WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Crime Rate Prediction using Machine Learning

Harsh Patel[1], Devansh Gupta[2], Ms. Nidhi Kaushik[3]

[1, 2] IMS Engineering College, Ghaziabad, India B-tech Information Technology

[3] Assistant Professor IMS Engineering College, Ghaziabad, India B-tech Information Technology

Index Terms—crime, crime rate, machine learning, knn, ran- dom forest

INTRODUCTION

Crime is a law-enforced act that can result in fines, imprisonment, or other legal consequences. It has been a significant part of human civilization since ancient times and is becoming increasingly prominent due to technological advancements and globalization. Crime prediction is crucial to determine the increase or decrease in crime rates from previous years. Machine learning is an emerging field of study that uses sophisticated algorithms and data-driven methods to detect and predict criminal activities. It can identify patterns in data that may indicate future crimes, such as past criminal activities, demographic information, and environmental factors. By lever- aging this data, machine learning can create predictive models that identify the likelihood of a crime occurring in a specific area or time frame. Additionally, machine learning can provide insights into criminal behavior, helping law enforcement professionals better understand and address criminal activity.

Problem statement

Crime is a global security threat, with increasing urban populations leading to higher crime rates. Officials face the challenge of accurately predicting future crime rates using large datasets. Predictive models, based on historical crime data, identify high-risk areas for future criminal activity, improving crime prevention and policing efforts.

Objective

The project aims to create a system that accurately predicts crime rates and identifies future trends, enabling officials to develop strategies to reduce crime and create a safer environment. Machine learning algorithms will be applied to predict crime rates based on year, location, and crime type. The system will also analyze crime patterns using multi-linear regression techniques based on territorial distribution and crime recognition.

LITERATURE REVIEW

Crime rate prediction has been a topic of interest in various studies. One such study by Muhammad Alkaff uses the Recurrent Neural Network (RNN) with the Gated Recurrent Unit (GRU) architecture to predict crime rates in Banjarmasin City. This model considers inflation rate and discretionary income and is more effective in adapting to different timescales and dealing with Vanishing Gradient problems. Another study by Wajiha Safat uses machine learning and deep learning techniques to analyze crime prediction in Chicago and Los Angeles datasets. The study found that XGBoost performed best with an accuracy of 94% and 88%.

Sakib Mahmud and Musfika Nuha proposed the relationship between crime and different features in criminology literature. They used various algorithms to reduce crimes and detect criminal activity, including Z-Crime Tools, Advanced ID3 algorithms, K-means clustering and deep learning, random forest and naïve Bayes algorithms, and multi-linear regression. They also used Apriori and Naive Bayes algorithms to identify and predict criminal trends and patterns.

Gaurav Hajela proposed a clustering-based hotspot identification approach for crime prediction, which includes crime hotspot identification, dataset preparation, and crime prediction approach. The decision tree approach achieved 83.95% accuracy and outperformed Nave Bayes.

Lastly, a system that converts crime information into a data-mining problem was proposed by Ms. Vrushali Pednekar, Ms. Trupti Mahale, Ms. Pratiksha Gadhave, and Prof. Arti Gore.

Masoomali Fatehkia and others have developed methods to improve crime rate predictions in urban areas. Fatehkia's study uses Facebook interests to measure the prevalence of interests among a ZIP code's Facebook population, revealing that a combination of demographic factors and Facebook interests had the strongest predictive strength for all crime types. Meanwhile, Pednekar, Mahale, Gadhave, and Gore's KNN system focuses on crime analysis, data pre-processing, and interpretation to help detectives solve crimes faster. Both methods use data mining techniques to predict crimefrequency based on territorial distribution but do not address parameters like outlier effects or feature value.

Here are the top ten countries with the highest crime rate per 100k in 2023

Country	**Crime rate (%)**
Venezuela	**83.76**
Papua New Guinea	**80.79**
South Africa	**76.86**
Afghanistan	**76.31**
Honduras	**74.54**
Guyana	**68.74**
El Salvador	**67.79**
Brazil	**67.49**
Jamaica	**67.42**
Syria	**67.18**

Murder crime rate per 100,000

Characteristic	**Murder rate**
Celaya, Mexico	**109.39**
Tijuana, Mexico	**105.15**
Ciudad Juarez, Mexico	**103.61**
Ciudad Obregon, Mexico	**101.13**
Irapuato, Mexico	**94.99**
Ensenada, Mexico	**90.58**
St.Louis, United States	**87.83**
Uruapan, Mexico	**72.59**
Fiera De Santana, Brazil	**67.46**
Cape Town, South Africa	**64**
Cumana, Venezuela	**62.42**
Fortaleza, Brazil	**62.28**
Mossoro, Brazil	**62.21**
Guyana, Venezuela	**62.1**
Zacatecas, Mexico	**59.22**
Baltimore, United States	**56.45**

