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

Nishant
CSE with specialization in artificial intelligence and machine learning
Apex Institute of Technology, Chandigarh University
Mohali, Punjab
gurinishant2002@gmail.com

Priya Karn
CSE with specialization in artificial intelligence and machine learning
Apex Institute of Technology, Chandigarh University
Mohali, Punjab

Jasskaran Singh Shanghari
CSE with specialization in artificial intelligence and machine learning
Apex Institute of Technology, Chandigarh University
Mohali, Punjab
jasskaransingh0007@gmail.com

Lakshay Bhardwaj
CSE with specialization in artificial intelligence and machine learning
Apex Institute of Technology, Chandigarh University
Mohali, Punjab
lakshaybhardwaj@gmail.com

Nikita
Assistant Professor
Chandigarh University
Mohali, Punjab

Abstract*— Predicting and analysing crime rates are essential elements of modern law enforcement and public policy, fulfilling multiple functions that are vital to the welfare of society. Authorities are able to effectively deploy resources by analysing crime rates, which provide them with crucial insights into the patterns and trends of criminal behaviour. By facilitating the deployment of law enforcement officers to locations and times when criminal activity is at its peak, this allocation maximises financial resources and improves the effectiveness of efforts to prevent crime. Furthermore, by locating possible hotspots for crime and new risks, predictive analysis supports proactive efforts by enabling law enforcement to put preventive measures in place that attempt to stop criminal activity before it gets out of control. Through focused interventions and heightened police presence, such projects not only improve public safety but also cultivate community trust and engagement. Moreover, evidence-based policymaking is guided by crime rate analysis, which also helps to build interventions that are specifically designed to address the issues and underlying causes of criminal behaviour. This method makes it easier to evaluate intervention programmes, which helps policymakers improve their plans and successfully modify their methods in response to changing conditions. In the end, crime rate analysis aids in long-term strategic planning by foreseeing difficulties and facilitating proactive measures to guarantee society's safety and well-being. Authorities can negotiate the complex dynamics of crime and strive towards creating safer and more resilient communities by collaborating with one another and using data-driven insights *Keywords: enforcement, policy making, data drive

I. Introduction

The rate of crime is rising daily as a result of current technology and high-tech techniques that enable criminals to carry out their illicit activities.The Crime Record Bureau reports that crimes such as robbery, sex abuse, rapes, and other crimes have increased, crimes like burglary and arson have decreased. We'll gather crime statistics from a variety of websites, blogs, and news sources. The massive amount of data serves as a record for building a vast database storing all crime reports. The generous information gained from data mining methodoloy will contribute to a decrease in crime because it will speed up the process of apprehending offenders and identify high-crime regions.When used to crime datasets, data mining yields strong results and speeds up the process of investigating crimes. The police department can benefit from the information collected by data mining techniques. The police have discovered that one particular strategy, known as "hot spot" identification, is helpful in identifying regions where there is a significant concentration of criminal activity. Crime report datasets can provide significant findings through the application of data mining techniques. Crime analysis is the initial stage of the study of crime. Investigating, connecting, and identifying relationships between different crimes and their features is the goal of crime analysis. This study facilitates the on-demand preparation of maps, queries, and statistics. Seeing whether a crime in a known ornamental form—or the need for a new one. Since criminals are busy and engage in their comfort zones, crimes can be very

frequent. After they flourish, they venture to commit the same crime under homologous circumstances. Crime rates were determined by a number of variables, including criminal intelligence, site security, etc. The project adhered to the procedures employed in data analysis, where pattern recognition, data classification, data gathering, prediction, and visualisation are the key stages. The suggested system makes use of a variety of machine learning algorithm-based visualisation techniques to display crime trends and predict crimes. To summarise, the study and forecast of crime rates are vital for improving public safety, making the best use of available resources, guiding policy decisions, and promoting efficient cooperation among law enforcement, politicians, and communities. Authorities can better comprehend, predict, and address the intricate dynamics of crime in society by utilising data-driven insights.

