

Crime Prediction & Prevention using Unsupervised Machine Learning

Prof. Rushali A.Deshmukh¹, Suraj Gaonkar², Vicky Khillare³, Aniket Rokade⁴

Department of Computer Engineering, Rajarshi Shahu College of Engineering¹²³⁴

*radesh19@gmail.com¹, surajgaonkar1998@gmail.com²,
khilarevicky19@gmail.com³, nabhsprushyadipam@gmail.com⁴*

Abstract

Crime rates are increasing day by day from the last 15 years in India due to various reasons. Maharashtra is a state in India's western peninsular region in which crimes are increasing from the last few years. We are trying to predict the rate of various types of crimes, such as robbery, theft, murder, etc. We will generate crime hotspots on a map of Maharashtra based on the rate of a particular crime according to the location of the crime. We are collecting crime data of the past 15 years (2004- 2019) from various mediums like newspaper, Kaggle, data.gov.in, etc. Our algorithm is k-means, which is the simplest clustering algorithm. It is helpful to establish different relationships, such as the age of criminals and the type of crime they committed. We will also add the reasons behind the crime, such as money, drug abuse, etc. Our project will help police authority as well as the government for taking proper action for the reduction of the crime. It will be helpful for the investors for investment in a particular area according to the crime rate of that area as well as it is also helpful for tourists to decide which area is safe for tourism.

Keywords: *K-means clustering, crime hotspots, crime prevention.*

1. Introduction

21st Century is the century of technology in which most of the things are digitalized. It is necessary to keep crime records in a digitalized format, which can be helpful for data analysis. With the help of high power computing machines, it is very useful for police authority for the effective prediction of crimes. Crime is one of the biggest problems in the country, affecting economic growth and development in the country, 5-10% of people commit 70% of crime. Different types of crimes committed by criminals from the last 15 years & their pattern of crime are also changing and difficult to explain. We aim to identify different types of crimes based on location, time, district, police station, age of criminal, etc. It will help police authorities increase patrol and increase CCTV cameras according to the crime rate in particular areas. Crime Prediction is the technique that will predict when the next crime will commit according to available data as well as with the help of statistical analysis. Crime prediction is helpful to decrease the crime rate. Machine learning plays an important role in crime prediction, helping the government decrease the crime rate. It helps in the prediction of crime before it happens. It is also useful for analyzing various crime patterns as well as for the classification of crime based on different locations. Crime hotspots are developed according to the rate of crime, which will help researchers and non-government organizations for further study.

2. Literature Survey

Researchers, as well as developers, use various techniques, tools for the prediction of the crime rate. The following are some reference papers which we can use to make our system more promising as well as efficient.

Suhong Kim et al.[1] proposed the KNN & boosted decision tree to forecast the rate of different types of crime. They used two different approaches. In that, each area and category of a crime had given a unique number. In another perspective, they gave a binary number for the particular region and the day where a particular crime was committed.

Sunil Yadav & Ajit Yadav[2] considered the last 14 years of data (2001-2014) & they developed the regression model. Various algorithms implemented on the available dataset by using Weka & R tool. Apriori algorithm was used for the generation of association rules between the attributes.

Romika Yadav et al.[3] used Auto Regression Technique for perfect prediction of crime with minimum error for such time-series data by identifying relationships between crime attributes. They aimed to find out the location of a particular criminal in advance with better accuracy.

Shoaib Khalid et al.[4] did their work in 4 phases. They collected the data with the help of a GPS survey. They used ArcGIS for analysis purpose. They used kernel density estimation for finding crime hotspots in urban areas. They followed the compstat model with correction for management of performance of the police authority.

I. Jayaweera et al.[5] presented the system to solve the problems in front of law enforcement such as large geographical diversity & changing patterns of crime etc. They generated a web-based system for hotspot detection & crime pattern visualization. They generated a focused crawler, a web crawler for the collection of web pages that satisfy some specific property.

B. Sivanagaleela[6] used fuzzy means, and fuzzy Means is mainly used for the clustering of data. It is widely used to understand various crime patterns. Fuzzy C handles the uncertainty condition of the data.

