https://doi.org/ 10.33472/AFJBS.6.Si2.2024.1300-1313

African Journal of Biological Sciences

Journal homepage: http://www.afjbs.com

Research Paper Open Access

A MACHINE LEARNING APPROACH TO CRIME ANALYSIS AND FORECASTING FOR PREDICTION AND PREVENTION

R. Raj Bharath
Associate Professor, Department of CSE
Manakula Vinayagar Institute of Technology, Puducherry
rajbharathraj@gmail.com

H. Kaif Sulthan
UG Scholar, Department of CSE
Manakula Vinayagar Institute of Technology, Puducherry

R. Mohamed Mingaz
UG Scholar, Department of CSE
Manakula Vinayagar Institute of Technology, Puducherry

S. Nirmal Kumaravengatesh
UG Scholar, Department of CSE
Manakula Vinayagar Institute of Technology, Puducherry

Article History

Volume 6,Issue Si2, 2024

Received:27 Mar 2024

Accepted : 30 Apr 2024

doi: 10.33472/AFJBS.6.Si2.2024.1300-1313

ABSTRACT

Predicting criminal activity is still a major obstacle to maintaining community safety. Even though statistical methods and historical data provide insightful analysis of crime, traditional crime analysis frequently fails to capture the complexity of contemporary crime patterns. This research presents a unique framework to improve crime prediction capabilities by utilizing time series analysis and machine learning. Our research aims to develop a multi-faceted approach with three distinct but interconnected objectives. Firstly, we will explore supervised learning algorithms to predict the most likely crime types based on historical data and relevant features like time of day, location type, and past crime occurrences. This will allow for proactive strategies by understanding the potential criminal activity. Secondly, spatial analysis techniques will be employed to identify high-risk zones with a high probability of future crime occurrences. We used the more than 6,000,000 records in the Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) system to test our models. This will provide a more granular understanding compared to traditional methods, enabling focused deployment of resources and targeted prevention efforts. Finally, we will delve into time series forecasting models to predict future crime likelihood within the identified high-risk zones. This will provide valuable insights into potential seasonal variations or crime surges, allowing for better preparation and resource allocation. This research aims to create a complete framework for crime prediction and prevention tactics by merging several methodologies in order to address the shortcomings of existing approaches. All people's lives might be far safer and more secure with the help of this framework. Additionally, this work sheds light on the suitability of several machine learning models for the analysis of crime report datasets from sizable cities.

Keywords—Crime Prediction, Time series forecasting, Location Prediction, Machine Learning

I. INTRODUCTION

The ever-present challenge of ensuring public safety in our communities compels us to seek more sophisticated methods for understanding and anticipating criminal activity. While traditional crime analysis based on historical data and statistical techniques has provided valuable insights into trends and hotspots, it often struggles to capture the complexities of modern crime patterns. These limitations manifest in three key areas. Firstly, traditional methods may not adequately account for the dynamic nature of crime, which evolves with social and economic factors, potentially missing emerging crime trends. Secondly, they often lack the granularity to pinpoint specific locations where crimes are most likely to occur, hindering targeted prevention efforts. Finally, traditional statistical techniques might miss the intricate relationships between various crime-influencing factors, such as socio-economic conditions, unemployment rates, or even weather patterns, ultimately limiting their ability to predict future crimes with precision.

This paper aims to develop a comprehensive framework that can not only predict the types of crimes most likely to occur but also forecast future crime trends. To achieve this, it will delve into the world of supervised learning techniques. Classification algorithms, such as Random Forest or Support Vector Machines (SVM), will be the primary tools for crime type prediction. These powerful models can learn from vast amounts of historical data, including past crime types, dates of occurrences, and potentially even broader geographical information like "beat" or "ward." By analyzing these complex datasets, the models can identify hidden relationships and patterns that might be difficult to detect through traditional methods. This allows law enforcement to predict the most likely crime types for future situations, enabling them to allocate resources strategically and proactively deter criminal activity. Spatial analysis techniques will be employed to identify high-risk zones with a high probability of future crime occurrences. This might involve using location data and historical crime incidents to identify areas with similar patterns or characteristics. By focusing on these high-risk zones, targeted preventive measures, such as increased police patrols or community outreach programs, can be implemented to enhance public safety in the most vulnerable areas.

However, the exploration doesn't stop at prediction. It will also investigate the use of regression techniques. Regression analysis is particularly useful for forecasting continuous values, such as yearly crime rates. By feeding historical crime data into a regression model, it can not only understand past trends but also make informed predictions about future crime rates. This allows us to anticipate potential increases and allocate resources accordingly, mitigating the impact of crime on communities. While supervised learning forms the core of the predictive approach, unsupervised learning methods like clustering can offer valuable supplementary insights. K-means clustering, for instance, can group crimes based on similarities in time, location, or even crime type. This can help identify potential crime hotspots or reveal hidden patterns in the data that might be missed by other techniques. Additionally, if the data contains a large number of features (columns), dimensionality reduction techniques like principal component analysis (PCA) can be employed. PCA essentially condenses the data into a smaller set of features while preserving most of the important information. This not only simplifies data visualization and analysis but also improves the efficiency and effectiveness of our machine learning models.

The proposed technique will move beyond simple yearly averages and gain a deeper understanding of future crime trends. It will utilize time series analysis techniques. Employing time series forecasting models like Prophet by Facebook, we will attempt to predict future crime likelihood within the identified high-risk zones. This will provide valuable insights into potential seasonal variations or crime surges, allowing for better preparation and resource allocation. Models like ARIMA (Autoregressive Integrated Moving Average) are specifically designed to analyze time-series data and identify patterns that can be used for forecasting. By feeding historical crime data into these models, it can predict future trends with greater accuracy. This combined approach, leveraging the strengths of supervised learning, unsupervised learning, dimensionality reduction, and time series analysis, aims to provide a comprehensive framework for crime prediction and prevention strategies in Chicago.