II. Literature Review

The goal of analysing crime rates is not just to comprehend historical patterns but also to influence governmental decisions, advance social justice, improve urban planning, stimulate economic growth, and empower communities. Societies may make surroundings for the present and future generations safer, more inclusive, and wealthier by utilising data-driven insights and cooperative ways.

[1] According to a study by Mehmet Sait, and Mustafa Gök introduced the prediction of criminal activities, which is used to identify the most likely offender of a specific offence incidence when the suspected list of offenders is given access to synthetic crime data created by the Gaussian Mixture Model. The authors scrutinized the on spot performance of the Naïve Bayes Classifier and Decision Tree for offender prediction. The comparison led the authors to the conclusion that the Naïve Bayes Classifier performed better, consuming less accomplishment time and achieving an accuracy of 78.05%.

[2] Sivaranjani,S.,S.Sivakumari,and M. Aasha Presented. An analysis of crime in six Indian cities in the Tamilnadu region The authors come to the conclusion that DBSCAN clustering outperforms other methods for clustering alike patterns to identify crimes in six Tamilnadu, India cities, with accuracy 0.95, recall 0.91, and F measure 0.93. Kmeans, DBSCAN, and agglomerative clustering are three clustering techniques they used to achieve this.

[3] Presented by Nafiz M. et al. Using past crime data dating back up to 30 years, the statistical modeling tool CRIMECAST projects future crime by analyzing historical trends, patterns, and variables that impact crime, including frequency, timing, kind, and victimization.

[4] Tahani A., et al. Decision Tree and Naïve Bayesian classifiers were used to analyse two separate crime data sets in order to identify the most likely crime sites and the frequency of those places through the use of the Apriori Algorithm. The authors combined demographic data with crime statistics to identify concerns that could compromise neighbourhood safety by introducing potential crimes that could occur in a certain location during a given time frame.

[5] Thongsatapornwatana U., studied 0, several studies that analyse and forecast criminal data via data mining techniques. In order to assist researchers studying criminal data, the author highlights research gaps and obstacles from various studies and suggests several data mining strategies for identifying patterns and trends in crime data.

[6] Vedhadharshan B, V.K, H.S, T.P. Chand, R. Kaladevi, and B.A. Preliminarily unknown, known data can be prioritised by the suggested method using the concepts of data mining and machine learning. Utilising the crime analysis technique lowers crime rates, which benefits the nation's growth and reputation. We can also lower the investment costs for the criminal investigation.

[7] Cesario, Cesario E, Catlett C, Talia D In order to estimate the crime level category, they examined two different data sets that were gathered from various sources.The Naïve Bayes classifier outperforms the neural network in terms of accuracy, with 90.2207%, according to the authors' analysis and comparison of the two methods for predicting the possbility of a bad fease that happens in a given place.

[8] Zhang Q, Yuan P, Zhou Q, Yang Z. Enhanced multiple crime category, known as "heat levels," was introduced and is divided into five categories to anticipate areas with high crime rates.The authors predicted areas with high crime rates using KNN and utilised LDA to reduce features.

[9] Sathyadevan,Shiju, and Surya Gangadharan. It predicts districts with a high probability of crime occurring on a given day, and Naïve Bayes crime data analysis and forecasting has an accuracy of 90%.The Apriori Algorithm is used to identify recurring patterns of crime in specific districts. The research used GraphDB and data from MongoDB to visualise neighbourhoods with a high crime rate.

[10] Yu CH, Morabito M, Ding W, Ward MW Chung-HsienY. et al. developed a prediction approach for areas with a high density of home burglaries and the prossibility that they would happen in the future. To preserve location and temporal data, as well as the overall count of offenses and the police officer's response to those offenses, it is important to use an appropriate data organisation technique.Secondly, to improve the accuracy of residential burglary prediction, an ensemble supervised data mining technique is used.

