

A Clustered Based Approach for Prediction of Crime Hotspots Using Machine Learning

¹**Dr.M.V.Rajesh**, Professor, Dept of Information Technology , Aditya Engineering College(A),
Surampalem

²**B. Revathi** , Department of Information Technology, Aditya Engineering College, Surampalem,
(Affiliated to JNTUK, Kakinada)

³**D. N. SivaSri** ,Department of Information Technology, Aditya Engineering College, Surampalem,
(Affiliated to JNTUK, Kakinada)

⁴**M. Satya Sri** ,Department of Information Technology, Aditya Engineering College,
Surampalem, (Affiliated to JNTUK, Kakinada)

⁵**P. Eswar Vijay**,Assistant Professor, Department of Information Technology, Aditya Engineering
College, Surampalem, (Affiliated to JNTUK, Kakinada)

Abstract

The project is entitled with “A Clustered Based Approach for Prediction of Crime Hotspots Using Machine Learning.” The main objective of this project is to implement machine learning techniques to predict the crime hotspot. The final result of this project is to efficiently analyze and predict the crime hotspot. In this project we propose machine learning algorithm K-means for crime hotspot prediction.

Introduction

In the present situation crime rates are increasing rapidly. Most of the harmful effects is caused due to serious crimes. A crime is an unlawful act which is punishable by a state or other authority. It is an act which is harmful not only to individual but also to community, state or society. A crime hotspot is generally defined as an area containing huge number of criminal activities. So identification of crime hotspots helps public safety sectors to allocate various measures for crime prevention.

Statement of the Problem

- To overcome these problems in existing system, we are proposing a crime prediction and monitoring framework.
- To analyze the data in a better way K-Means machine learning technique is used.
- Crime dataset used in this work is reliable, real, and authentic as data is acquired from the official site of Indian government.

- The dataset contains nearly 10000 records and different attributes like location, year and crime type.
- It takes history of crimes from past few year's as training dataset.

Advantages:

- It can accurately determine the crime hotspots.
- In this algorithm will help to predict the crime hotspots and fast up the process of solving crime.

Counter measures are chosen to prevent crimes.

Literature Survey

The various researchers have addressed the problems regarding crime control and have proposed different crime-prediction algorithms. The accuracy of prediction depends on the attributes selected and the dataset used as a reference.

Rizwan Iqbal [1] has proposed
“**Experimental study of**

classification algorithms for prediction” gives comparison between two classification algorithms, Decision Tree and Naive Bayesian, was performed using WEKA, an open-source data mining software, and 10-fold cross-validation. The socio-economic, law-enforcement, and crime data sets for this study were compiled from the 1990 US Census, the 1990 US LEMAS survey, and the 1995 FBI UCR, respectively. Experimental results shows that decision tree has better performance for crime dataset, using WEKA.

Shyam Varan Nath [2] proposed a **“Crime pattern detection”** using K-means, Law Enforcement, Semi-supervised learning. The authors discussed about the use of clustering algorithm for detecting the crime pattern and used semi-supervised learning for knowledge discovery. This technique is used to identify the crime patterns from a large number of crimes and makes job of crime detectives easy. One of the limitations is data mining is sensitive to quality of input data which is inaccurate.

Dr. Zakaria [3] proposed a **“Data Mining Techniques to Analyze Crime patterns in the Libyan National Crime Data”** for crime data and criminal and analyses using simple K-means algorithm for clustering and Apriori algorithm association rules. The goal of this project is to help the government of Libya to take strategic decision about prevention in the increasing crime rates. They have used WEKA software and Excel software to preprocess and analyse the collected crime data as well as criminal data. K-means clustering algorithm gives the overall statistical knowledge the criminal age and crime type.

A major challenge regarding crime prediction is analyzing large crime datasets accurately and efficiently. Data mining is utilized to find hidden patterns in large crime datasets quickly and efficiently. The increased efficiency and reduced errors in crime data-mining techniques increase the accuracy of crime prediction.