Xiangyu Zhao[7] proposed Spatio-temporal patterns in urban. Their methodology consisted of 3 stages: extraction of features, STF based strategy for the development of predictive model & last step was the forecasting of crime. Their Feature extraction process extracts the features from a huge dataset such as dataset of various crime complaints & so many things.

Alkesh Bharati and Sarvanguru[8] introduced how machine learning can be used to solve crime in less time and how faster crimes can be solved, which can help police authority decrease the crime rate. Comparative study of different algorithms such as KNN, support vector machine, etc.

Vrushali Pednekar et al.[9] proposed the use of KNN for the prediction of crime. Available information is used to extract various crime patterns. The main focus was to analyze the rate of crime type according to their location. Weighting technique is used to remove low-value attributes in order to compete with the high dimension of data.

L McClendon et al.[10] used linear regression, Additive regression & Decision stump algorithm to analyze various violent crime patterns. Linear regression was best among all regression techniques in terms of accuracy. Five metrics were considered to check the effectiveness of algorithms such as the Correlation coefficient, Mean absolute error, Relative absolute error, etc.

3. Existing System

In existing system classification algorithms, such as KNN, boosted decision tree, etc. are used. Crime rate prediction accuracy is less as well as a dataset, or the data considers historical data such as (2000-2015). The latest data or latest crime types were absent in their system. Reasons behind crimes such as drug abuse, money, political issues, etc. were absent. They were not displaying the age group who is doing more crime. A conclusion like increase patrolling in a high crime rate area was not there. The existing system is time-consuming as well as not user-friendly.

4. Proposed System

We are going to use k-means, which is a clustering algorithm for the prediction of the crime rate. Our dataset is not in the generalized format. We are making our dataset by own. We are adding the latest data as well as different types of a new crime in the dataset. We are trying to add different types of crime information, including the type of crime, location of a crime, time of the specific crime, district of particular crime, police station near the location of a crime, etc. We also include victim details, including the name of the victim. For authentication purpose, we are generating a login page, including the name. Mobile number, Aadhar number, mail id, etc. OTP is sent to the user who is interested to see the rate of crime as well as the type of crime in a particular location. Our system is helpful for users like foreign tourists, investors, and police authorities for various purposes. We also include reasons behind the crime, such as drug abuse, money, extortion, etc. It helps to identify the reason behind the increasing crime rate & the government can able to take proper measures or actions to decrease the crime rate. We are identifying Crime hotspots in the map of Maharashtra according to the crime rate in a different region. A red box shows the more dangerous area as well as less dangerous areas according to the rate of crime. The more number of Red box, means higher the crime rate; less the number of Red box means lower is the crime rate.

5. System Architecture

1) Data Collection & preprocessing- We are collecting data from various sources such as newspapers, news channels, social media, etc. Our dataset has attributes like name of the criminal, area, type of crime, date of crime, age of the criminal, time of the crime, police station, district, victim details including the name of the victim, address of the victim, reasons behind the crime etc. preprocessing technique- we are using binarization technique to transform the data in binary format so that it can be processed easily.

2) Clustering–Clustering is used for similar grouping kind of things. We have formed 2 clusters along with the centroid of the cluster. We are forming clusters of types of crime & the number of times a particular crime is committed. We are also forming the clusters of the age of criminal & type of crime & every time centroid is updated for new data.

3) Prediction- We are also predicting each crime type's attention level, such as low, high, medium, based on the number of times a particular crime type takes place in a particular region. We are also predicting the age group who is doing more crimes.

4) Visualization- We used pictorial representation techniques such as bar graphs, pie charts, and histograms to display our results.

