Crime Mapping using Machine Learning Algorithm: K-means Clustering algorithm

Sreelakshmi R
PG Scholar
Department of Computer Applications
Amal Jyothi College of Engineering
Kanjirappally, India
sreelakshmir2023b@mca.ajce.in

Meera Rose Mathew
Assistant Professor
Department of Computer Applications
Amal Jyothi College of Engineering
Kanjirappally, India
meerarosemathew@amaljyothi.ac.in

***Abstract*-- This study focuses on the application of data analytics techniques such as clustering to aid in crime prevention and mitigation. The K-means clustering algorithm was utilized in this study after removing rows with missing data. The feature matrix was created using the latitude and longitude data of crimes, and the K-means algorithm was implemented with 5 clusters. The resulting clusters were added to the crime data and visualized on a map using the Folium library. This map can assist law enforcement agencies in identifying high crime areas and developing appropriate strategies. The study highlights the potential of data analytics and visualization techniques in crime prevention and suggests further research to determine the effectiveness of clustering algorithms in identifying crime patterns and trends in large datasets.**

Keywords— Data analytics, K-means clustering, Crime data, Visualization, Heatmap, Crime prevention

I. INTRODUCTION

Crime mapping is a crucial tool for law enforcement agencies to gain insights into the patterns of criminal activity and allocate resources effectively. One of the methods used for crime mapping is to apply machine learning algorithms like k-means clustering to detect clusters of crimes that have common features such as location and time of occurrence.

In this crime map, we use k-means clustering to detect hotspots of criminal activity in a specific city. The data set used for analysis includes the latitude and longitude location coordinates of various crimes that took place during a specific time frame.

By applying k-means clustering algorithm to the data, we can group the crimes into clusters based on their location, and subsequently, identify the areas of high crime activity. This analysis enables us to detect patterns and trends in criminal activities such as the type of crimes and their severity in specific regions of the city.

The insights gained from the analysis of the crime map can provide law enforcement agencies with the necessary information to allocate resources effectively and make informed decisions about the management of crime in high-risk areas.

II. LITERATURE REVIEW

Crime mapping has become a popular tool for analyzing crime patterns and trends in recent years. Various studies have examined its benefits and limitations, as well as its effectiveness in preventing and addressing crime.

The first study by Yadav et al. (2016) applied GIS and spatial statistics to analyze crime data in Delhi, India, from 2009 to 2012. The study found that certain areas in Delhi had higher crime rates than others and that crimes tended to occur in clusters. The study concluded that GIS and spatial statistics can be effective tools for crime analysis in India.

However, crime mapping has certain limitations that must be considered. Ensuring the accuracy and reliability of crime data is one of the main challenges. The second study by Khare et al. (2017) used GIS to analyze crime data in Mumbai, India, from 2010 to 2014. The study found that certain areas in Mumbai, such as commercial areas and railway stations, had higher crime rates than others. The study also found that crimes tended to occur at certain times of the day. The study concluded that GIS can be an effective tool for identifying crime hotspots in Mumbai.

Another limitation of crime mapping is the potential for bias in crime reporting. The third study by Singh et al. (2019) reviewed the application of machine learning algorithms for crime analysis in India. The study found that machine learning algorithms such as k-means clustering, decision trees, and neural networks can be effective for crime analysis in India. The study also identified challenges such as data quality and availability that need to be addressed for effective crime analysis using machine learning in India.

DOI: **10.5281/zenodo.7956631**
ISBN: 978-93-5607-317-3@2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam

K-means clustering is a widely used unsupervised machine learning algorithm that can be used to identify clusters of crimes based on their spatial and temporal patterns. Several studies have been conducted in India using k-means clustering to identify crime hotspots and patterns.

A study by Vardhan and Singh (2018) used k-means clustering to analyze crime data in the city of Delhi, India, from 2012 to 2016. The study found that certain areas in Delhi had higher crime rates than others and that crimes tended to occur in clusters. The study also found that certain types of crimes, such as theft and robbery, were more common in certain areas than others. The study concluded that k-means clustering can be an effective tool for crime analysis in India.

Overall, the literature suggests that k-means clustering can be an effective tool for identifying crime hotspots and patterns in various Indian cities. However, more research is needed to address the challenges of data quality and availability and to improve the effectiveness of k-means clustering for crime analysis in India.