Md Abu Saleh [4] proposed **“Crime Data Analysis”** use machine learning K-means algorithm for prediction and to analyze the crime in Chicago town and the dataset is accessible at Kaggle website. The next intention of this project is to check the feasibility of k-means algorithm to determine and solve the crime problems. In this, attributes focused for analysis are type, time, location of crime and arrest made and results are represented using 2d graphs.

Tahani Almanie [5] proposed **“Crime Prediction Based on Crime Types And Using Spatial And Temporal Criminal Hotspots”** which is focused to find criminal hotspots. They made use of Apriori algorithm for the production of interesting frequent pattern to observe the criminal hotspot and they used decision tree and naive Bayesian classifier for predicting crime categories. They applied the algorithms for Denver crime data and Los Angeles crime data. They combined the findings of Denver crimes datasets with demographics information for further analysis.

Alkesh Bharati [6] has proposed a **“Crime Prediction and Analysis Using Machine Learning.”** which predicts the type of crime in a particular area and improve the accuracy using machine learning algorithms like logistic regression, Decision tree, Random forest classification, KNN etc. In this it takes

Chicago crime dataset which is extracted from the official portal of Chicago police. This will give a clear idea of how machine learning can be used by the law enforcement agencies to detect, predict and solve crimes at a much faster rate and thus reduces the crime rate. Using the concept of machine learning they have built a model using training dataset which has undergone data cleaning and transformation. The model predicts the type of crime with accuracy of 78.9.

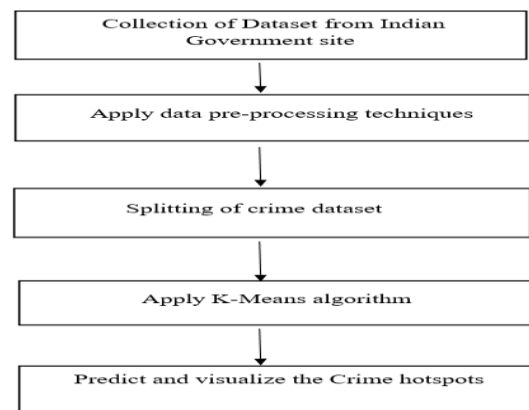
Prajakta Yerupude [7] proposed a “Predictive modeling of crime dataset” using Decision tree, Naive Bayes, Random forest, Linear Regression. In this communities and crime dataset from UCI repository is used. The authors explained, algorithms like decision tree, Naive Bayes are applied on the dataset to predict features that affect the high crime rate. Supervised and Unsupervised learning are used to get the output. Random forest gives most balanced results when compared to other algorithms.

Md. Abdul Awal [8] has proposed “Using Data Mining Technique to Analyze Crime of Bangladesh”. It uses linear regression algorithm to predict future crime trends in Bangladesh. The algorithm helped in forecasting the crimes such as kidnapping, women and child repression, robbery, burglary, dacoit theft and murder. The paper is concluded from the result has been categorized into two parts such as metropolitan region result and divisional region result for better and easier understanding. It is from the result as most of the crimes are increasing with the growth of population.

Deepika K [9] has proposed “Crime analysis in India using data mining”. It includes steps such as data pre-processing, clustering, classification and visualization.

Crime is identified using k-means clustering and the clusters are formed based in the similarity of the crime attributes. The Randomforest algorithm and neural networks are applied on the data for classification. Visualization is achieved using the Google marker clustering and the accuracy is verified using WEKA tool. The proposed model is very useful for both the investigating agencies and the police officials in taking necessary steps to reduce crime.

Research Methodology Architecture:



Modules

Our project includes 2 basic modules they are

1. Admin
2. Crime Hotspot Prediction
3. Analysis

Admin

Store User Details

Admin login to the application and he performs certain operations like storing details of the all the registered users.

Crime Hotspot Prediction

State-Wise Prediction

State-wise prediction provides crime hotspot areas predicted in a state. For this first we need to select state and then crime type and click on go. It will display the

heat map of the selected state with crime hotspot area being focused.