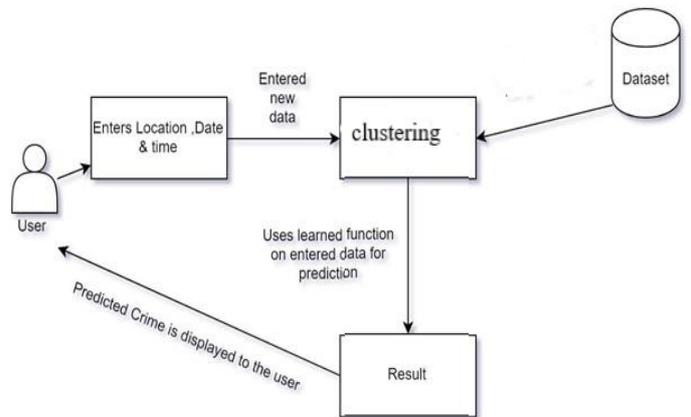


Figure 1. System Architecture for crime prediction

Algorithm: K-means

K means clustering is an unsupervised machine learning algorithm. It can work with the help of following steps:

- First, we randomly choose k data points from our dataset. This is nothing but initialization steps. We are taking the value of k=2 in our system. The centroid is the center of a cluster in this step.
- All data points which are very closest to the centroid that data points will form the cluster. We are using Euclidian distance for the formation of the cluster.
- We formed the cluster. We are updating the value of centroid every time when new entry or data comes. The new value of centroid is nothing but the average of all records available in the cluster.
- We are repeating step 2 and step 3 until the dataset doesn't become empty.

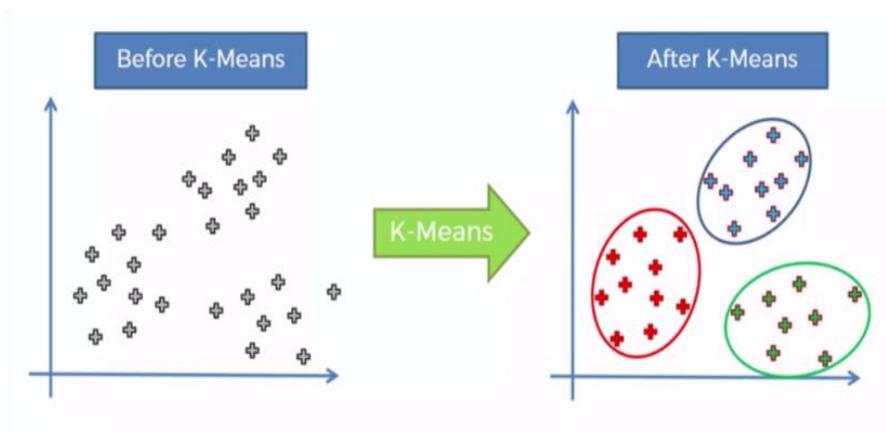


Figure 2. Simple Demonstration of K-means

Input – k is the number of clusters

D is the data elements

Output- clusters are formed according to time, location, type

Method – select k objects from data elements as centers of the cluster

Repeat – assign the next elements to the cluster with similar properties and repeat this until no data elements are remaining

Results

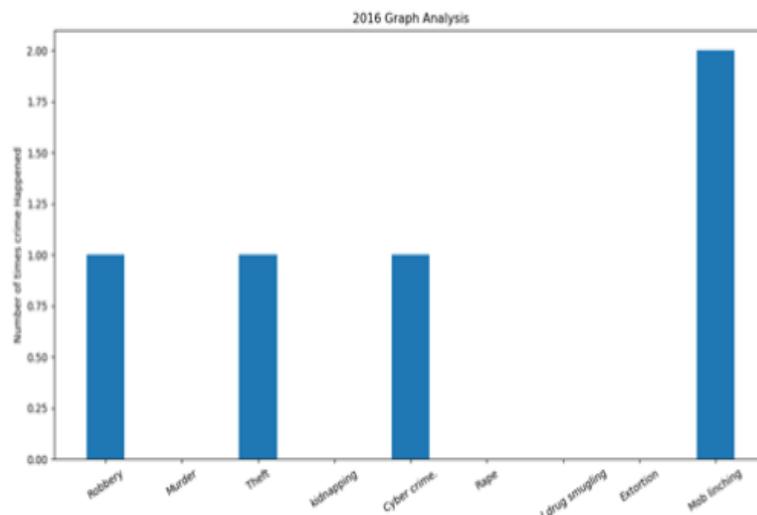


Figure 3. Type of crime & number of times particular crime happen

It shows that number of times particular crime types such as Robbery, Kidnapping, etc. happen in graphical format from the available set of records.

