JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Crime Rate Prediction using Machine Learning - A Predictive Framework

Pranali Manwar [1], Sushma Shinde[2]

[1]Student, Siddhant College of Engineering, Sudumbre, Pune

[2]Professor, Savitribai Phule Pune University 2023-24

Abstract: *The global rise in criminal activities necessitates a concerted effort to lower crime rates, as they have a direct impact on a nation's social and economic progress. Consequently, there is an immediate requirement for security agents and agencies to combat and diminish crime within society. Additionally, by creating efficient crime detection systems, law enforcement agencies can actively respond to incidents, thwarting criminal acts and safeguarding communities.*

The proposed system helps to predict the type of crime in a particular region of the Chicago based on patterns of criminal activity. Random forest is most accurate algorithm providing maximum accuracy of 90.23% in predicting the crime rate based on different factors like crime date, location, time, crime type. domestic, ward, description.

Keywords: Crime rate prediction, Machine learning, Random Forest, Crime type, Crime location, Crime type, Location.

I. Introduction

Chicago Police Department data indicate that as crime rates returned to pre-pandemic levels in 2023, the number of shootings and homicides registered in the city decreased by 13%. Furthermore, the number of homicides in the city last year remained among the highest in the previous 20 years. In Chicago in 2023, there were 2,450 gunshots and 617 homicides [1]. Thus, it could potentially state that Chicago's crime rate is continually high. The city has seen a significant increase in robberies, despite a decline in shooting events year over year. The amount of motor vehicle thefts (29,287) and robberies (11,051) that were reported in 2023 both increased dramatically, by 23% and 37%, respectively [2]. Thus, predicting crime rates in Chicago is crucial for enhancing public safety, optimizing resource allocation, preventing crime, fostering community engagement, informing policy decisions, mitigating economic impacts, and upholding civil liberties and equity.

Crime rates rise along with population growth, making it difficult for authorities to forecast crime rates accurately. The authorities may not be able to anticipate future crimes because they are focusing on so many different topics. Even with their greatest efforts, the government and police officers are unable to guarantee a decrease in crime. It might be challenging for them to predict the amount of crime in the future. Several investigations have been carried out in relation to crimes in order to support police enforcement.

These days, a variety of approaches, including statistical methods, supervised learning techniques, and unsupervised learning methods, can be used to forecast crime analysis. These days, improving crime prediction is aided by the development of categorization systems, particularly machine learning algorithms [3]. In the field of crime prediction, machine learning techniques are frequently employed. In our study, we have demonstrated the application of machine learning techniques, such as random forests, to predict crimes based on several parameters, such as time, date, location, arrest, or description. Thus the main contribution of this paper are:

1. Implementing a crime rate detection project demonstrates a commitment to community safety.
2. To predict the type of crime which will happen in a particular area.
3. To improve the investigation efforts for any type of crime
4. To support community policing initiatives, encouraging collaboration between law enforcement and community members to collectively address safety concerns

II. Related Work

Related work in crime prediction involves reviewing existing research, methodologies, and technologies used in the field of predictive policing and crime analysis.

In their study, P. Poornima et al. [4] analyze the effectiveness of various machine learning algorithms in predicting public property crime using historical data from a specific area in southeast China. The findings indicate that the LSTM model outperforms other algorithms such as KNN, random forest, support vector machine, naive Bayes, and convolutional neural networks when solely relying on historical crime data for prediction. Additionally, the research highlights that the Decision tree, random forest, and Extra tree classifier algorithms exhibit superior performance with optimal training and high accuracy.

In their research, Mahima Chowdary et al.[5] proposed a solution to address the issue of crime rate prediction. They introduced a method called Assemble-stacking-based crime prediction method (SBCPM), which utilizes SVM algorithms. This method is implemented in MATLAB and aims to effectively select appropriate crime predictions through coordinated learning-based techniques.