Country-wise Prediction

Country-wise prediction provides crime hotspot areas predicted in a country. For this first we need to select crime type and click on go. It will display the heat map of the country with crime hotspot area being focused.

Crime Analysis

Women crime analysis

Women crime analysis provides analysis of crimes against women. In the present situation we see that crime against women occurs for every three minutes. So we are providing analysis of them to create awareness among people. In this we need to select crime type, year and district. After this when we click on go button it will provide graph analysis regarding crime against women.

Children crime analysis

Children crime analysis provides analysis of crimes against children. In this we need to select crime type, year and district. After this when we click on go button it will provide graph analysis regarding crime against children.

The following stages will help us understand how the K-Means clustering technique works. K-means clustering is one of the methods in cluster analysis which aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean.

Process:

- Initially, the number of clusters must be known let it be k .
- The initial step is to choose a set of K instances as centers of the clusters.
- □ Next, the algorithm considers each

instance and assigns it to the cluster which is closest.

- The cluster centroids are recalculated either after whole cycle of reassignment or each instance assignment.
- This process is iterated.

K-means implements the Expectation-Maximization strategy to solve the problem. The Expectation-step is used to assign data points to the nearest cluster, and the Maximization-step is used to compute the centroid of each cluster.

When using the K-means algorithm, we must keep the following points in mind:

- It is suggested to normalize the data while dealing with clustering algorithms such as K-Means since such algorithms employ distance-based measurement to identify the similarity between data points.
- Because of the iterative nature of K-Means and the random initialization of centroids, K-Means may become stuck in a local optimum and fail to converge to the global optimum. As a result, it is advised to employ distinct centroids' initializations.

Implementation of K Means Clustering Graphical Form

STEP 1: Let us pick k clusters, i.e., $K=2$, to separate the dataset and assign it to its appropriate clusters. We will select two random places to function as the cluster's centroid.

STEP 2: Now, each data point will be assigned to a scatter plot depending on its distance from the nearest K -point or centroid. This will be accomplished by establishing a median between both centroids.

STEP 3: The points on the line's left side are close to the blue centroid, while the points on the line's right side are close to the yellow centroid. The left form cluster

has a blue centroid, whereas the right Form cluster has a yellow centroid.

STEP 4: Repeat the procedure, this time selecting a different centroid.

STEP 5: After that, we'll re-assign each data point to its new centroid. We shall repeat the procedure outlined before (using a median line). The blue cluster will contain the yellow data point on the blue side of the median line.

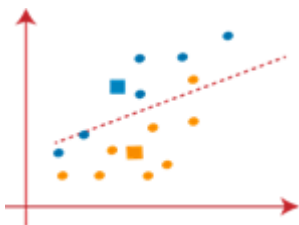


Figure: K-Means Graph1

STEP 6: Now that reassignment has occurred, we will repeat the previous step of locating new centroids.

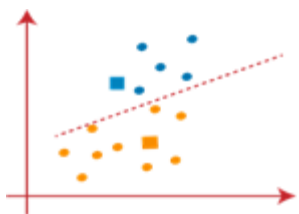


Figure: K-Means Graph2

STEP 7: We will repeat the procedure outlined above for determining the center of gravity of centroids, as shown below.



Figure: K-Means Graph3

STEP 8: Similar to the previous stages, we will draw the median line and reassign the

data points after locating the new centroids.

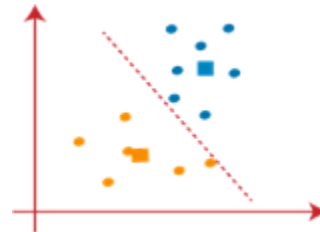


Figure: K-Means Graph4

STEP 9: We will finally group points depending on their distance from the median line, ensuring that two distinct groups are established and that no dissimilar points are included in a single group.

