Available online @ https://jjem.jnnce.ac.in
https: www.doi.org/10.37314/JJEM.SP0208
Indexed in International Scientific Indiexing (ISI)
Impact factor: 1.395 for 2021-22
Published on: 08 December 2024

Predictive Modeling of Criminal Activity Hotspots

Maithili Bhat[1*], Arunkumar K L[2],

[1*]Student, Department of Computer Applications, Assistant Professor, [2]Department of Computer Applications,

JNN College of Engineering, Shimoga

maithlibhat31@gmail.com,arunkumarkl@jnnce.ac.in

Abstract

In our society the major cause in disturbing the peace of the country is Crime. The rate of crimes in society is an ever-increasing problem. Crime is a deliberate act of physical or physiological harm, damage or loss. Early prediction of the crime hotspot may help the law enforcement to take necessary action; in this regard the researchers are working to identify the crime zone which is affecting the peace and morality of the society using machine learning. Crime prediction helps the law enforcement committee to identify the most crime prone subjected areas and its pattern of crime. The previously recorded datasets helps in finding the pattern of the crime and speculate which are the high risk areas with the style of crime. This can be done using a KNN algorithm has potential to find spatial and temporal patter which makes it simple and transparency. KNN is ease to deploy with minimum training dataset. Our approach will help to find the crimes and its awareness in the country. Our work gives the 60-80% accurate results and can predict the crime hotspot. This work shows the potentiality of crime clustering, providing a solid foundation for next studies and useful applications in crime prevention techniques.

Keywords: *K-means, One-hot encoding, label Encoder, Elbow method, K-NN.*

1. Introduction

A crime is intentional act of physical or psychological harm, which damage the once mental health and peace in life. There are lots of crimes happening in our environment where everyone would have undergone through this once in their lifetime. Crime affects people from all backgrounds, locations and ages. According to the National Crime Records Bureau (NCRB), India's crime rate was422.2 per lakh population in 2022, down from 445.9 in 2021. After analyzing the data, police force agencies can forecast about future criminal activities. The study diverse with multiple forms which help the law enforcement agency to allocate the resources and prevent the crime from occurring.

The vast datasets analyzes the crime pattern identification, predictive policing and for effective law enforcement strategy, resource optimization of machine learning is used. A sub branch of AI which imitates the human behavior is called the machine learning. Predictive analytics can include machine learning to analyze data quickly and efficiently. The KNN is used all over in predicting the crime as it is simple, adaptive and yet more effective algorithm. It is undemanding to deploy and develop because of its straightforward nature, which requires a minimal assumption about the data distribution. This transparency helps the law enforcement committee to get a clear communication of results. k-NN is used in capturing the complex and non linear patterns of the crime data, where it is determined by the socio-economic and environmental factor. This algorithm focuses on local neighborhoods, which helps in identifying the crime hotspots with recent incidents and local conditions. Algorithm is said flexible in handling both regression and classification tasks. Classification task predicts whether a crime takes place in specified area likewise the regression task, estimates the wholesome of crimes in an area. Additionally, k-NN encompasses spatial coordinates and temporal factors results in anticipating evaluation. The forecasting of crime hotspot facilitates the visualization and analysis by Geographical Information System(GIS) which is integrated by spatial. The law enforcement agencies allocate resources more efficiently by the data-driven approach which helps in the implementation of targeted interventions and the addressing of potential crime areas enhances the safety of the public.