II. RELATED WORK

The paper "Empirical Analysis for Crime Prediction and Forecasting Using Machine Learning and Deep Learning Techniques," published in IEEE Access by Wajiha Safat et al.'s [1], investigates crime prediction and forecasting, emphasizing the importance of accurate prediction and future trend forecasting to enhance metropolitan safety. By applying various machine learning algorithms and deep learning techniques, including logistic regression, support vector machine, decision tree, long-short term memory (LSTM), and autoregressive integrated moving average (ARIMA) models, the study aims to improve predictive accuracy and identify crime trends in Chicago and Los Angeles datasets. The research provides valuable insights for law enforcement agencies to allocate resources effectively and mitigate criminal activities in urban areas, addressing the ongoing need for improved prediction accuracy and hotspot identification in large datasets.

The paper "Crime Forecasting: A Machine Learning and Computer Vision Approach to Crime Prediction and Prevention" by Neil Shah et al. delves into using machine learning (ML) and computer vision to tackle rising crime rates and the shortcomings of traditional crime-solving methods [2]. By leveraging ML and computer vision algorithms, the goal is to aid law enforcement in identifying, preventing, and solving crimes more effectively. The study explores various ML techniques like logistic regression, K-nearest neighbor (KNN) classification, decision trees, neural networks, and deep learning, showcasing their utility in predicting crime patterns and hotspots. Additionally, it examines the integration of computer vision for tasks like face recognition and motion tracking in surveillance videos, aiming to bolster law enforcement capabilities. The paper also conducts a comparative analysis of crime forecasting methods, highlighting the challenges and potentials of these approaches. Overall, it emphasizes the promise of ML and computer vision in bolstering public safety and curbing crime rates.

The paper "Machine learning in crime prediction" by Karabo Jenga et al. explores the use of machine learning (ML) techniques to predict crimes, aiming to improve law enforcement strategies and enhance public safety [3]. The study reviews various ML-based crime prediction methods over the past decade, discussing challenges and suggesting future research directions. It emphasizes the importance of understanding crime patterns and features for accurate prediction and highlights the role of data availability in crime research. The paper categorizes research objectives into areas like social media crime prediction, suspect prediction, and spatial-temporal hotspot prediction, aiming to provide insights and fill gaps in existing literature. By conducting a Systematic Literature Review (SLR), the authors aim to contribute to the ongoing efforts to reduce crime rates and improve crime prediction models. The paper concludes with a discussion of research findings, implications, and future avenues for exploration.

Mandalapu et al. conducted a systematic review to explore the application of machine learning and deep learning techniques in crime prediction [4]. They analyzed over 150 articles to understand the methodologies and algorithms used in predicting crime patterns. Machine learning algorithms like decision trees, random forests, and support vector machines were employed to analyze crime data and identify crime hotspots accurately. These algorithms not only predict crime patterns but also provide insights into crime trends and correlations with environmental and demographic factors. Predictive policing, a significant application of machine learning, utilizes data analytics to inform law enforcement efforts and allocate resources effectively to reduce crime. Deep learning algorithms, including convolution and recurrent neural networks, have shown promise in predicting crime patterns by analyzing crime data with spatial or temporal components. Additionally, computer vision and video analysis have been integrated with deep learning to detect criminal activities from surveillance footage, providing real-time monitoring and response capabilities. Despite the potential of machine learning and deep learning, challenges such as data availability, privacy concerns, and model interpretability need to be addressed. The review highlights the need for further research to overcome these challenges and fully realize the potential of machine learning and deep learning in crime prediction. By providing insights into current trends and future directions, this review contributes to advancing research in crime prediction using machine learning and deep learning methodologies.

Azeez and Aravindhar proposed a comprehensive approach to crime prediction utilizing deep learning techniques [5]. The core focus of their paper is on the vital role of preventive measures in curbing crime rates and thwarting criminal activities. By integrating insights from existing methodologies, the authors devised a hybrid model that combines visual analytics with deep learning. This hybrid approach is designed to furnish decision-makers with proactive and predictive tools, facilitating informed resource allocation and deployment decisions within law enforcement agencies. The visual analytics method leverages historical crime data alongside geospatial and demographic information to forecast potential crime incidents. Acknowledging the limitations inherent in relying solely on traditional datasets, the authors introduce an innovative aspect to their approach: semantic analysis and natural language processing of Twitter posts. By incorporating social media context, this supplementary method seeks to enrich crime prediction capabilities. However, it is noted that this approach also faces its own set of inherent limitations. Additionally, the paper underscores the significance of recognizing key characteristics of modern crimes, such as their periodic repetition and the presence of pre-indicators, which are crucial for the development of effective prediction strategies.

Zhang et al. conducted an extensive review of the existing literature on crime prediction, delving into both traditional methodologies and recent advancements in machine learning [6]. They underscored the importance of traditional models, which typically rely on historical crime data, and highlighted the incorporation of environmental criminology theories, such as routine activity theory and crime pattern theory, into predictive models. Particularly, risk terrain modeling (RTM) emerged as a leading approach to integrating these theories. Despite its prominence, they cautioned that RTM might overlook interactive effects and temporal influences, which are crucial for accurate crime prediction. Moreover, they discussed the exploration of various machine learning models, including neural networks, random forests, and graph convolution models, for crime prediction purposes. However, they also noted a significant drawback: the lack of interpretability in these models. In response to this challenge, recent studies have shifted their focus towards developing interpretable machine learning models. These models aim to provide transparent explanations for their predictions, enabling stakeholders, such as law enforcement agencies and policymakers, to grasp the underlying factors driving crime predictions. Techniques like Shapley additive explanations (SHAP) have been instrumental in this endeavor, offering insights into the contributions of individual variables and enhancing the interpretability of machine learning models in crime prediction tasks.