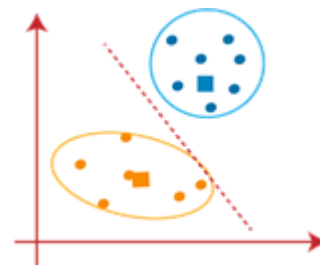


Figure: K-Means Graph5

The final Cluster is as follows:

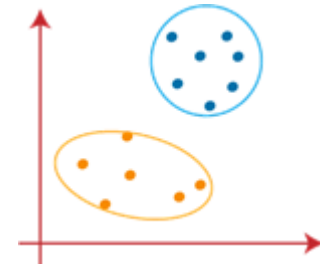


Figure: K-Means Graph6

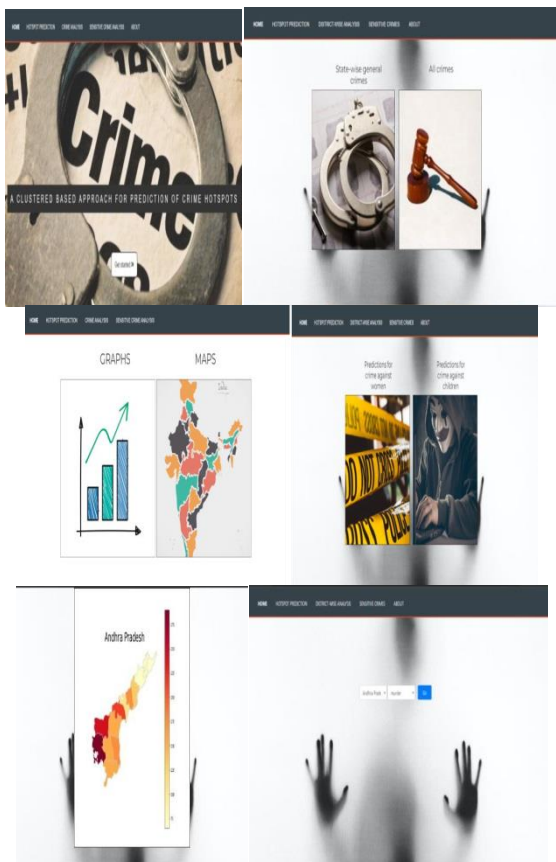
Crime Hotspot prediction using k-means clustering

The procedure is given below:

- First, we take crime dataset.
- Filter dataset according to requirement and create new dataset which has attribute according to analysis to be done

- Open excel and read excel file of crime dataset and apply and replace Missing value operator on it and execute operation.
- Perform k means clustering on resultant dataset formed.
- Then hotspots are visualized in heat maps.
- Analysis can be done on cluster formed.

Results and Discussion



Conclusion

- In this approach, we are developing a prediction method for detecting crime
- This initiative helps the law enforcement agencies to predict and analyze crimes in India with improved accuracy and thus reduces the crime rate.
- It will also create awareness among the society regarding dangerous areas in a country.
- In this project we proposed a method for crime hotspot prediction which is a vital asset for the police department as well as the society

References

1. Zhu, R., Zhang, X., Shi, M., & Tang, Z. (2020). Secure neural network watermarking protocol against forging attack. *EURASIP Journal on Image and Video Processing*, 2020(1), 1-12.
2. Rey, C., & Dugelay, J. L. (2002). A survey of watermarking algorithms for image authentication. *EURASIP Journal on Advances in Signal Processing*, 2002(6), 1-9.
3. Barni, M., Bartolini, F., Piva, A., & Salucco, F. (2002). Robust watermarking of cartographic images. *EURASIP Journal on Advances in Signal Processing*, 2002(2), 1-12.
4. Chen, N., & Zhu, J. (2007). A robust zero-watermarking algorithm for audio. *EURASIP Journal on Advances in Signal Processing*, 2008, 1-7.
5. Comesana, P., & Pérez-González, F. (2007). Breaking the BOWS watermarking system: key guessing and sensitivity attacks. *EURASIP Journal on Information Security*, 2007, 1-8.