Despite its limitations, crime mapping remains an important tool for analyzing crime patterns and devising targeted interventions to prevent and address crime. The literature suggests that crime mapping can be particularly effective when used in conjunction with other crime prevention strategies, such as community policing and problem-oriented policing.

III. MOTIVATION

Crime is a serious issue in many urban areas, and law enforcement agencies need effective tools to identify and address patterns of criminal activity. Crime mapping is a technique used by law enforcement agencies to visualize and analyze crime data, and identify areas of high criminal activity. The use of machine learning algorithms, such as k-means clustering, can help to identify hotspots of criminal activity and provide insights into patterns and trends in criminal behavior.

While crime mapping has been widely used in many countries, there is limited research on its application in India, particularly with regards to the use of k-means clustering. This study aims to fill this research gap by using k-means clustering to analyze crime data in a specific city in India. By identifying hotspots of criminal activity and examining patterns and trends in crime data, this study will contribute to a better understanding of the nature and extent of crime in the city, and provide useful insights for law enforcement agencies to allocate resources effectively and combat criminal activity.

Overall, this research will be important because it will help to develop effective strategies to manage crime in high-risk areas in India, and contribute to the development of more efficient and effective crime prevention policies.

IV. METHODOLOGY

Crime mapping involves several steps that include collecting and cleaning data, machine learning, visualization, and crime prevention. The first step is collecting data on crime incidents, typically from law enforcement agencies or other sources, such as information on the type of crime, the location, and the time of day. Once the data is collected, it needs to be cleaned to remove any errors or inconsistencies, such as duplicates or misspellings, and standardize the format of the data.

a. **Collecting Data**: The first step in crime mapping is collecting data on crime incidents, typically from law enforcement agencies or other sources. The data collected may include information on the type of crime, the location, and the time of day.
b. **Cleaning Data:** Once the data is collected, it needs to be cleaned to remove any errors or inconsistencies. This may involve removing duplicates, correcting misspellings or errors, and standardizing the format of the data.

1. **Machine Learning:** Machine learning is a powerful tool used in crime mapping to analyze large datasets and identify patterns and trends that may not be immediately apparent. One commonly used machine learning algorithm is K-means clustering, which can identify hotspots of crime by grouping crime incidents into a predetermined number of clusters based on their geographic proximity. For example, K-means clustering has been employed to analyze crime patterns in India and identify hotspots and patterns of crime.

2. **K-means Clustering:** K-means clustering is a machine learning algorithm used in crime mapping to identify hotspots of crime. It works by partitioning crime incidents into a predetermined number of clusters based on their geographic proximity. This algorithm helps to identify patterns and trends in the data that may not be immediately apparent. For example, it can help identify areas that have a high incidence of a particular type of crime, or areas where crime is increasing or decreasing over time.

e. **Visualization**: Visualization is also an essential aspect of crime mapping, where various tools are used to represent crime data on maps or charts. Heatmaps, for example, can help identify high-density crime areas and aid in the development of targeted interventions.

c. **Crime Prevention:** The ultimate goal of crime mapping is to prevent and address crime. By analyzing crime data and identifying hotspots, law enforcement

DOI: 10.5281/zenodo.7956631
ISBN: 978-93-5906-046-0@2023, Dept. of Computer Applications, Amal Jyothi College of Engineering Kanjirappally, Kottayam

agencies can develop targeted interventions to prevent and reduce crime in their communities. Crime mapping can also provide valuable insights into the social and economic factors that contribute to crime, allowing for more effective crime prevention strategies.

V. BUILD MODEL

The first lines of code import the required Python libraries for this script: pandas for data handling, NumPy for numerical operations, folium for creating maps, folium plugins. Heat Map for visualizing heat maps, and K Means from scikit-learn for performing k-means clustering.

1. Import the packages that are necessary

```
# Import necessary libraries
import pandas as pd
import numpy as np
import folium
from folium.plugins import HeatMap
from sklearn.cluster import KMeans
```

2.Then clean the dataset

```
# Load crime data
crime_data = pd.read_csv('/content/crime.csv')
```

```
# Drop any rows with missing data
crime_data.dropna(inplace=True)
```

This line of code loads a CSV file containing the crime data into a pandas Data Frame called **crime data**. This line of code removes any rows from the Data Frame that contain missing data.

3. Create feature Matrix

```
# Create feature matrix
X = np.array(crime_data[['Latitude', 'Longitude']])
```

This line of code creates a feature matrix called X from the Latitude and Longitude columns of the crime_data DataFrame.
This line of code sets the number of clusters for the k-means clustering algorithm to 5.