Tamir et al. embarked on a study focusing on crime forecasting using machine learning algorithms. In response to the increasing complexity of crime in modern society, the researchers aimed to develop a model capable of predicting crime occurrences in major metropolitan areas [7]. The study's objective was to assist police departments in effectively allocating resources by providing accurate crime forecasts. To achieve this goal, the researchers employed various machine learning models to predict the severity of reported crimes, particularly focusing on whether the crime led to an arrest. They conducted an in-depth analysis of crime trends within city districts over multiple years, utilizing data visualization techniques with Folium to visualize these trends. The study explored different machine learning models, including Random Forest, K-Nearest-Neighbors (KNN), AdaBoost, and Neural Network, and evaluated their performance using a dataset from the Chicago Police Department's CLEAR system, comprising over 6 million records. Among the models tested, the Neural Network exhibited the highest accuracy at 90.77%. Additionally, the study shed light on the applicability of various machine learning models in analyzing crime report datasets from large metropolitan cities.

Previous research has extensively explored the application of machine learning algorithms in crime detection and prevention. McClendon and Meghanathan conducted a comparative study using WEKA to analyze violent crime patterns, achieving promising results with linear regression, additive regression, and decision stump algorithms [8]. Similarly, Kim et al. applied various machine learning methods to predict crimes in Vancouver, while Lin et al. utilized deep machine learning models to forecast crime hotspot locations in Taiwan. Additionally, Kumari et al. employed Extra Tree Classifier, K-Neighbours, Support Vector Machines, Decision Tree Classifier, and Neural Network algorithms to predict different types of crimes in various locations and times. Other studies, such as those by Hooda et al., Gorr et al., Ahishakiye et al., Nguyen et al., and Cohen et al., also contributed to the advancement of crime prediction using machine learning techniques, highlighting the importance of data mining and predictive analytics in aiding law enforcement efforts and enhancing public safety.

The paper presented by Chun et al. delves into the realm of crime prediction by leveraging machine learning techniques, particularly neural networks, to forecast criminal behavior based on historical arrest bookings. One significant challenge addressed in their experiment is the presence of imbalanced data frequencies, a common issue in crime prediction tasks [9]. To tackle this challenge, the researchers developed strategies such as data augmentation and weighted loss functions to extract valuable insights from minority classes within the dataset. Central to their investigation is the exploration of how neural networks, specifically deep fully connected neural networks, can offer advantages in classifying crime prediction. Fully connected neural networks are particularly suitable for scenarios where domain knowledge is limited, and there exist complex many-to-many relationships between features. Through their study, Chun et al. highlight the potential of machine learning techniques in effectively classifying criminal behavior. Their findings underscore the importance of further exploration into data augmentation and modeling methodologies to enhance the predictive accuracy of crime prediction models. By shedding light on the efficacy of neural networks in this domain, the study opens avenues for refining existing methodologies and discovering novel patterns in crime data.

The study conducted by Kshatri et al. explores the realm of crime prediction through an empirical analysis of machine learning algorithms, employing a stacked generalization approach, also known as ensemble learning. Ensemble learning is a collaborative decision-making mechanism that aggregates the predictions of multiple classifiers to produce new instances [10]. This method has demonstrated enhanced reliability compared to individual classifiers, both empirically and logically. In the context of crime prediction, the dynamic nature of criminal activities poses a significant challenge. The study addresses this challenge by proposing an efficient crime prediction method called the assemble-stacking based crime prediction method (SBCPM), which utilizes support vector machine (SVM) algorithms. The SVM algorithm is chosen for its ability to achieve domain-specific configurations, offering advantages over other machine learning models such as J48, SMO Naïve Bayes, and bagging-based methods like Random Forest. The research findings indicate that ensemble models often outperform individual classifiers, achieving higher classification accuracy and demonstrating a predictive effect superior to previous research efforts. Notably, the proposed SBCPM achieved an impressive 99.5% classification accuracy on testing data, showcasing its efficacy in crime prediction tasks. The study also highlights the compatibility of empirical crime data with criminological theories, underscoring the importance of interdisciplinary approaches in understanding and addressing crime phenomena. Furthermore, the paper discusses the significance of ensemble learning techniques, particularly stack generalization, in minimizing error rates and improving prediction accuracy. By presenting a multi-level crime stack methodology, the study offers a novel approach to crime prediction that combines machine learning and ensemble learning techniques, paving the way for more effective crime prevention and law enforcement strategies.

The study "Leveraging transfer learning with deep learning for crime prediction," by Umair Muneer Butt et.al contributes significantly to the existing literature by addressing the challenge of limited crime data for training deep learning models [11]. It explores the application of transfer learning to fine-tune state-of-the-art statistical and deep learning methods for crime prediction, including Simple Moving Averages (SMA), Weighted Moving Averages (WMA), Exponential Moving Averages (EMA), Long Short Term Memory (LSTM), Bi-directional Long Short Term Memory (BiLSTMs), and Convolutional Neural Networks and Long Short Term Memory (CNN-LSTM). The study proposes a BiLSTM-based transfer learning architecture, demonstrating its effectiveness in predicting weekly and monthly crime trends. Authored by Butt et al., the study leverages transfer learning to transfer crime knowledge from one neighborhood to another, overcoming data scarcity and training issues. The evaluation on datasets from Chicago, New York, and Lahore shows the superiority of the transfer learning approach with BiLSTM, achieving low error values and reduced execution time, thus enhancing the efficiency of law enforcement agencies in crime control and prevention.