[11] Emmanuel A., et al. Support vector machines, naïve bayes, neural networks, and J48 were used to analyse crime data. The methods were compared using execution times and accuracy to forecast offence levels as "Low," "Medium," and "High." The decision tree (J48) performed better with 100% accuracy for crime predictions and took less execution time—0.06 seconds—according to the authors' analysis of the contrast.

[12] Lin, Wen-Yang, and Sy-Yen Kuo a comparison of various crime prediction techniques, consisting of support vector machines, random forests, and decision trees as examples of machine learning algorithms. Using data from actual crimes, the researchers assess how well these techniques work and talk

about their advantages and disadvantages. The goal of the study is to cellist some light on how well different prediction methods work for various kinds of crimes.

[13] Adnan, Muhammad, et al. investigated several machine learning strategies for crime prediction, including ensemble methods, decision trees, and neural networks. The paper outlines possible directions for further research and addresses the prospects and problems of utilising machine learning to forecast crime.

[14] Srivastava, Rajeev, and Sangeeta Mittal explains the use of classification, clustering, and affinity analysis techniques to analyse criminal data and find trends. The paper looks at the benefits and drawbacks of various data mining techniques and emphasises how crucial feature selection and model assessment are for jobs involving crime prediction.

[15] Khan, Fahad Shahbaz, et al**.** explains how to analyse crime data and create predictive models using ensemble methods and supervised and unsupervised learning algorithms. The paper also explores methods for enhancing the efficacy of machine learning-based crime prediction systems and tackles issues including feature selection, algorithmic bias, and data quality.

III. Proposed Methodology

Creating a Machine Learning model for predicting crime rates involves several consideration about our dataset and selecting a perfect model which can provide us high accuracy with good score: accuracy, precise, recall and F1 score. Here's a structured approach:

1. Data Collection

Collecting data is an initial stage also an important stage which helps in gathering data on which we can apply the ML algorithms. Gathering data for crime rate on the basis of history can be found on some government sites or can be from Kaggle. We always have a requirement of the good dataset because a good dataset have a good number of feature and the quality feature which are always required. Gather socioeconomic information for the areas of interest, such as population demographics, income and unemployment rates, educational attainment, and housing data.

2. Data Preprocessing:

Data preprocessing is an indispensable stage for cleaning the gathered data so that we can handle the missing values, outliers and reducing the inconsistencies. It also helps in Normalizing or Standardizing the numerical features so that we can get a data with the same scale. It also helps in encoding the features which have values in categories with techniques of one-hot encoding, label encoding and some other techniques.

3. Feature Selection and Engineering:

Examine the gathered information to find characteristics that have a strong correlation with crime rates. It helps in using statistical analysis and domain expertise, choose pertinent attributes. Create additional features, such crime density per capita or temporal trends, that could reveal hidden patterns or linkages in the data.

4. Model Selection:

When choosing machine learning algorithms to forecast crime rates, we have strict requirements at this point that include performance, scalability, and interpretability. Algorithms that may be used for the same purpose are decision trees, random forests, and linear regression.

5. Model Training:

This stage used to assess the effectiveness of the model, divide the preprocessed data towards training and testing sets. We will use methods like cross-validation to tweak hyperparameters as necessary while training the chosen models on the training set. Depending on the type of prediction task, evaluate models using relevant metrics like accuracy, mean squared error, or mean absolute error.

6. Model Evaluation:

It helps in evaluating the generalization performance of the training models by validating them against the testing set. Examine the model's performance using a variety of metrics like accuracy, precise, recall, F1 score, and contrast it with benchmarks or baseline models. It also used to find out how modifications to the input variables impact the model's predictions, perform sensitivity analysis.

After using these stages we will get our Machine learning based model which will able to predict the crime rates based on the past data with high performance. Performance of a model is very necessary thing for the model because if the performance is low then it can be dangerous for us so, creating model with thinking and using a best methodology is necessary.