2. Literature Survey:

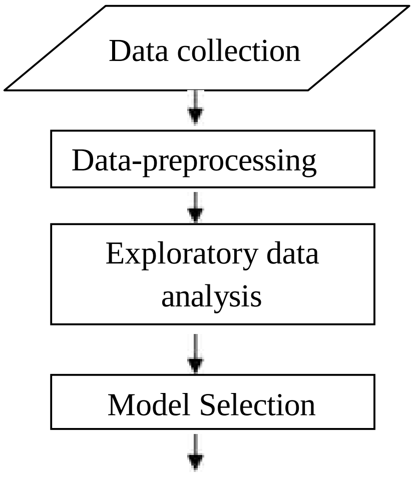
Akash Kumar et.al [1] works on K-Nearest Neighbor algorithm in forecasting crime. This project composed through data collecting and preprocessing by KNNs pattern recognition skills.F1-score, accuracy, precision, and recall matrix helps in the implementation of K-NN. By using this algorithm the author says that this is the best outcome when compared to other models in competitive manner. The crime prediction and data analysis on exploring the clustering algorithm that is Fuzzy C-Means (FCM) by B. Sivanagaleela and S. Rajesh[4]. The author described how the FCM has effectively identified and analyzed by enabling the prediction of crime hotspot in future. The algorithm highlights the complex datasets and handles the large data set of crime prediction by improving the accuracy and reliability. Krishnendu S.G et.al [9] proposed K-means clustering algorithm for analyzing and predicting crime patterns. It is acceptable for large datasets by enhancing the quality of the clusters and reduces the computational complexity using the optimized algorithm. The work can predict future occurrence by analyzing the historical data, which helps the law enforcement agencies to prevent the crime on resource allocation. Karabo Jenga et.al [10] this paper composed the various machine learning techniques in predicting the criminal activities. The prediction can be done using decision trees,(SVM), neural networks, and ensemble methods using ML. The challenges faced in this domain are privacy concerns, data imbalance, and the need of real time prediction. Nurul Hazwani et.al[11]composed a comprehensive review of various techniques use to predict criminal activities. The author worked on the decision tree, neural networks, regression as key methods and K-Means including Fuzzy C-means are used as clustering algorithm. The geographic information system (GIS) signifies an vital role in visualizing and analyzing the spatial crime patterns. Romika Yadav and SavitaKumari,

[14] describes the Autoregression technique to predict the crime using the Time Series Data. The researchers have proctored many prediction techniques before using autoregressive model. These include the MAE and RMSE to show the effectiveness of the

model. The AR model outperforms as the best method in the small limit when compared with other. Also notes that the work on Time Based Series should increase to resolve the ethical issue and integrate the data in real time. The employ of machine learning in forecasting and examining the crime trends is discussed by Suhong Kim et.al[17] Data collection includes the methods like decision trees, support vector machines, and neural networks to implement and essay assess. This performs a major role in data quality and requirements for useful application. Varshitha D N et.al [18] works on Although the study of different ML techniques like decision trees and neural networks results in promising accurate prediction even though the statistical method is simpler. To improve the performance the strategies are combined into hybrid modes. The conclusion of the paper suggests that choosing the right methodologies with data qualities and operational requirements.

3. Methodology:

The analysis of crime prediction by law enforcement helps to find the highest crime prone areas. This can be done using the historical dataset containing the states and types of crime which predicts the high subjected crime area of India.



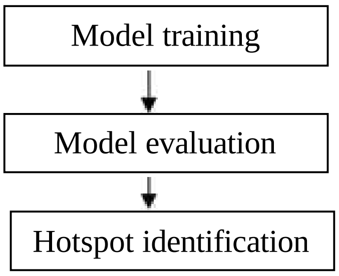


Figure1: Crime analysis flowchart

Data collection: In this section a relevant data collected from the large amount of historical data. The essay way is to collect the data from law-enforcement or from the kaggle dataset named State_wise_crime_India_2001_2012.

Data preprocessing: The collected datasets are formatted in csv.file type and performs these 3 different steps: read the dataset, manage the missing data and replace the missing entries using Mean, One hot encoding which helps in effective crime prediction system.

Mean: This helps in imputing missing values of numeric columns. If there are outliers, then this method is not appropriate, because our dataset is normal and the missing values are random. Through this method is said to finest in preprocessing the data.

On hot encoding: This is a machine learning used to convert the categorical data into numeric data. These assign 1 or 0 value to the column and each integer values are represented as a binary vector that entirely a value zero except for the index is marked as 1. The model improves its accuracy by allowing it to understand and differentiate between various states, includes the multiform unique crime pattern.

Exploratory Data Analysis: This contains the elbow method which helps in clustering analysis for determining the optimal number of cluster dataset. Elbow method is non parametric and do not make any sub-positions in primary data. This recognizes the regions having maximum crime rates or specific crime patterns.

Model selection: The following trend used is clustering, by using a method K-Mean. K-Means is technique of grouping the similar crime incidents or areas into clusters based on their characteristics. This helps in identifying the highly concentrated crime areas by revealing the pattern in the type of crime and their location.

Model training: The label encoder algorithm builds each cluster and trained completely. The K-Mean model is selected as a model and trained it by the algorithm label encoder, which ensures that the labels are encoded in sequential integer starting from 0 and then data is cut for model evaluation.

Model evaluation: once the data is transformed by splitting it into the clusters, the trained data is evaluated whether the data trained is meeting the expected crime rates. This can be done using the standardization method. In this, models are trained constantly without a bias of feature scale.