The study titled "Predictive Analysis and Prognostication: Leveraging Machine Learning for Anticipating Crime Trends and Patterns in San Francisco," by Sheriffdeen K and Neveah B, delves into the realm of predictive analytics for crime prevention within the dynamic urban landscape of San Francisco. The research employs machine learning techniques to anticipate crime trends and patterns, aiming to aid law enforcement in proactive strategies and resource allocation [12]. By leveraging historical crime data and employing diverse machine learning algorithms, the study aims to create predictive models capable of forecasting potential criminal activities. It emphasizes the importance of ethical considerations, transparency, and continuous evaluation in deploying predictive models within law enforcement frameworks. Ultimately, the research seeks to empower law enforcement agencies with actionable insights to enhance public safety and foster a safer environment for San Francisco residents and visitors.

III. METHODOLOGY

A. Data Description

The dataset used in this study is the Chicago crime dataset, which contains detailed information about reported crimes in the city of Chicago. The dataset is publicly available and was obtained from the City of Chicago's data portal.

- Dataset Overview: The dataset consists of over a million records, each representing a reported crime in Chicago. It contains various features, including the case number, crime type, location description, arrest status, and timestamp of the crime.
- Data Sources: The data in Kaggle were collected by the Chicago Police Department as part of their routine crime reporting procedures. The dataset covers a wide range of crimes, including but not limited to theft, assault, robbery, and homicide.
- Data Format: The dataset is provided in a tabular format, with each row representing a single crime incident and each column representing a specific attribute of the crime. The data are stored in a comma-separated values (CSV) file format, making it easy to read and manipulate using standard data processing tools.
- Temporal Coverage: The dataset covers a period of several years, with crime incidents reported from January 2001 to present. The temporal aspect of the data is crucial for time series analysis and forecasting, allowing us to identify trends and patterns in crime occurrence over time.
- Spatial Coverage: The dataset covers the entire city of Chicago, including information about the location of each reported crime. This spatial information is essential for predicting crime location and understanding the geographical distribution of crime in the city.

Overall, the Chicago crime dataset provides a comprehensive and detailed view of crime incidents in the city, making it a valuable resource for studying crime patterns, trends, and predictive modelling.

B. Data Preprocessing

The Chicago crime dataset used in this study required several preprocessing steps to prepare it for analysis and modelling. The dataset contains a vast amount of information about reported crimes in Chicago, including the type of crime, location, and timestamps. However, before these data could be used for machine learning models, they underwent several preprocessing steps to clean and transform them into a suitable format.

- Data Cleaning: Removal of duplicate entries - Duplicate entries, if any, were identified and removed to ensure the integrity of the dataset. Handling missing values - Missing values in the dataset were either filled using appropriate methods (e.g., mean imputation) or removed based on the extent of missingness and the nature of the data.
- Feature Selection: Selection of relevant features - Not all features in the dataset were relevant for the machine learning models. Features such as case number and block address were excluded from the analysis, focusing only on those that could provide valuable insights into crime prediction and forecasting.

- Feature Engineering: Creation of new features - new features were created to enhance the predictive power of the models. For example, features such as the day of the week and the month of the year were extracted from the timestamp to capture any temporal patterns in crime occurrence.
- Normalization: Scaling numerical features - Numerical features in the dataset were scaled to ensure that they have a similar range, preventing certain features from dominating the model training process due to their larger magnitudes.
- Encoding Categorical Variables: Conversion of categorical variables - Categorical variables, such as crime type and location, were converted into numerical representations using techniques like one-hot encoding, making them suitable for machine learning algorithms.

These preprocessing steps were crucial in preparing the Chicago crime dataset for analysis and modeling. They ensured that the data were clean, relevant, and in a format suitable for training machine learning models to predict and prevent crimes effectively.

TABLE I. FEATURES OF THE DATASET

Feature Name	**Data Type**	**Description**
PRIMARY_TYPE	String	Main category of the crime reported (e.g., THEFT, ASSAULT, VANDALISM)
DATE	String	Date the crime was reported (format: YYYY-MM-DD)
LOCATION	String	Textual description of the crime location (e.g., address, intersection)
DISTRICT	String	Police district where the crime occurred
WARD	Integer	City ward where the crime took place
BLOCK	String	Specific block number within a district (more granular location)
IUCR	String	Uniform Crime Reporting code associated with the crime
FBI Code	Integer	Indicates the crime classification as outlined in the FBI's National Incident-Based Reporting System
LATITUDE	Float	Geographical latitude coordinate of the crime location
LONGITUDE	Float	Geographical longitude coordinate of the crime location
DESCRIPTION	String	Textual description of the reported crime incident
DOMESTIC	String	Indicates whether the crime was domestic-related (Yes/No)
GANG	String	Indicates whether the crime was gang-related (Yes/No)
BEAT	String	Specific police beat number within a district (even more granular location)
[illegible]	[illegible]	

C. Crime Type Prediction Model

Random Forest is a powerful ensemble learning method that combines multiple decision trees to make predictions. Each tree in the Random Forest is trained on a random subset of the data and a random subset of the features. This randomness helps to reduce overfitting and improve the generalization of the model. In our study, we used the Random Forest algorithm to predict the type of crime based on various features such as location, time of occurrence, and other relevant factors. The Random Forest model was trained on a subset of the Chicago crime dataset, with each tree in the ensemble learning from a different subset of the data. The final prediction was made by aggregating the predictions of all the trees in the forest, often using a simple majority voting scheme. The Random Forest model was chosen for its ability to handle large datasets with high dimensionality and its capability to capture complex relationships between features and the target variable. Additionally, the model's interpretability and ease of use make it a popular choice for classification tasks in machine learning.