2016 Crime Analysis			2015 Crime Analysis		
Name of Crime	Percentage (of total)	Attention Level	Name of Crime	Percentage (of total)	Attention Level
Robbery	20.0	Low	Robbery	75.0	High
Murder	0.0	Low	Murder	0.0	Low
Theft	20.0	Low	Theft	0.0	Low
kidnapping	0.0	Low	kidnapping	0.0	Low
Cyber crime	20.0	Low	Cyber crime	0.0	Low
Rape	0.0	Low	Rape	25.0	Moderate
Illegal drug smuggling	0.0	Low	Illegal drug smuggling	0.0	Low
Extortion	0.0	Low	Extortion	0.0	Low
Mob lincing	40.0	High	Mob lincing	0.0	Low

Figure 4. Percentage of different type of crimes & its attention level

6. Conclusion & Future Scope

We conclude that by using the k-means algorithm, our system becomes more efficient as well as user-friendly & it will be helpful for the government, police authority to understand the various reasons behind the crime & it will also help for the reduction of the crime rate. It will be helpful for the tourist also to decide which area is safe for tourism as well as it is also helpful for investors. In the future, we can consider different states, cities, as well as a crime against women, small children, etc. We can also use the concept of face recognition if a person commits a crime before it happens. Crime pattern identification and pattern generation can be done while emphasizing on time; for example, we can generate patterns yearly, monthly, and weekly. Frequent criminals and types of crime can be identified.

References

- [1] Kim, S., Joshi, P., Kalsi, P. S., & Taheri, P. "Crime Analysis Through Machine Learning", In *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), IEEE*, (2018) November (pp. 415-420).
- [2] Yadav, S., Timbadia, M., Yadav, A., Vishwakarma, R., & Yadav, N. "Crime pattern detection, analysis & prediction", In *2017 International conference of Electronics, Communication and Aerospace Technology (ICECA) IEEE*, (2017) April (Vol. 1, pp. 225-230).
- [3] Yadav, R., & Sheoran, S. K. "Crime Prediction Using Auto Regression Techniques for Time Series Data", In *2018 3rd International Conference and Workshops on Recent Advances and Innovations in Engineering (ICRAIE) IEEE*, (2017) April (pp. 1-5). IEEE.
- [4] Khalid, S., Wang, J., Shakeel, M., & Nan, X. "Spatio-temporal Analysis of the Street Crime hotspots in faisalabad city of Pakistan", In *2015 23rd International Conference on Geoinformatics, IEEE*, (2015) June (pp. 1-4).
- [5] Jayaweera, I., Sajeewa, C., Liyanage, S., Wijewardane, T., Perera, I., & Wijayasiri, A. "Crime analytics: Analysis of crimes through newspaper articles", In *2015 Moratuwa Engineering Research Conference (MERCCon) IEEE*, (2015) April (pp. 277-282).
- [6] Sivanagaleela, B., & Rajesh, S.. Crime Analysis and Prediction "Using Fuzzy C-Means Algorithm", In *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI) IEEE*, (2019) April (pp. 595-599).
- [7] Zhao, X., & Tang, J. "Exploring transfer learning for crime prediction", In *2017 IEEE International Conference on Data Mining Workshops (ICDMW) IEEE*, (2017) November (pp. 1158-1159).
- [8] Vaquero Barnadas, M. (2016) "*Machine learning applied to crime prediction*" (Bachelor's thesis, Universitat Politècnica de Catalunya) (2016).
- [9] Pednekar, V., Mahale, T., Gadhve, P., & Gore, A "Crime Rate Prediction using KNN. *International Journal on Recent and Innovation Trends in Computing and Communication*, 6(1), (2018) pp. 124-127.
- [10] McClendon, L., & Meghanathan, N. "Using machine learning algorithms to analyze crime *Machine Learning and Applications*", *An International Journal (MLAIJ)*, 2(1), (2015) pp. 1-12.