Hotspot Identification: K-NN (K-nearest neighbor) is one of the simplest Machine learning technique based on supervised learning. This shows the similarity between the new data case and available data case. The new data case is compared with that of the available categories. As this is non-parametric it doesn’t make any assumption or doesn’t learn from the training set, instead it stores the dataset at a time of classification and action is performed, so this is also called as lazy learner algorithm. With the help of K-NN we can easily identify the category or class of a particular dataset.

4. Implementation

The initial step is collecting the data from the Kaggle dataset named State wise crime India 2001-2012.Then the data is preprocessed through the collected dataset in 3 steps contains reading the dataset, managing the dataset, replace the missing entities using a method. The methods used are mean and one hot encoding. In mean the missing values are turn into numerical data, in the same way one hot encoding converts the dataset from categorical data to numeric data and this allows the dataset to understand and differentiate between the historical dataset place contains multiple crime pattern. The third step of the implementation is to cluster the optimal number of dataset. By optimizing the dataset they do not form any sub-position in dataset. Another method used in grouping of similar crime incident in accordance with the characteristics is called K-Means. Theses similar datasets are trained using a method label encoder which checks for the sequential number starting with 0. The trained dataset is then clustered by splitting it and standardizing it to avoid bias feature.

Next step is to create the ANN model. Before creating it initialize the ANN with essential libraries that is tensorflow keras, then input layer is added to the hidden layer of the data, this informs the model about the structure of the data. The same step is followed by the second layer and the output layer. Before training the model, this needs to be specify how the model should be trained and evaluated, the evaluation can be done using the following formula

$$Loss=-[ylog(p)+(1-y)\log(1-p)] \quad (1)$$

Where y is true label (0-1) and p is predicted probability using 1.

After this the dataset is trained which converts the NumPy to specify the correct data type and prediction on the test data is done. The forecasting labels are then differentiated with actual label which are concatenated and prints it aside. Prior to testing of accuracy the confusion matrix results in true positive, true negative, false positive and false negative prediction by understanding the performance of the classification. The calculation of accuracy is done using the following formula.

$$Accuracy =TP+TN+FP+FNTP+TN \quad (2)$$

The next step is the most important step where K-NN is applied to predict the best accuracy rate of the crime. K-NN finds the crime areas based on the clusters created by dataset. The K-NN trains and tests the dataset with prediction value containing confusion matrix in it. The models result represented in the below table:

Sl.No	Dataset	Algorithm	Accuracy
1.	Sate wise crime India 2001-2012	ANN	86%

Table1: Result comparison

5. Experimental Results:

The proposed system has resulted with the expected outputs through the csv.file applied to the state wise crimeIndia2001-2012.By applying the K-means to the dataset the bar graph is plotted comparing the state crime rates.

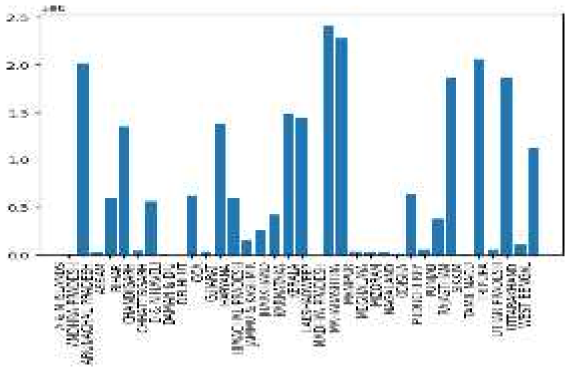


Figure 2:K-Means safe and unsafe zone of all state crime rates

Even the Density-Based Spatial Clustering Of Applications With Noise is applied on the similar dataset collected from the kaggle named state wise crime India 2001. The clustered crime incident data can identify the crimes which are highly concentrated and results in the below graph.

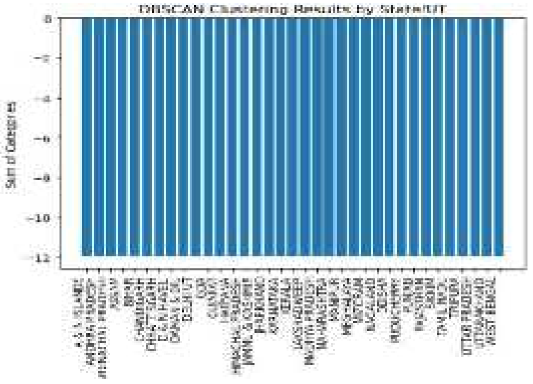


Figure 3: DBSC showing all the unsafe zone state which is in high crime rate

6. Conclusion:

The paper brings to a close comparison of two methods, the K-NN predicts that bars with high range are unsafe zone and bars with range low are considered as safe. Likewise when applied the DBSAN method the graph plots by predicting all the states are unsafe. So we conclude that K-Means is the best method as it finds the centric within the minimum variance clusters which identifies in high concentrated area of crime and predicts the graph with safe and unsafe zone.