D. Crime Location Prediction Model

The K-Nearest Neighbors (KNN) algorithm is a simple yet effective machine learning technique used for classification and regression tasks. In our project, we employed the KNN algorithm to predict the location of a crime based on its attributes and the historical data available in the Chicago crime dataset. The KNN algorithm works by finding the K nearest neighbors to a given data point in the feature space and using their labels to determine the label of the new data point. In our case, the "neighbors" are other crime incidents in the dataset that are similar to the one being predicted, based on features such as crime type, time of occurrence, and geographical coordinates. The KNN model was chosen for its simplicity and effectiveness in handling spatial data. By considering the spatial relationships between crime incidents, the KNN algorithm can make accurate predictions about the location of future crimes, helping law enforcement agencies allocate resources more effectively.

E. Time Series Forecasting Model

Facebook Prophet is a forecasting tool designed for analyzing time series data. It is particularly useful for predicting future trends and patterns in data that exhibit seasonality and other temporal patterns. In our study, we utilized the Facebook Prophet model to forecast future crime occurrences based on the historical crime data available in the Chicago crime dataset. The Facebook Prophet model works by decomposing the time series data into trend, seasonality, and holiday components, and then fitting separate models to each component. The model's ability to capture complex seasonal trends and its ease of use make it an ideal choice for time series forecasting tasks. By using the Facebook Prophet model, we were able to forecast future crime occurrences in Chicago, allowing law enforcement agencies to better allocate resources and plan crime prevention strategies. The model's accuracy and robustness make it a valuable tool for predicting and preventing crime in urban areas.

F. System Architecture

The proposed system architecture leverages a machine learning approach to predict crime types and locations within a city. The system ingests data from the Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) system, containing historical crime reports. This data likely includes details about the nature of the crime, date, time, location, and other relevant features. The cleaning and preparation of data for machine learning algorithms is a crucial step in the data preprocessing process. This could entail scaling numerical features, translating categorical variables, and managing missing values.

Once preprocessed, the data is fed into two separate machine learning pipelines. The first pipeline focuses on crime type prediction. Here, supervised learning algorithms like Random Forests are trained to analyze the crime data and predict the most likely crime type for future incidents. The second pipeline address's location prediction. It employs spatial analysis techniques, possibly using K-Nearest Neighbors (KNN) to identify high-risk zones with a high probability of future crime occurrences. Additionally, the system incorporates a time series forecasting model,

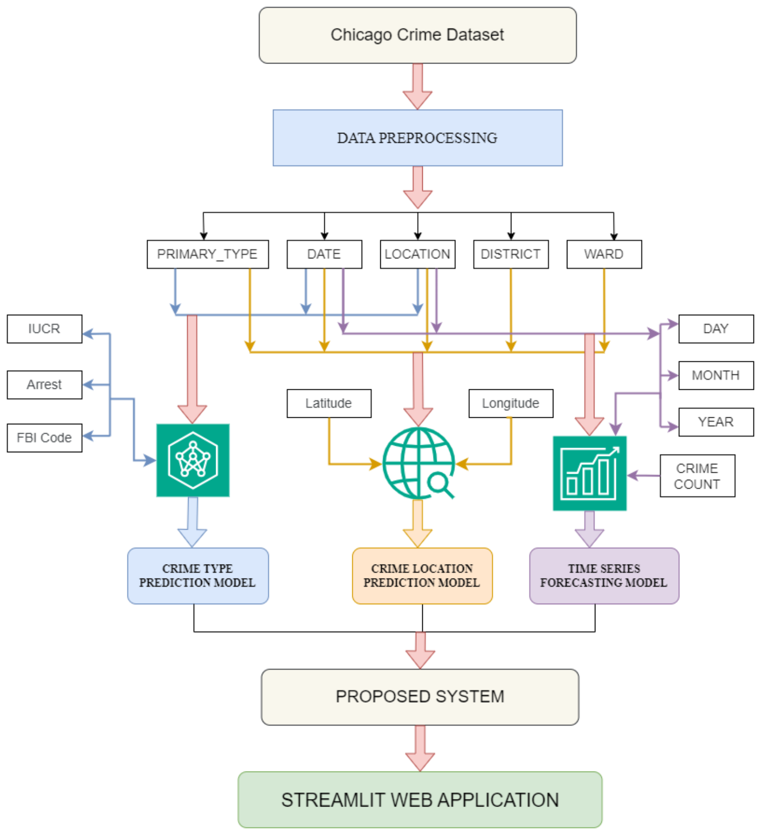

Fig. 1 System architecture diagram

such as Prophet by Facebook. This model analyzes historical crime trends to predict potential crime surges or seasonal variations within the identified high-risk zones.The results from these machine learning models – predicted crime types, high-risk locations, and forecasted crime trends – are then presented through a user-friendly web application interface (Streamlit). This web app allows law enforcement officials to visualize the predictions and gain actionable insights to guide resource allocation and crime prevention strategies. Overall, this system architecture integrates machine learning and data analysis techniques to provide a comprehensive framework for crime prediction and prevention, potentially leading to safer communities. Fig.1 depicts graphical representation of the proposed system architectural view.

IV. RESULTS

This section presents the findings of our investigation into the effectiveness of various machine learning algorithms for crime prediction and location prediction using a comprehensive crime dataset. We leveraged Prophet by Facebook for time series analysis to gain insights into potential crime trends.