Kingston, Jamaica	**54.46**
Acapulco, Mexico	**54.13**
Caracas, Venezuela	**52.82**
Vitoria da Conquista, Brazil	**52.47**

The 20 most peaceful countries in the world according to the Global Peace Index 2023
(the lower the index value, the higher the peacefulness)

METHODOLOGY

Ensemble learning is a machine learning technique that combines multiple models to improve predictive performance. It works by building multiple models from the same training data set and combining them to make more accurate predictions. This technique is popular in various applications, such as computer vision and natural language processing. Random Forest is an ensemble learning technique that combines multi-ple decision trees to generate more accurate predictions. It duces variance and improves prediction accuracy by randomly sampling data points, selecting features, and building multiple decision trees on the data. The predictions from each tree are then averaged to produce the final prediction. Random Forests are more accurate than traditional decision trees and are robust even when dealing with large datasets. The algorithm involves starting with a dataset of observations and labels, randomly selecting 'k' features, choosing the best-split point, creating a tree using the chosen split points, repeating steps 2-4 for each feature, combining the trees, making predictions on new data, and calculating the accuracy of the predictions

Characteristic	Index points
Iceland	1.12
Denmark	1.31
Ireland	1.31
New Zealand	1.31
Austria	1.32
Singapore	1.33
Portugal	1.33
Slovenia	1.34
Japan	1.34
Switzerland	1.35
Canada	1.38
Czechia	1.4
Finland	1.45
Croatia	1.46
Germany	1.49
Netherlands	1.5
Bhutan	1.51
Hungary	1.51
Malaysia	1.52
Belgium	1.53

A. Proposed System

- Initially, the dataset is prepared manually based on the publication available on the National Crime Rate Bureau (NCRB) official website.
- Data Preprocessing: The data is prepared in the correct format for analysis. Some columns are removed or transformed, and label encoding is used to convert the categorical data into numeric for better prediction.
- Random Sampling: After feature selection, the data has been split into two parts: training data (70%) and testing data (30%).

- Model Creation: The model algorithms are imported from sklearn. Model is build using model.fit(). The dataset has been analyzed using five different models: support vector machine, nearest neighbor, decision tree, random forest, and neural network.
- Model Selection: Based on the defined goals and model performance, random forest model has been selected. Predic- tion is done using model.predict(). The model accuracy is determined using accuracy_score imported from metrics.
- Model Deployment: The model has been deployed for prediction using various web technologies.
- Model
- Bllock diagram Data flow diagram .

RESULT

Of the five models that were chosen, the Random Forest Regression model shows the highest accuracy in predicting test results. The model forecasts the crime rate values for 19 Indian metropolitan cities: Ahmedabad, Bengaluru, Chennai, Coimbatore, Delhi, Ghaziabad, Hyderabad, Indore, Jaipur, Kanpur, Kochi, Kolkata, Kozhikode,

