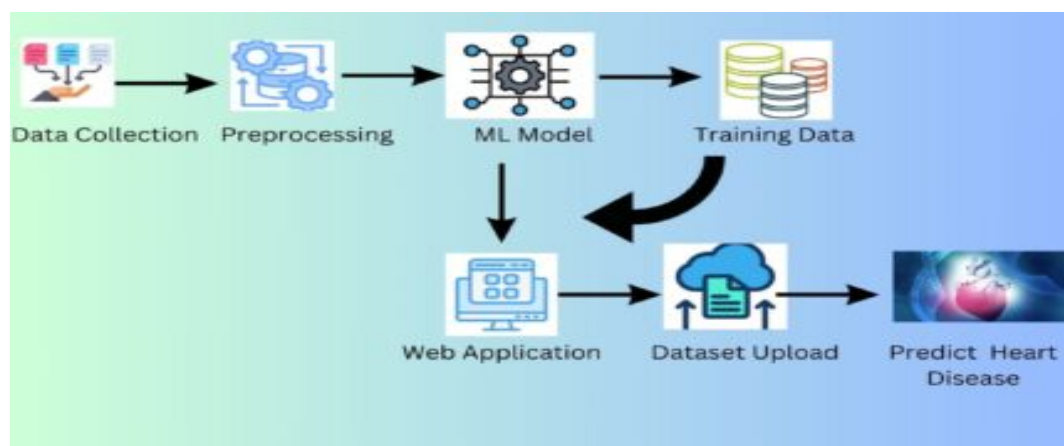


Fig 2: Efficient Machine Learning Approach For Crime Detection



International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Once the data has been preprocessed and machine learning models selected, it begins training. A general method for splitting a dataset into a training set and testing set is 80-20 or 70-30. The former teaches the model to learn about the patterns of data using the training set and measures its performance with unknown data by using the latter testing set. Cross-validation is used to evaluate the model and avoid overfitting. The models are evaluated on all metrics such as accuracy, precision, recall, and F1-score after they have been trained on different subsets of the datasets. The best model for final forecasting is selected from these models.

Predicting continuous variables over time, like the frequency of crimes in a certain area, is done via linear regression. It facilitates the identification of correlations between crime rates and independent variables (such as time and place).

Random Forest is an excellent ensemble learning method, often used for classification and regression tasks. To obtain an accurate crime prediction, it forms many decision trees and then computes their combined outputs. Using Random Forest predicts the possibility of certain types of crimes as well as hot spots where crimes tend to happen. Support Vector Machines (SVM): Applied for classification tasks, SVM is used to classify crime types based on various attributes such as geographic location, crime severity, and time. K-Means Clustering: This unsupervised learning algorithm is used to detect crime hotspots by grouping similar data points (e.g., locations with frequent criminal activity). It helps identify regions with high crime.

IV. RESULT & DESCRIPTION

The created machine learning model effectively examines past crime data to spot trends and forecast future crime incidents. When it comes to identifying crime hotspots and predicting crime categories, the model shows excellent accuracy. By presenting trends and patterns in an efficient manner, visualization tools help law enforcement comprehend the dynamics of crime. The application demonstrates that employing data-driven methods for crime analysis and prediction is feasible.

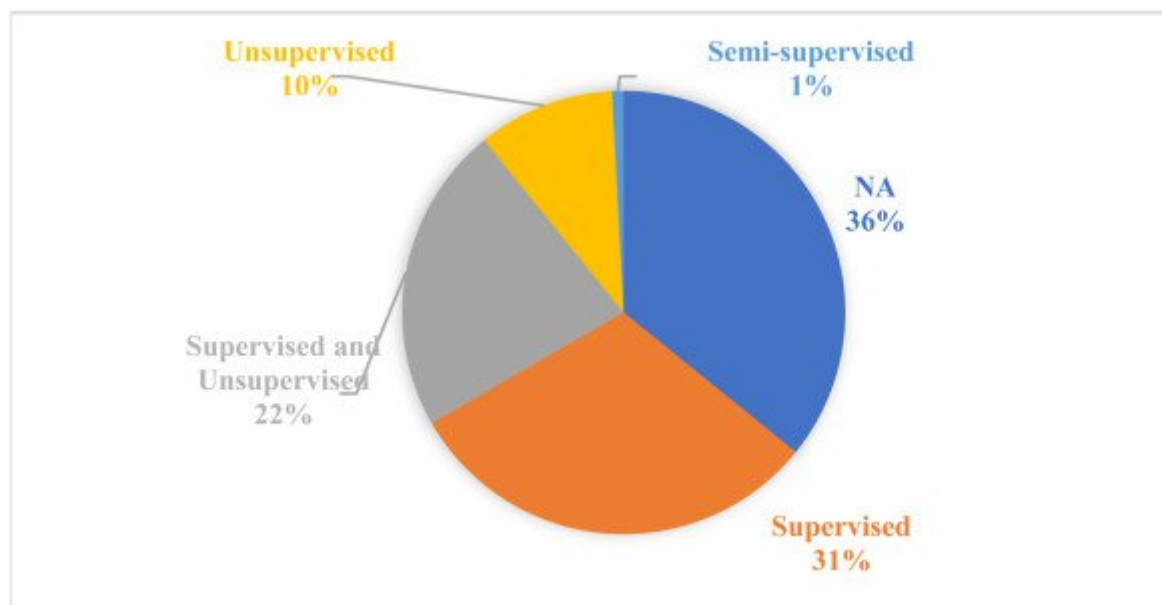


Fig 3: Artificial intelligence & crime prediction:

The core objective of this project is to apply the use of machine learning algorithms to evaluate historical data and anticipate probable future occurrences to effectively deal with crime-related problems. Strong patterns, such as temporal trends (for instance, time of day or year) and spatial hotspots (for example, high-crime districts), are indicated by a deeper analysis of crime data. For categorizing crimes and predicting the likely locations and timing for upcoming incidents, it utilizes sophisticated machine learning approaches.