IV. Results and Discussions

The successful extraction of significant insights from the dataset is demonstrated by the excellent accuracy scores of the Random Forest (0.9984), Logistic Regression (0.9415), and Decision Tree (0.9999) algorithms in the crime rate prediction model. Contributing to the overall predicted performance, each algorithm has certain strengths and abilities:

1. Random Forest:

The Random Forest Algorithm is assembled from many (many) decision trees with identical nodes that use various data sets to produce distinct leaves. To ascertain the solution, that can be thought of as the standard of all decision trees, it cojoins the decisions of various decision trees. The way nodes branch in a decision tree may be ascertained using entropy.

It excels in handling large datasets with numerous features and can effectively deal with both numerical and categorical data. In the context of crime rate prediction, Random Forest is capable of detecting intricate linkages and nonlinear interactions between several elements impacting criminal conduct. Its ability to handle high-dimensional data and mitigate overfitting makes it particularly suitable for analysing diverse socioeconomic, demographic, and environmental variables associated with crime dynamics. Mathematically, the prediction of a random forest can be represented as:

$$\hat{y} = \frac{1}{N} \sum_{i=1}^{N} \hat{y}_i$$

Where, N is the of trees in random forest and $\hat{y}_i$ is the prognosticate of i-th tree.

Furthermore, Random Forest offers feature significance rankings that let researchers determine which elements have the most impact on crime rates. As we can see with the figure 1.

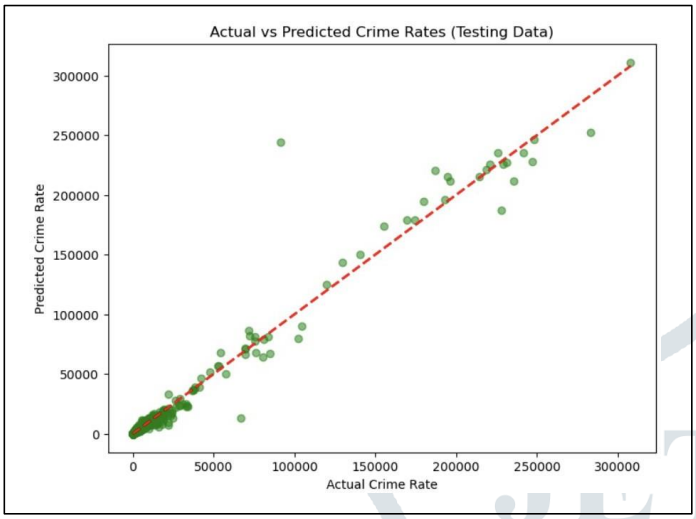

Figure: 1

2. Linear Regression

In order to model the affiliation between a variety of few independent variables and a dependent variable, statisticians employ the technique known as linear regression. It is assumed that there is a connection between the independent variables (predictors) and the dependent variable (outcome). In short, the objective of linear regression is to fit a straight line to the data points in order to minimize the gap between the observed data points and the line.

The basic form of a linear regression model is represented by the equation:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \varepsilon$$

where x1, x2,..., xn are the independent variables and y is the dependent variable. Additionally, ε is the error term, and β are the coefficients, which stand for the linear relationship's intercept and slopes.

Many disciplines, including economics, finance, the social sciences, and machine learning, employ linear regression extensively for projects like forecasting sales, home values, stock prices, and a wide range of other continuous outcomes. It is an easy-to-use yet effective method that lets you make predictions based on observed data and offers insightful information about the correlations between variables. We can use this algorithm on Crime rate prediction where we foresee the rate of a dependent variable on the bases of the multiple independent variables which give us a result on testing as given below in figure 2.