A study by McClendon et al.[7] compares real crime statistics for the state of Mississippi, provided by neighborhoodscout.com, with the violent crime patterns found in the Communities and Crime Unnormalized Dataset from the University of California-Irvine repository. Using the same finite set of features, the study applied the Linear Regression, Additive Regression, and Decision Stump algorithms to the Communities and Crime Dataset. Out of the three techniques chosen, the linear regression algorithm yielded the best results overall.

Hitesh Kumar et al. [8] talk about the identification of criminals engaged in crimes at Redeemer's University. The Directorate of Students and Services Development (DSSD) provided historical data on crimes and offenders, which was pre-processed to guarantee clean and accurate data in order to detect crimes. To analyse and train the data, the Iterative Dichotomiser 3 (ID3) decision tree technique was used. Using decision trees for crime prediction, the trained model was then utilised to create a system that exposed the hidden linkages within the crime-related data.

Tyagi et al.'s study [9] demonstrates how various data mining and machine learning approaches are used in criminal investigations. This study's primary goal is to clarify the approaches used in crime data analytics. A number of machine learning techniques, including as clustering, naïve Bayes, SVM, and KNN, were used to understand, categorise, and analyse datasets according to preset standards. It is feasible to determine the type of crime and anticipate future crime hotspots by comprehending and analysing the data found in crime records.

H. Zheng, Y. Li, and Y. Zhu's paper "Crime Prediction Based on LSTM Recurrent Neural Networks" looks into the use of LSTM recurrent neural networks in crime prediction. For the purpose of predicting crime, the authors suggest a deep learning structure that combines fully linked and LSTM layers. The efficacy of the suggested strategy in collecting temporal patterns in crime data is demonstrated by experimental results.

III. Proposed system

The creation of models that can analyse past crime data, identify patterns, and forecast upcoming criminal activity is known as crime prediction. Thus, the approach aids in lowering the nation's crime rate. By assisting in the identification of criminal activity inside a particular region, the proposed approach improves understanding of the patterns of crime that are common in the area.

Data Uploading: the historical crime dataset from Kaggle id downloaded, which contains details about the time, place, kind of crime, and other pertinent characteristics including demographics, environmental conditions, and other contextual data that might have an impact on crime. The algorithm uses this data as input to forecast and solve crimes considerably more quickly.

Data Pre-processing: missing values, outliers, and inconsistencies in the data are handled from Kaggle in order to identify significant features that are quite natural for forecasting the crime.

i. Crime occur_ Date
ii. Location
iii. Crime Type
iv. Case_ Number
v. Arrest
vi. Case_ Description
vii. Domestic
viii. Ward
ix. Description