7. Reference

1 Akash Kumar, Aniket Verma, Gandhali Shinde, Yash Sukhdeve, and Nidhi Lal,“CrimePredictionUsingK-Nearest Neighboring Algorithm”IEEE International Conference on Emerging Trends in Information Technology and Engineering 2020.

2 Arunkumar K L, Ajit Danti "A NOVEL APPROACH FOR VEHICLE RECOGNITION BASED ON THE TAIL LIGHTSGEOMETRICALFEATURESIN THE NIGHT VISION", International Journal of Computer Engineering and Applications, Volume XII, Issue I, Jan. 18, www.ijcea.com ISSN 2321- 3469

3 Manjunatha H T Arunkumar KL, Ajit Danti, "A Novel Approach for Detection and Recognition of Traffic Signs for Automatic Driver Assistance System Under Cluttered Background" Springer 1035 (CCIS), pp 407-419

4 B. Sivanagaleela and S. Rajesh "Crime Analysis and Prediction Using Fuzzy C-Means Algorithm" Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI 2019).

5 C. C. K. Jenga and G. Ka, “Machine learning in crime prediction.” https://www.ijrpr.com/, Feb 2023.

6 Arunkumar K L, Ajit Danti, Manjunatha HT , D Rohith, “Classification of Vehicle Type on Indian Road Scene Based on DeepLearning", ,Springer,Singapore1380(2021),1-10

7 Arunkumar K L, AjitDanti, ManjunathaHT Estimation of vehicle distance based on feature points using monocular vision", IEEE 8816996(2019),1-5

8 Manjunatha H T, Ajit Danti, ArunKumar K L, D Rohith "Indian Road Lanes Detection Based on Regression and clustering using Video Processing Techniques", , Springer, Singapore 1380 (CCIS), 193-206

9 Krishnendu S.G, Lakshmi P.P, and Nitha L "Crime Analysis and Prediction using Optimized K-Means Algorithm" International Journal of Engineering Research &Technology(IJERT)ISSN:2278-0181 Publishedby, www.ijert.orgNCETEIT – 2017.

10 KaraboJenga,CagatayCatal,andGorkem Kar "Machine Learning in Crime Prediction" Journal of Ambient Intelligence and Humanized Computing (2023).

11 Nurul Hazwani Mohd Shamsuddin, Nor Azizah Ali, and Razana Alwee "An Overview on Crime Prediction Methods"6th ICT InternationalStudent Project Conference (ICT-ISPC)2017.

12 CM Nrupatunga, KL Arunkumar, “Peruse and Recognition of Old Kannada Stone Inscription Characters”, Springer, Singapore, 2020

13 Arunkumar K L, Ajit Danti, "Recognition of Vehicle using geometrical features of a tail light in the night vision", National Conference on Computation Science and Soft Computing (NCCSSC-2018)

14 RomikaYadav and SavitaKumari, “Crime Prediction Using AutoRegression Techniques

for Time Series Data” IEEE 3rd International Conference and Workshops on Recent Advances and Innovations in Engineering, 22-25 November 2018.

15 R. M. Saeed and H. A. Abdulmohsin, “A study on predicting crime rates through machine learning and data mining using text,” JournalofIntelligent Systems,vol.32,no.1,p. 20220223, 2023

16 R.J.D.D.S.R.G.SomaSekhar, Puvvada Abhinaya, “Criminality data scrutiny using logistic regression algorithm,” IEEE Xplore, 2023

17 Suhong Kim, Param Joshi, Parminder SinghKalsi, and Pooya Taheri's study "Crime Analysis Through Machine Learning" 978-1-5386-7266-2/18/$31.00 ©2018 IEEE.

18 Varshitha D N, Vidyashree K P,Aishwarya P, Janya T S, K R Dhananjay Gupta, and Sahana R.'s work "Different Approaches for Crime Prediction System" International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 Publishedby, www.ijert.orgNCETEIT – 2017.