Data Modelling

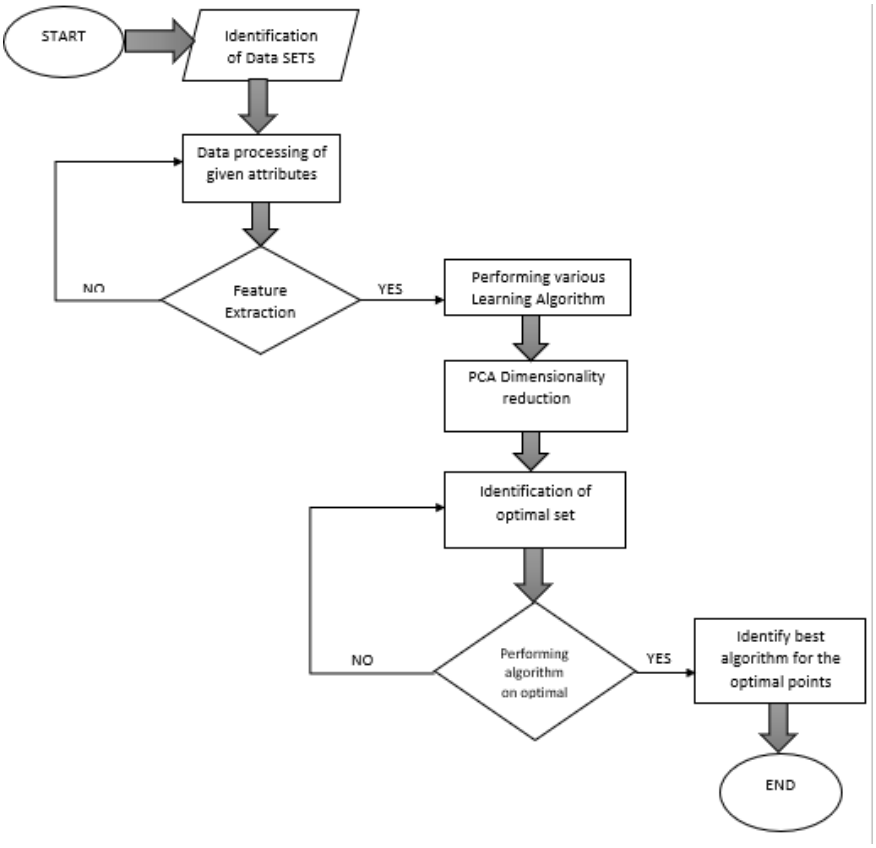

Fig. 1. Activity Diagram

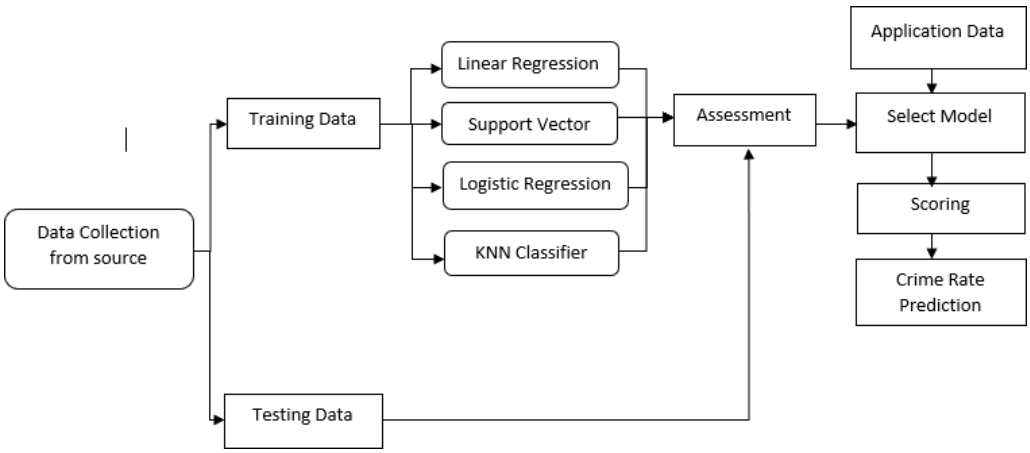

Fig. 2. Block Diagram

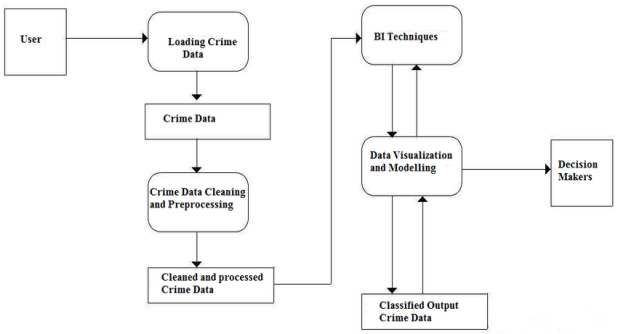

Fig. 3 Data Flow Diagram

REFERENCES

1. W. Safat, S. Asghar and S. A. Gillani, “Empirical Analysis for Crime Prediction and Forecasting Using Machine Learning and Deep Learning Techniques,” in IEEE Access, vol. 9, pp. 70080-70094, 2021, doi: 10.1109/ACCESS.2021.3078117.
2. Shah, N., Bhagat, N. & Shah, M. Crime forecasting: a machine learning and computer vision approach to crime prediction and prevention. Vis. Comput. Ind. Biomed. Art 4, 9 (2021). https://doi.org/10.1186/s42492-021-00075-z
3. Caldwell, M., Andrews, J.T.A., Tanay, T. et al. AI-enabled future crime. Crime Sci 9, 14 (2020). https://doi.org/10.1186/s40163-020-00123-8
4. S. Kim, P. Joshi, P. S. Kalsi and P. Taheri, “Crime Analysis Through Machine Learning,” 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, Canada, 2018, pp. 415-420, doi: 10.1109/IEMCON.2018.8614828.
5. Mittal, M., Goyal, L.M., Sethi, J.K. et al. Monitoring the Impact of Economic Crisis on Crime in India Using Machine Learning. Comput Econ 53, 1467–1485 (2019). https://doi.org/10.1007/s10614-018-9821-x
6. S. S. Kshatri, D. Singh, B. Narain, S. Bhatia, M. T. Quasim and G. R. Sinha, “An Empirical Analysis of Machine Learning Algorithms for Crime Prediction Using Stacked Generalization: An Ensemble Approach,” in IEEE Access, vol. 9, pp. 67488-67500, 2021, doi: 10.1109/ACCESS.2021.3075140.
7. X. Zhang, L. Liu, L. Xiao and J. Ji, "Comparison of Machine Learning Algorithms for Predicting Crime Hotspots," in IEEE Access, vol. 8, pp. 181302-181310, 2020, doi: 10.1109/ACCESS.2020.3028420.
8. P. Tamilarasi and R. U. Rani, “Diagnosis of Crime Rate against Women using k-fold Cross Validation through Machine Learning,” 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2020, pp. 1034-1038, doi: 10.1109/ICCMC48092.2020.ICCMC-000193.
9. Kumar, A. Verma, G. Shinde, Y. Sukhdeve and N. Lal, "Crime Prediction Using K-Nearest Neighboring Algorithm," 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), Vellore, India, 2020, pp. 1-4, doi: 10.1109/ic-ETITE47903.2020.155.

10. Rajapakshe, S. Balasooriya, H. Dayarathna, N. Ranaweera, N. Walgampaya and N. Pemadasa, "Using CNNs RNNs and Machine Learning Algorithms for Real-time Crime Prediction," 2019 International Conference on Advancements in Computing (ICAC), Malabe, Sri Lanka, 2019, pp. 310-316, doi: 10.1109/ICAC49085.2019.9103425.