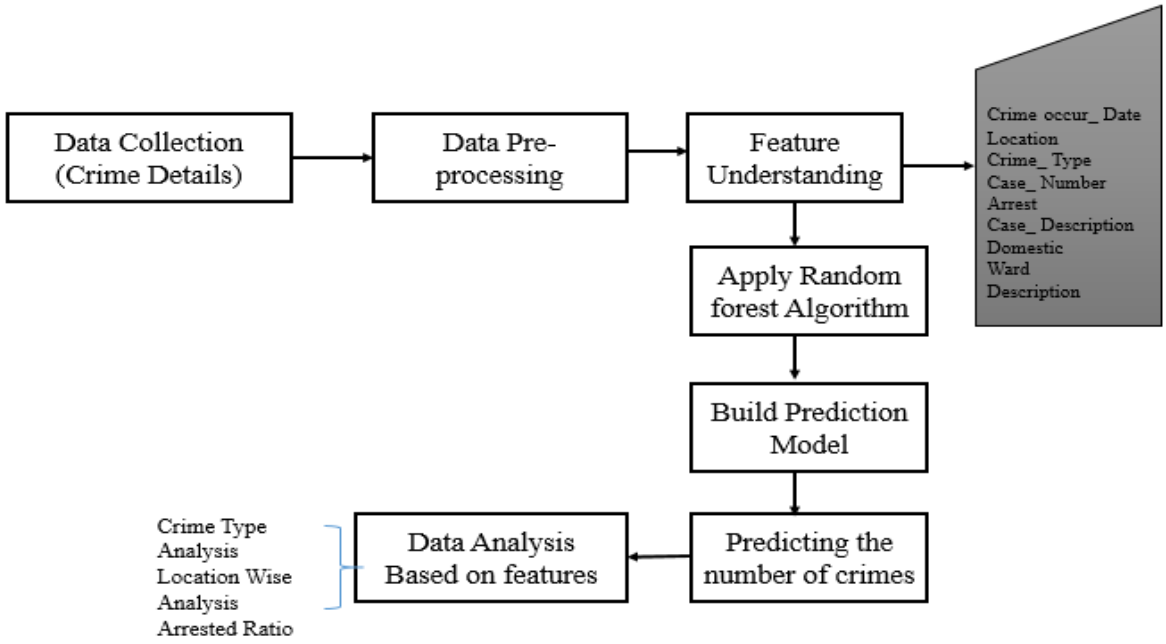

Figure 1: System Block diagram.

Feature Selection and Understanding: Relevant characteristics that can help in the prediction of crime are chosen, such as the geographic area's ability to reveal the underlying patterns of the crime that the region experiences. Additionally, dependent variables are used to identify the characteristics that are mostly reliant on variables connected to crime, like date, time, and location. Consequently, this approach aids in choosing the most important features.

Prediction Model: The random forest technique is used to create a prediction model. This method predicts crucial elements of crime detection and prevention by using word-formation vectors and dependent variable analysis. The method will produce several decision trees, each trained with a distinct random subset of features and on a distinct subset of the data.

Data Analysis: Building a prediction model requires the identification of important features, such as the type of crime and its likely location, which can be determined by data analysis.

A. Workflow Diagram

A workflow diagram for crime prediction using machine learning typically involves several stages. Figure 2 illustrating how data flows through each step of the process from collection to deployment.