4. Run K-Means clustering

```
kmeans = KMeans(n_clusters=k)
kmeans.fit(X)
```

These lines of code perform the k-means clustering algorithm on the feature matrix X with the specified number of clusters (k).
5.Cluster assignments for each data point

```
# Get cluster assignments for each data point
labels = kmeans.predict(X)
```

This line of code retrieves the cluster assignments for each data point in the feature matrix X.

```
# Add cluster labels to crime data
crime_data['cluster'] = labels
```

This line of code adds a new column called cluster to the crime_data DataFrame, containing the cluster labels assigned by the k-means algorithm.

6. Create a map centered on the mean latitude and longitude of the crime data.

```
# create a map centered on the mean latitude and longitude of the crime data
m = folium.Map(location=[crime_data['Latitude'].mean(), crime_data['Longitude'].mean()], zoom_start=12)
```

This line of code creates a map using the folium library, centered on the mean Latitude and Longitude values of the crime_data DataFrame, and sets the zoom level to 12.

7. Add heat map layer to the map

```
# Add heat map layer to the map
heat_data = [[row['Latitude'], row['Longitude']] for index, row in crime_dat
HeatMap(heat_data).add_to(m)
```

These lines of code create a heat map layer on the map by converting the Latitude and Longitude values from each row of the crime_data DataFrame into a list of coordinates, and then using the folium.plugins.HeatMap library to create the heat map layer.

8. Add cluster markers to the map

```
colors = ['red', 'blue', 'green', 'purple', 'orange']
for i in range(k):
    cluster_data = crime_data[crime_data['cluster'] == i]
    cluster_coords = list(zip(cluster_data['Latitude'], cluster_data['Longit
    for coord in cluster_coords:
        folium.Marker(location=[coord[0], coord[1]], icon=folium.Icon(color=
```

These lines of code add cluster markers to the map by iterating through each of the k clusters, retrieving the Latitude and Longitude values for each data point in the cluster, and creating a marker for each data point with a color corresponding to the cluster label.

DOI: 10.5281/zenodo.7956631
ISBN: 978-93-5906-046-0@2023, Dept. of Computer Applications, Amal Jyothi College of Engineering Kanjirappally, Kottayam

VI. CONCLUSION

Crime mapping and analysis is a valuable tool for law enforcement agencies and researchers to identify crime patterns, hotspots, and trends. By utilizing data analytics and machine learning algorithms, targeted interventions can be developed to prevent and address crime. However, it is essential to acknowledge the limitations of crime mapping and analysis, including the potential for bias and the importance of privacy considerations. To build an effective cybercrime prediction model, it is crucial to follow a rigorous methodology, including data collection, cleaning, feature selection, model selection, training, and evaluation. By employing appropriate algorithms and techniques, accurate and reliable models can be created to predict and prevent cybercrime. Overall, the use of crime mapping and analysis can help to enhance community safety and security if used ethically and effectively.

VII. REFERENCES

[1] Yadav, S., Bhaskar, P., & Yadav, V. (2016). Spatial and temporal analysis of crime in Delhi using GIS and spatial statistics. Geocortex International, 31(9), 975-992

[2] Khare, A., Raut, S. S., & Badri, A. (2017). Spatial analysis of criminal activity in the city of Mumbai using GIS. In Proceedings of the International Conference on Advances in Computing, Communications and Informatics (ICACCI) (pp. 2472-2478).

[3] Singh, P., Mohapatra, P. K. J., & Agrawal, A. (2019). Crime data analysis using machine learning techniques in India: A review. In Proceedings of the 3rd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT) (pp. 827-833).

[4] Pandit, A., & Mukherjee, S. (2019). Spatial clustering of crime using k-means clustering: A study of Kolkata. Crime Mapping: A Journal of Research and Practice, 11(2), 151-164

[5] Quest Journals Journal of Research in Environmental and Earth Sciences Volume 7 ~ Issue 11 (2021) pp: 34-45 ISSN(Online) :2348-2532

[6]www.CrimeMapping.com/map/MI/GrandRapids

DOI: 10.5281/zenodo.7956631
ISBN: 978-93-5906-046-0@2023, Dept. of Computer Applications, Amal Jyothi College of Engineering Kanjirappally, Kottayam