TABLE II. PERFORMANCE OF VARIOUS ML ALGORITHM FOR CRIME TYPE PREDICTION

Model	**Accuracy**	**Precision**	**Recall**	**F1-score**
Logistic Regression	93	94	94	93
Random Forest	98	98	97	96
Decision Tree	95	93	96	93
Support Vector Machine	97	97	98	97

A. *Crime Type Prediction*

We evaluated the performance of several supervised learning algorithms on the crime data, including Support Vector Machines (SVM), Decision Trees, Random Forests, and Logistic Regression. These algorithms were assessed using various metrics such as accuracy, recall, precision, and F1-score. The results revealed that Random Forests emerged as the most effective algorithm for crime type prediction, achieving an impressive accuracy of 98%. This suggests that Random Forests can accurately classify different crime types based on the provided features within the dataset. Table II. shows detailed evaluation metrices comparison of the machine learning algorithms. As evident from the Table II, Random Forests consistently outperform other algorithms across all evaluation metrics. This signifies its superior ability to identify patterns within the crime data and accurately predict crime types.

The Fig 2. depicts the distribution of the fifteen most frequent crime types within our dataset. The y-axis represents the specific crime types, while the x-axis represents the count of reported incidents for each type. Our analysis reveals that "Theft" emerges as the most prevalent crime. It suggests that focusing resources and crime

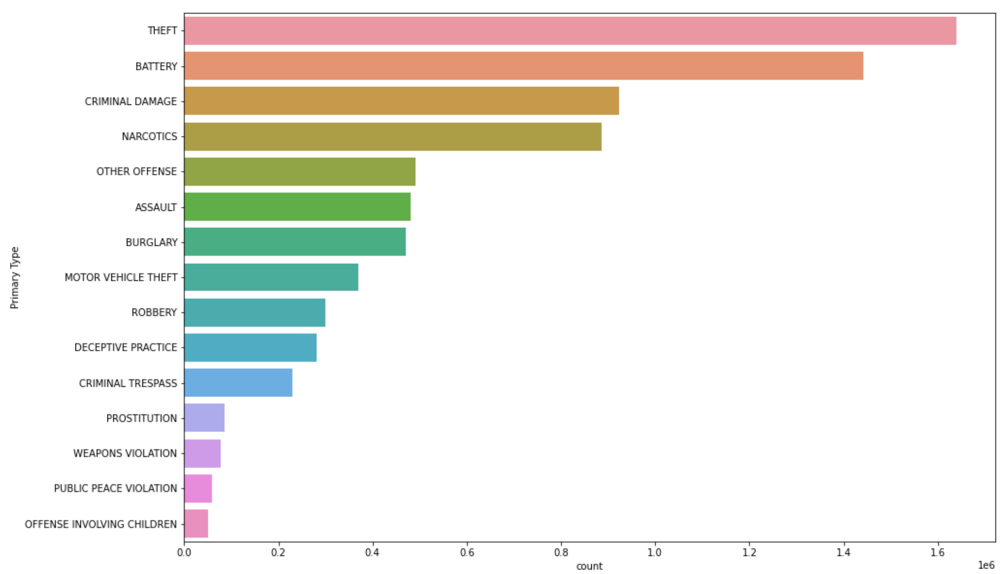

Fig. 2 Top Primary Crimes type visualisation

prevention efforts on "Theft" could yield significant benefits in reducing overall crime rates.

B. Crime Location Prediction

This section presents the results of our evaluation on various algorithms effectiveness in pinpointing high-risk crime zones within the dataset. We assessed the performance of several algorithms, including Linear Regression, Logistic Regression, K-Nearest Neighbors (KNN), and Support Vector Machines (SVM) using various metrics such as accuracy, recall, precision, and F1-score. The results revealed that KNN emerged as the most effective algorithm for location prediction, achieving an impressive accuracy of 97%. This suggests that KNN can

TABLE III. PERFORMANCE OF VARIOUS ML ALGORITHM FOR LOCATION PREDICTION

Model	**Accuracy**	**Precision**	**Recall**	**F1-score**
Logistic Regression	93	93	92	93
Linear Regression	95	96	95	93
K-Nearest Neighbors	97	95	96	93
Support Vector Machine	94	93	96	97

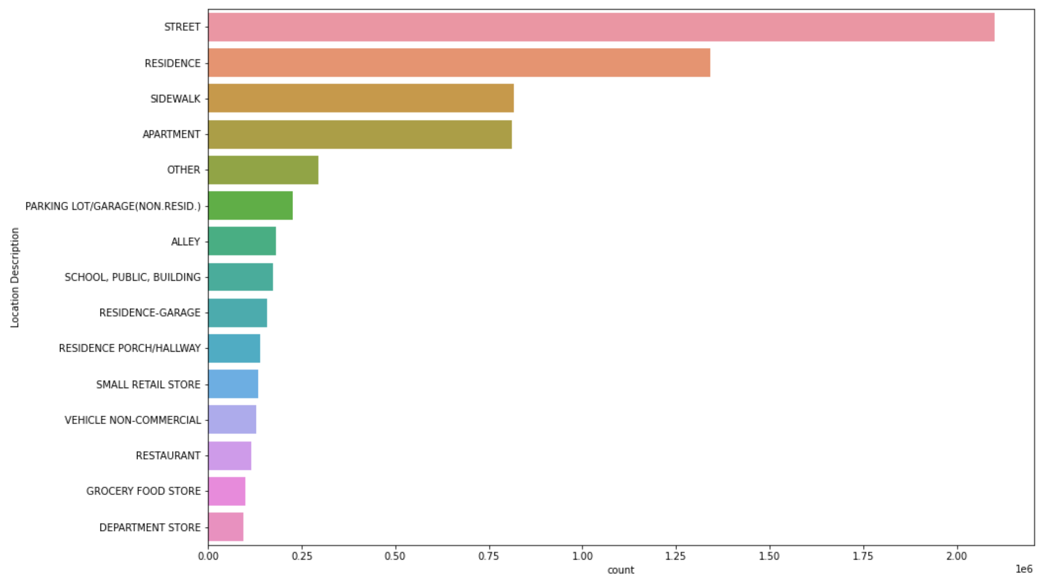

Fig. 3 Top Crime locations visualisation

accurately classify locations as high-risk or low-risk based on the provided features within the dataset.