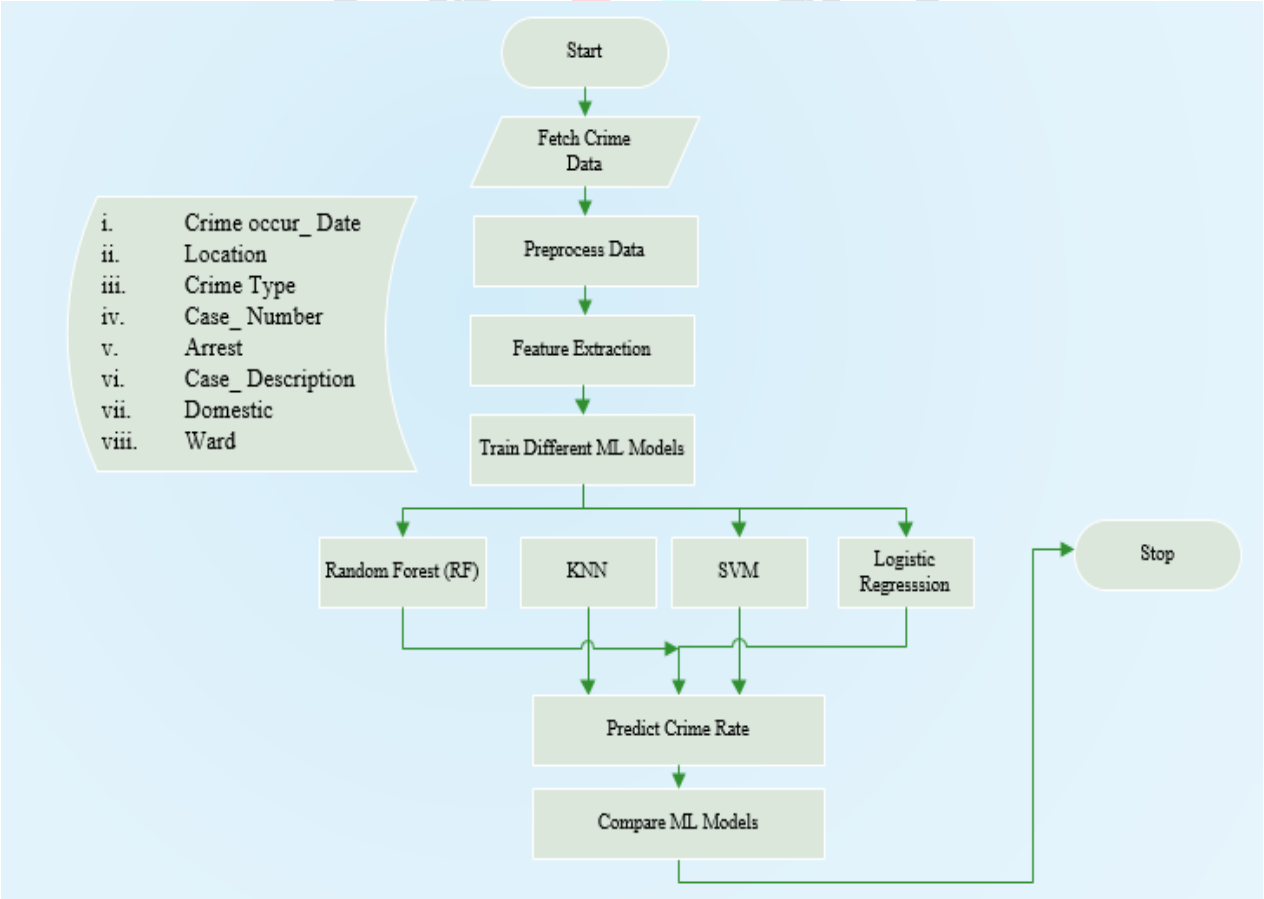

Figure 2: System Workflow

IV. Algorithm Used

A. Random Forests Algorithm

Random Forest is a popular machine learning algorithm used for classification and regression tasks, making it suitable for crime prediction. Here's an explanation of how Random Forest works in the context of crime prediction with a Chicago dataset.

Data Collection and Pre-processing: Crime data for Chicago are gathered from Kaggle website and from reliable sources such as the Chicago Police Department or government data portals. The dataset is pre-processed by cleaning it, handling missing values, encoding categorical variables, and selecting relevant features.

Feature Selection: Features are identified that are likely to be informative for crime prediction, such as location (latitude and longitude), time of day, day of week, type of crime, historical crime rates, demographic characteristics of the area, weather conditions, and socioeconomic factors.

Training and Testing Data: the dataset is splited into training and testing sets. The training set is used to train the Random Forest model, while the testing set is used to evaluate its performance. Consider using techniques like cross-validation to ensure robustness of the model.

Model Training: The Random Forest classifier is trained using the training data. Hyperparameters are tuned such as the number of trees in the forest, maximum depth of the trees, and minimum number of samples required to split a node to optimize the model's performance.

Model Evaluation: Trained Random Forest classifier is evaluated using the testing data. Evaluation metrics are evaluated such as accuracy, precision, recall, F1-score, and area under the ROC curve to assess the model's performance in predicting crime occurrences.

Feature Importance Analysis: The feature importance rankings are analysed by the Random Forest model to understand which features contribute most significantly to crime prediction. This information can help to identify important predictors and provide insights into the underlying factors driving criminal activity in Chicago.

Model Interpretation: Random Forest model are interpreted the result to understand how different features influence crime prediction. Visualizations such as feature importance plots, confusion matrices, and ROC curves can help interpret the model's decisions and assess its strengths and weaknesses.

Deployment and Monitoring: Trained Random Forest model are deployed for crime prediction in Chicago. Performance is monitored over time and update the model as new data becomes available. Ethical considerations and privacy concerns are addressed in the deployment process. By following these steps, Random Forest can be used for crime prediction with a Chicago dataset, helping law enforcement agencies and policymakers make informed decisions to enhance public safety and reduce crime in the city.