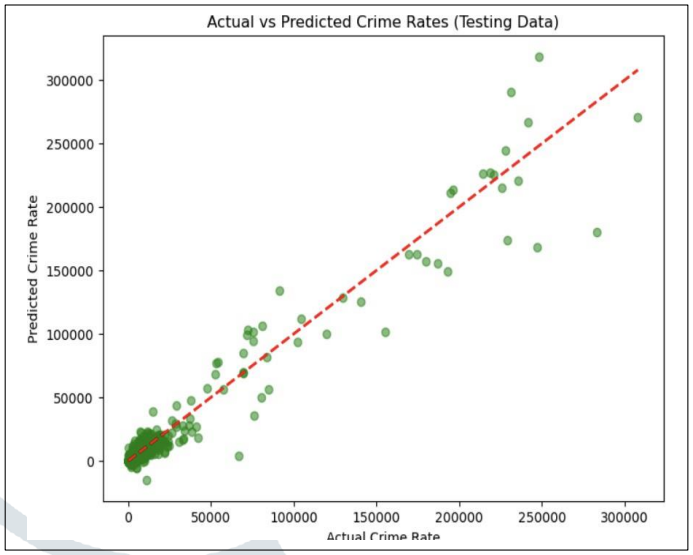

Figure: 2

3. Decision Tree

Decision Tree algorithms are intuitive and easy to interpret models that partition the feature space into hierarchical decision rules based on the most informative features.

Decision trees are useful for investigating intricate linkages and spotting nonlinear correlations between predictors and crime rates since they can manage both numerical and category data with ease.

Law enforcement organizations and legislators can obtain practical insights into the temporal and spatial patterns of criminal behaviour by utilizing Decision Tree algorithms in crime rate prediction models. This allows for the development of resource allocation plans and targeted interventions that improve community safety and well-being.

As we can use the Decision tree algorithm for predicting the crime rates with different features as nodes of the decision tree which give result as given in below figure 3.

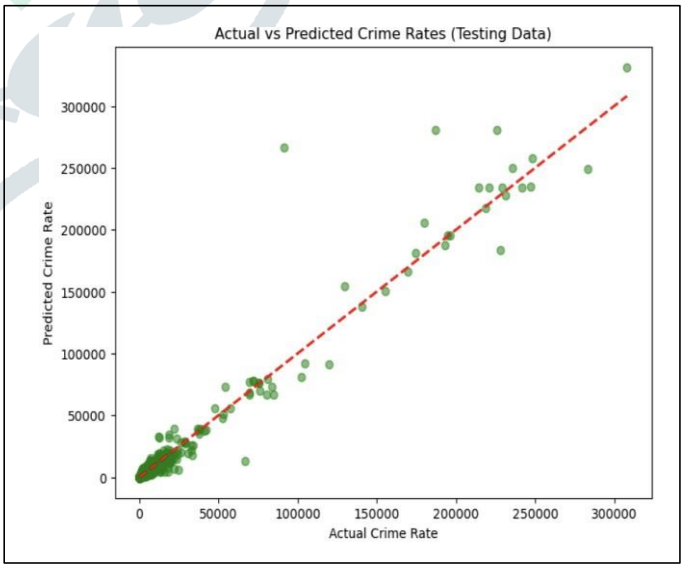

Figure: 3

Model	Predicted Value	Variation
Linear Regression	15921.91786263	8905.91786263
Random Forest	8333.48	1317.48
Decision Tree	8547	1531

With Decision Tree model, giving the result closest to the original value and the least variation, we can see that this model is superior against the others for crime rate prediction.

V. Conclusion and Future Scope

In conclusion, crime rate prediction powered by machine learning and spatial analysis offers invaluable insights for law enforcement and policymakers. Advanced algorithms like linear regression and Random Forest, coupled with integration of diverse data sources, enable accurate predictions by identifying influential factors. Continuous refinement of models and collaborative efforts between data scientists, law enforcement, and communities are vital for enhancing prediction accuracy and effectiveness. The future holds promise with advancements in real-time data integration, predictive policing technologies, and interdisciplinary research. By harnessing these developments, we can create proactive crime prevention strategies, fostering safer communities and enhancing public safety.