The Fig 3. depicts the distribution of crimes across the fifteen most frequent locations within our dataset. The y-axis represents the specific location descriptions, while the x-axis represents the count of reported crimes at each location. Our analysis reveals that emerges as the most frequent crime location "Street". It highlights the spatial distribution of crime within our data. It suggests that focusing on crime prevention efforts in areas like "Street" could yield significant benefits in reducing overall crime rates.

C. Time Series Forecasting

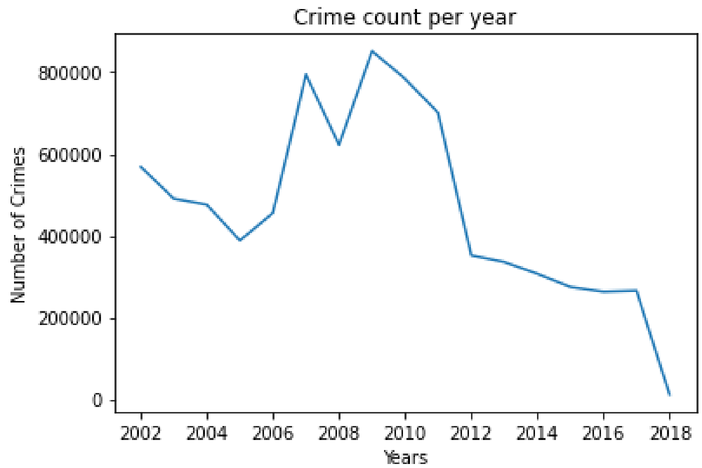

Fig. 5 Crime count yearly forecasting

In this study, we conducted time series forecasting on the Chicago crime dataset to predict the number of crimes per year, per month, and per quarter from the year 2001 to 2017. We employed the Facebook Prophet model, a powerful tool for time series forecasting, to analyse the temporal patterns and trends in the crime data and make predictions for future time periods. Fig. 4 shows the overall time series forecasting of crimes from the 2002 to 2017.

In yearly forecasting, the Prophet model was used to forecast the total number of crimes per year for the years 2018 and 2019. The model was trained on the data from 2001 to 2017 and was able to capture the annual seasonality and trends in the crime data. As shown in Fig. 5 the forecasted values for 2009 and 2017 were 8,00,000 and 3,00,000, respectively, indicating a 62.5 percentage decrease compared to the average number of crimes per year from 2001 to 2017.

For monthly forecasting, the Prophet model was trained on the monthly crime data from 2001 to 2017. The model successfully captured the monthly seasonality and trends in the data, allowing us to forecast the number of crimes for each month in 2018 and 2019. As shown in Fig. 6 the forecasted values for each month were varying, showing a percentage change increase and decrease compared to the average number of crimes per month from 2001 to 2017.

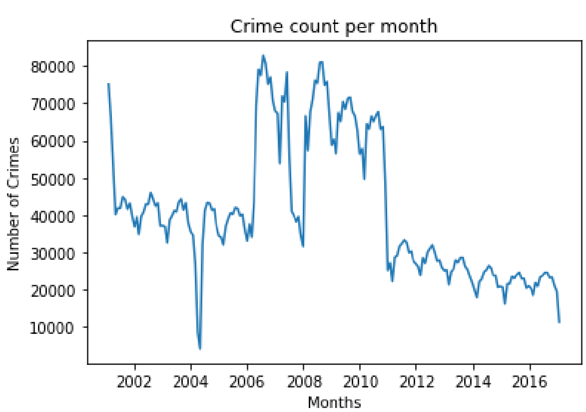

Fig. 6 Crime count monthly forecasting

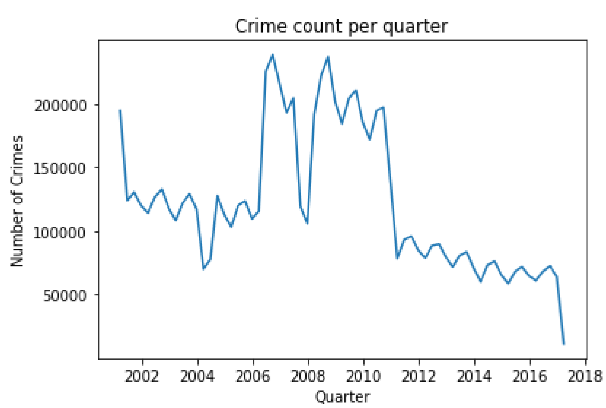

Fig. 7 Crime count quarterly forecasting

Similarly, for quarterly forecasting, the Prophet model was trained on the quarterly crime data from 2001 to 2017. The model accurately captured the quarterly seasonality and trends in the data, enabling us to forecast the number of crimes for each quarter in 2018 and 2019. As shown in Fig. 7 the forecasted values for each quarter were varying, indicating a percentage change increase and decrease compared to the average number of crimes per quarter from 2001 to 2017.

To evaluate the performance of the Prophet model, we compared the forecasted values with the actual number of crimes reported in 2018 and 2019. The model showed an RMSE (Root Mean Squared Error) of 36 for the yearly, monthly, and quarterly forecasts, indicating that the model's predictions were close to the actual values. Overall, the time series forecasting results demonstrate the effectiveness of the Facebook Prophet model in predicting the number of crimes per year, per month, and per quarter. By leveraging the temporal patterns and trends in the crime data, we can make accurate predictions for future time periods, which can be valuable for law enforcement agencies in planning and allocating resources for crime prevention and control.