V. Datasets Used

The Chicago Crime dataset contains a summary of the reported crimes occurred in the City of Chicago from 2001 to 2017. We have used 25025 records for the experimentation. Data is extracted from the Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) system. In order to protect the privacy of crime victims, addresses are shown at the block level only and specific locations are not identified. Different features such as case_number, date, block, primary_type, location, arrests, domestic, district, community area.

VI. Results

After uploading the Chicago crime dataset from Kaggle, we employed the WEKA tool to perform the Random Forest Classification algorithm. By inputting the dataset into the WEKA program, a classification report is generated, dividing the data into two classes, as presented in Table I. the experimentation has used 25025 records from the dataset.

Class 0 - Crime Occurrence

Class 1- No Crime Occurrence

Table I: Classification Report

Class	a	b
a	21169	4040
b	869	24156

The values of true positive (TP), true negative (TN), false positive (FP), and false negative are essential in calculating the confusion matrix. This matrix is then represented in a table II format based on these values.

Table II: TP, TN, FP and FN

	TP	TN	FP	FN
Class 0	21169	24156	869	4040
Class 1	24156	21169	4040	869

The values of TP, FP, and FN i.e truly positive, false positive and false negative can be used to calculate the accuracy, precision, sensitivity and specificity using formulas mentioned below:

Precision=TP / (TP + FP)

Sensitivity = TP / (TP + FN)

Specificity = TN / (FP + TN)

Accuracy = TP+ TN/ TP+ TN+ FP+ FN

Table III: Confusion Matrix

Class	Accuracy	Precision	Sensitivity	Specificity
Class 0	90.23%	96.06%	83.97%	96.53%
Class 1	90.23%	85.67%	96.53%	83.97%

The graphical representation of the above table is represented in following graph. Thus the average accuracy of the random forest for prediction of employee attrition rate is 90.23%