The future scope for crime rate prediction is promising, with advancements in machine learning, artificial intelligence, and data analytics. Integration of real-time data streams, such as social media and IoT devices, can enhance prediction accuracy and enable proactive crime prevention measures. Furthermore, the adoption of predictive policing technologies and the development of user-friendly crime prediction tools accessible to law enforcement agencies and communities hold significant potential. Additionally, interdisciplinary research collaborations focusing on understanding the socio-economic and behavioral factors influencing crime rates will further refine predictive models. Overall, the future holds opportunities for more sophisticated and effective crime prediction systems that contribute to safer and more secure communities.

The future scope of our crime rate prediction project involves exploring ways to make our model even better and using it to help communities stay safe. This could mean improving the accuracy of our predictions by adding more data or refining our algorithms.

We could also expand our project to cover different areas or time periods to see how crime trends vary. Overall, our goal is to keep using data and technology to help prevent crime and make neighborhoods safer for everyone.

VI. References

[1] Mehmet Sait, and Mustafa Gök. "Criminal prediction using Naive Bayes theory." Springer 28.9 (2016): 2581-2592.

[2] Sivaranjani, S., S. Sivakumari, and M. Aasha. "Crime prediction and forecasting in Tamilnadu using clustering approaches." Emerging Technological Trends (ICETT), International Conference on. IEEE, 2016

[3] Mahmud, Nafiz, et al. "Crimecast: A crime prediction and strategy direction service." 19th International Conference on. IEEE, 2016.

[4] Tahani Almanie, Rsha Mirza, Elizabeth Lor, "Crime prediction based on crime types" International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.5, No.4, July 2015.

[5] Thongsatapornwatana U. A survey of data mining techniques for analyzing crime patterns. InDefence Technology (ACDT), 2016 Second Asian Conference on 2016 Jan 21 (pp. 123-128). IEEE.

[6] Vedhadharshan B, V.K, H.S, T.P. Chand, R. Kaladevi, and B.A. "Crime analysis using linear regression" 2022 published in 2nd International Conference on Innovative Sustainable Computational Technologies (CISCT)

[7] Cesario, Cesario E, Catlett C, Talia D "Forecasting Crimes Using Autoregressive Models" 2016 IEEE 14th Intl Conf on Dependable, Autonomic and Secure Computing

[8] Zhang Q, Yuan P, Zhou Q, Yang Z. Mixed spatial-temporal characteristics based Crime Hot Spots Prediction. InComputer Supported Cooperative Work in Design (CSCWD, 2016 May 4 (pp. 97-101). IEEE.

[9] Sathyadevan, Shiju, and Surya Gangadharan. "Crime analysis and prediction using data mining." Networks & Soft Computing (ICNSC), 2014 First International Conference on. IEEE, 2014.

[10] Yu CH, Morabito M, Ding W, Ward MW Chung-HsienY. et al."CRIME FORECasting Using Data Mining Techniques" Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on, Vancouver, BC, Canada, December 11, 201

[11] Anton Emmanuel et. al "Tough on crime, tough on the cases of the crime" , https://doi.org/10.7861/clinmed.ed.23.4.1.

[12] Wang, Yuqian, et al. "Crime prediction using spatial autocorrelation analysis, hot spot analysis, and neural network models: A GIS approach." Computers, Environment and Urban Systems, vol. 36, no. 3, 2012, pp. 207-217.

[13] Adnan, Muhammad, et al. " Cybercrimes and Role of Law Enforcement Agencies: A Critical Analysis", October 2021. Journal of Educational Management & Social Sciences 2(1):79-89

[14] Srivastava, Rajeev, and Sangeeta Mittal (2013). "Use of Classification, Clustering, and Association Rule Mining Techniques in Crime Data Analysis." Journal of Crime Analysis

[15] Khan, Fahad Shahbaz, et al. (2016). "Analyzing Crime Data and Creating Predictive Models Using Ensemble Methods and Supervised and Unsupervised Learning Algorithms.