11. N. Kanimozhi, N. V. Keerthana, G. S. Pavithra, G. Ranjitha and S. Yuvarani, "CRIME Type and Occurrence Prediction Using Machine Learning Algorithm," 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India, 2021, pp. 266-273, doi: 10.1109/ICAIS50930.2021.9395953.
12. Singh, M. K., & Kumar, A. (2023). Cucumber Leaf Disease Detection and Classification Using a Deep Convolutional Neural Network. Journal of Information Technology Management, 15(Special Issue: EIntelligent and Security for Communication, Computing Application (ISCCA-2022)), 94-110.
13. S.K. Upadhyay, & A. Kumar, "Early-Stage Brown Spot Disease Recognition in Paddy Using Image Processing and Deep Learning Techniques" Traitement du Signal, 2021, Vol. 38, No. 6, pp. 1755-1766. https://doi.org/10.18280/ts.380619.
14. S.K. Upadhyay, & A. Kumar, "A novel approach for rice plant diseases classification with deep convolutional neural network" Int. j. inf. tecnol. 14, 185–199 (2022). https://doi.org/10.1007/s41870-021-00817-5
15. Rukhsar and S. K. Upadhyay, "Rice Leaves Disease Detection and Classification Using Transfer Learning Technique," 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2022, pp. 2151-2156, doi: 10.1109/ICACITE53722.2022.9823596.
16. S. K. Upadhyay and A. Kumar, "An Accurate and Automated plant disease detection system using transfer learning based Inception V3Model," 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2022, pp. 1144-1151, doi: 10.1109/ICACITE53722.2022.9823559.
17. Rukhsar and S. K. Upadhyay, "Deep Transfer Learning-Based Rice Leaves Disease Diagnosis and Classification model using InceptionV3", 2022 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES), Greater Noida, India, 2022, pp. 493-499, doi: 10.1109/CISES54857.2022.9844374.
18. S. K. Upadhyay and A. Kumar, "Automatic Recognition and Classification of Tomato Leaf Diseases Using Transfer Learning Model", Future Farming: Advancing Agriculture with Artificial Intelligence (2023) 1: 23. https://doi.org/10.2174/9789815124729123010005
19. Cite Upadhyay, S., Jain, J., & Prasad, R. (2024). Early Blight and Late Blight Disease Detection in Potato Using Efficientnetb0. International Journal of Experimental Research and Review, 38, 15-25. https://doi.org/10.52756/ijerr.2024.v38.002