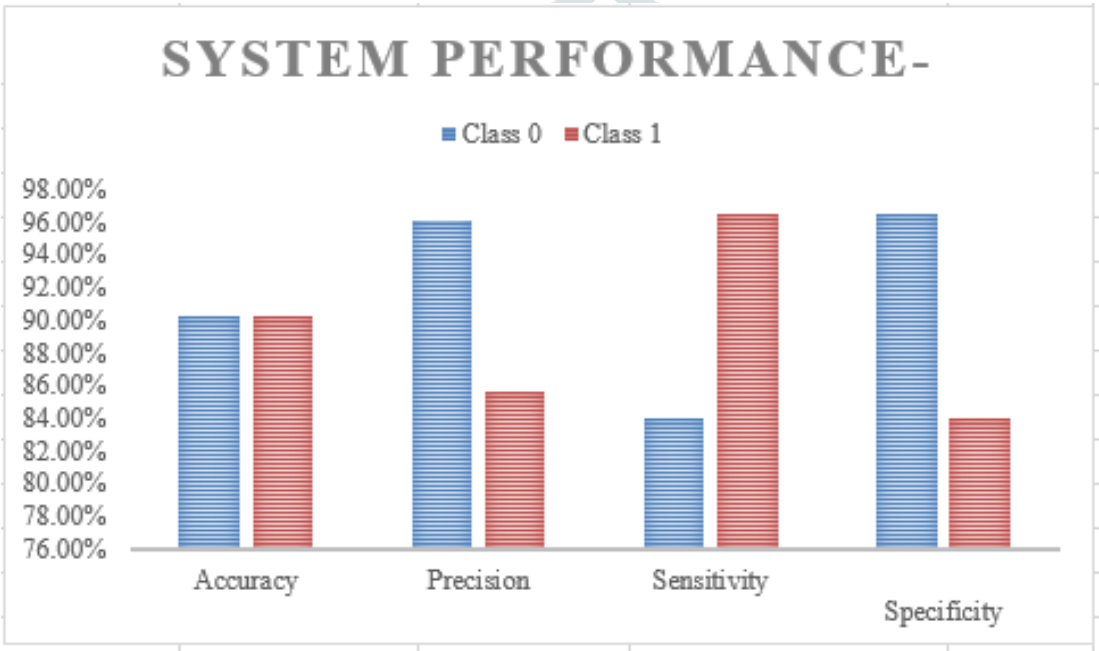

Figure 3: System Performance

VII. Conclusion

Crime prediction became the hot research area nowadays because of its correlation benefits to any society or nation's security. It is found that many studies adopted supervised learning approaches to the field of crime prediction compared to others. It is obviously concluded, that machine learning technique achieved the highest crime prediction accuracies. By making it easier to analyse time series crime data, the suggested approach enables us to find patterns in the data. Since time affects crime, time-dependent data analysis can be used to predict future crime incidents. We can identify the times when crime is more likely to occur and look at annual trends in crime by using the available crime data. Random forest is most accurate algorithm in predicting the crime rate in an accuracy of 99.23 %.

References

[1] "Latest 1.Matt Masterson , Shootings, Homicides in Chicago Drop 13% in 2023 and Returned to Pre-Pandemic Levels, But Violence Numbers Remain Among Highest in Recent Decade https://news.wttw.com/2024/01/02/shootings-homicides-chicago-drop-13-2023-and-returned-pre-pandemic-levels-

violence#:~:text=There%20were%20617%20homicides%20and,assault%20saw%20increases%20in%202023.| January 2, 2024

[2] Tahman Bradley,"Chicago crime stats 2023: Murders, shootings down but crime still stubbornly high",https://wgntv.com/news/chicagocrime/chicago-crime-stats-2023-murders-shootings-down-but-crime-still-stubbornly-high/

[3] Ruaa Mohammed Saeed EMAIL logo and Husam Ali Abdulmohsin,"A study on predicting crime rates through machine learning and data mining using text", Journal of Intelligent Systems https://doi.org/10.1515/jisys-2022-0223

[4] P.Poornima1 ,L.Tharun kumar, S.Sumera, K.Deepika, ,K.Yoshitha5,"Crime Rate Prediction Using Machine Learning",International Journal of Engineering Technology and Management Sciences Website: ijetms.in Special Issue: 1 Volume No.7 April – 2023

[5] K.Mahima Chowdary, 2 M. Vaishnavi, 3A.Saketh Rao, 4B.Hemath, 5Dr.M.Padmaja,"CRIME RATE PREDICTION AND ANALYSIS SYSTEM USING MACHINE LEARNING ALGORITHMS",2023 JETIR March 2023, Volume 10, Issue 3

[6] V. Chandra Shekhar Rao,Kallepelly Spandhan,C. Srinivas,M. Sujatha,"An Adaptive Technique for Crime Rate Prediction using Machine Learning Algorithms",https://ijritcc.org/index.php/ijritcc/article/view/7954

[7] Lawrence McClendon and Natarajan Meghanathan*,"USING MACHINE LEARNING ALGORITHMS TO ANALYZE CRIME DATA",Machine Learning and Applications: An International Journal (MLAIJ) Vol.2, No.1, March 2015

[8] Hitesh Kumar Reddy Toppi Reddya,, Bhavna Sainia, Ginika Mahajana ,"Crime Prediction & Monitoring Framework Based on Spatial Analysis ",International Conference on Computational Intelligence and Data Science (ICCIDS 2018).

[9] Tyagi D, Sharma S (2018) An approach to crime data analysis: a systematic review. Int J Eng Technol Manag Res 5(2):67–74. https://doi.org/10.29121/ijetmr.v5.i2.2018.615

[10] Nigus Asres Ayele, Yidnekachew Kibru,Tsion Eshetu Meskela, "Designing Time Series Crime Prediction Model using Long Short-Ter Memory Recurrent Neural Netw",November 2020 . International Journal of Recent Technology and Engineering (IJRTE) 9(4):402-405