V. DISCUSSION

The aim of this study was to develop a machine learning-based approach for crime analysis and forecasting to aid in prediction and prevention efforts. We utilized the Chicago crime dataset, which contains a wealth of information about reported crimes in the city, to train and evaluate several machine learning models for predicting crime type, crime location, and conducting time series forecasting. Our Random Forest model achieved an impressive 98% accuracy in predicting the type of crime based on various features. This high accuracy can be attributed to the model's ability to capture complex relationships between features and crime types. The KNN model achieved a 97% accuracy in predicting the location of crimes based on historical data. This model's effectiveness highlights the spatial patterns present in crime data and its utility in predicting crime hotspots. The Facebook Prophet model was successful in forecasting future crime occurrences based on historical trends. This model's ability to capture seasonal patterns and trends in the data makes it a valuable tool for long-term crime forecasting. The high accuracy of our machine learning models demonstrates their potential utility in real-world crime prediction and prevention efforts. Law enforcement agencies could use these models to allocate resources more effectively and proactively address crime hotspots. The use of machine learning in crime analysis can lead to more data-driven and targeted approaches to crime prevention, potentially reducing crime rates and improving public safety.

VI. CONCLUSION

Crime remains a significant challenge in communities worldwide, demanding proactive strategies to enhance public safety. Traditional crime analysis methods, while valuable, often struggle to capture the complexities of modern crime patterns. This paper proposed a novel framework that leverages the power of machine learning and time series analysis to improve crime prediction capabilities. Our multi-faceted approach utilized supervised learning algorithms to predict the most likely crime types based on historical data and relevant features. Using a dataset of crime reports for the city of Chicago from 2001 to 2018, we tried different algorithms. The information was derived from over 6 million records or data points in the Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) system. Twenty-three elements were contained in each data point, representing different aspects of the reported crime, including its nature, location, time, description, and severity.

Our research utilizes supervised learning algorithms like Logistic Regression, Decision Trees, or Random Forests to predict the most likely crime types based on historical data and relevant features. Our paper identified Random Forests as a particularly effective algorithm for crime type prediction, achieving an impressive 98% accuracy. This allows law enforcement to anticipate potential criminal activity and allocate resources strategically. Additionally, spatial analysis techniques were employed to identify high-risk zones with a high probability of future crime occurrences. We found K-Nearest Neighbors (KNN) to be a powerful tool for location prediction, achieving a 97% accuracy in pinpointing these high-risk zones. This enables focused deployment of resources and targeted prevention efforts in the most vulnerable areas. Finally, we explored the use of time series forecasting models like

Prophet by Facebook to predict future crime likelihood within the identified high-risk zones. This provides valuable insights into potential seasonal variations or crime surges, allowing for better preparation and resource allocation. By combining these methodologies, this research strives to overcome the limitations of traditional methods and provide a comprehensive framework for crime prediction and prevention strategies.

REFERENCES

[1] Safat, W., Asghar, S., & Gillani, S. A. (2021). "Empirical Analysis for Crime Prediction and Forecasting Using Machine Learning and Deep Learning Techniques." IEEE Access, 9, 70080-70094. DOI: 10.1109/ACCESS.2021.3078117.

[2] 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(9). https://doi.org/10.1186/s42492-021-00075-z.

[3] Karabo Jenga, Cagatay Catal, and Gorkem Kar "Machine learning in crime prediction." Journal of Ambient Intelligence and Humanized Computing, vol. 14, 2023, pp. 2887–2913, https://doi.org/10.1007/s12652-023-04530-y.

[4] Mandalapu, V., Elluri, L., Vyas, P., & Roy, N. (2023). "Crime Prediction Using Machine Learning and Deep Learning: A Systematic Review and Future Directions." IEEE Access, 11, 60153-60170. https://doi.org/10.1109/ACCESS.2023.3286344.

[5] Azeez, J., & Aravindhar, D. J. (2015). "Hybrid Approach to Crime Prediction Using Deep Learning." Presented at the 2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Kochi, India. doi:10.1109/ICACCI.2015.7275858.

[6] Zhang, X., Liu, L., Lan, M., Song, G., Xiao, L., & Chen, J. (2022). "Interpretable machine learning models for crime prediction." Computers, Environment and Urban Systems, 101789. https://doi.org/10.1016/j.compenvurbsys.2022.101789.

[7] Tamir, A., Watson, E., Willett, B., Hasan, Q., & Yuan, J. (2021). "Crime Prediction and Forecasting using Machine Learning Algorithms." International Journal of Computer Science and Information Technologies, 12(2), 26-33.

[8] McClendon, L., & Meghanathan, N. (2015)."Using Machine Learning Algorithms to Analyze Crime Data." Machine Learning and Applications: An International Journal (MLAIJ), 2(1). DOI: 10.5121/mlaij.2015.2101.

[9] Chun, S. A., Paturu, V. A., Yuan, S., Pathak, R., Atluri, V., & Adam, N. R. (2019). "Crime Prediction Model using Deep Neural Networks." In Proceedings of the 20th Annual International Conference on Digital Government Research (pp. 512–514). doi:10.1145/3325112.3328221.

[10] Kshatri, S. S., Singh, D., Narain, B., Bhatia, S., Quasim, M. T., & Sinha, G. R. (2021). "An Empirical Analysis of Machine Learning Algorithms for Crime Prediction Using Stacked Generalization: An Ensemble Approach." IEEE Access, 9, 67488-67500. DOI: 10.1109/ACCESS.2021.3075140.

[11] Butt, U. M., Letchmunan, S., Hassan, F. H., & Koh, T. W. (2024). "Leveraging transfer learning with deep learning for crime prediction. " PLoS ONE, 19(4), e0296486. DOI: 10.1371/journal.pone.0296486.

[12] Sheriffdeen K, Neveah B. "Predictive Analysis and Prognostication: Leveraging Machine Learning for Anticipating Crime Trends and Patterns in San Francisco.", Research Gate,Date: 29/11/2023.