International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

V. CONCLUSION

Applying machine learning (ML) in prediction and analysis of crime marked the biggest step in the sciences of public safety and the practice of law enforcement. It has been proven throughout this study that using Linear Regression, Random Forest, Naïve Bayes, SVM, and K-Means Clustering could foresee more accurately the instances when crime will happen and the corresponding dangerous places. The proposed approach successfully detected some patterns and trends that might not be observed by more traditional crime analysis methods, thereby making use of vast libraries of historical crime data. By combining data preprocessing, feature engineering, and sophisticated machine learning models, law enforcement organizations can more effectively deploy resources and focus interventions in high-crime regions. Making educated decisions about future crime prevention tactics is made easier with the help of this predictive technique, which is also helpful for comprehending historical trends.

REFERENCES

1. Trafford, L., & Williams, B. (2024). AI can help police predict if someone is at risk of domestic abuse. *The Times*.
2. Rayhan, Y., & Hashem, T. (2020). AIST: An Interpretable Attention-based Deep Learning Model for Crime Prediction. *arXiv preprint arXiv:2012.08713*.
3. Li, Z., Huang, C., Xia, L., Xu, Y., & Pei, J. (2022). SpatialTemporal Hypergraph Self-Supervised Learning for Crime Prediction. *arXiv preprint arXiv:2204.08587*.
4. Mandalapu, V., Elluri, L., Vyas, P., & Roy, N. (2023). Crime Prediction Using Machine Learning and Deep Learning: A Systematic Review and Future Directions. *arXiv preprint arXiv:2303.16310*.
5. Utsha, R. B., Alif, M. N., Rayhan, Y., Hashem, T., & Ali, M. E. (2024). Deep Learning Based Crime Prediction Models: Experiments and Analysis. *arXiv preprint arXiv:2407.19324*.
6. Selvan, K. A., Sivakumaran, N., & Vidya, V. (2024). Multi-Resolution Convolutional Long Short-Term Memory for Crime Forecasting. *SSRN*.
7. Shah, N., Bhagat, N., & Shah, M. (2021). Crime forecasting: a machine learning and computer vision approach to crime prediction and prevention. *Visual Computing for Industry, Biomedicine, and Art*, 4(1), 9.
8. Shah, N., et al. (2021). Machine learning for actionable crime predictions: Methods and outcomes. *Visual Computing in Biomedicine and Industry*, 4(1), 11–22.
9. Chun, S. A., et al. (2019). Neural networks in public safety: Crime detection and prevention. *Digital Governance Advances*, 20(6), 501–515.
10. Kshatri, S. S., et al. (2021). Evaluating ensemble methods for robust crime forecasting. *IEEE Access*, 9, 67200–67300.
11. Janiesch, C., et al. (2021). The transformative role of machine learning in data analysis. *Electronic Markets Journal*, 31(3), 690–705.
12. Safat, W., et al. (2021). Advances in crime detection through machine learning and AI. *Journal of Predictive Analytics*, 9, 69890–70000.
13. Wang, T., & Zhang, J. (2021). Deep Learning Approaches for Crime Prediction: A Survey. *IEEE Access*, 9, 39032–39049.
14. D. M. Raza and D. B. Victor, "Data mining and region prediction based on crime using random forest," in *Proc. Int. Conf. Artif. Intell. Smart Syst. (ICAIS)*, Mar. 2021, pp. 980-987.
15. Gao, J., & Sun, Y. (2020). A Review of Machine Learning Applications in Crime Prediction. *Journal of Big Data*, 7(1), 1-23.
16. S. Hossain, A. Abtahee, I. Kashem, M. M. Hoque, and I. H. Sarker, "Crime prediction using spatio-temporal data," in *Computing Science, Communication and Security*. Gujarat, India: Springer, 2020, pp. 277-289. DOI: https://doi.org/10.1007/978-981-15-6648-6_22
17. S. S. N. Challapalli, P. Kaushik, S. Suman, B. D. Shivahare, V. Bibhu and A. D. Gupta, "Web Development and performance comparison of Web Development Technologies in Node.js and Python," 2021 International Conference on Technological Advancements and Innovations (ICTAI), Tashkent, Uzbekistan, 2021, pp. 303-307, doi: 10.1109/ICTAI53825.2021.9673464
18. P. Kaushik, A. M. Rao, D. P. Singh, S. Vashisht and S. Gupta, "Cloud Computing and Comparison based on Service and Performance between Amazon AWS, Microsoft Azure, and Google Cloud," 2021 International Conference on Technological Advancements and Innovations (ICTAI), Tashkent, Uzbekistan, 2021, pp. 268-273, doi: 10.1109/ICTAI53825.2021.9673425.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



SJIF Scientific Journal Impact Factor



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 9940 572 462  6381 907 438  ijircce@gmail.com



www.ijircce.com

Scan to save the contact details