Crime Spot Detection: Identifying Safety Percentage of Location using FCM

[1]P. Rambabu, [2]Dr. M V S S Nagendranath, [3]M. Anantha Lakshmi,
Sasi Institute of Technology & Engineering, Andhra Pradesh, India
[1]rambabupasumarthy@sasi.ac.in, [2]nagendranath@sasi.ac.in, [3]anantha.cse@sasi.ac.in

Abstract

Crime is one of the central and most serious societal issues, and preventing it is a critical duty. Law enforcement agencies are also facing major issues in preventing crimes. In order to solve this and make our society a crime-free society, there is a need to develop a system that detects crime. Nowadays, one of the methodical techniques to reduce illegal activities all over the world is to analyse and identify crime rates. Crime does not happen in all the areas, it follows some patterns. There are some factors that cause crime to happen in that location, for example, locations that are far from the city usually have a higher crime rate. By preventing traveling to those areas crimes can be reduced. To know which place is safe to travel to, there is a need to find the safety percentage of the location. This can be achieved by using machine learning techniques. AI strategies were utilized to plan a model for recognizing the security level of the area. In order to implement this problem, clustering approaches were used i.e, Fuzzy c-means clustering and k-means clustering. Tests were led with various clustering approaches on the wrongdoing dataset and found that FCM gives improved results. Exploratory outcomes show that fluffy c-implies grouping gives 38% more precise outcomes than k-implies bunching. So, the idea of using a machine learning method called Fuzzy c-means clustering for identifying the safety percentage of the location is very useful and also gives more accuracy.

Keywords: *Crime detection, Clustering, Fuzzy-c-means clustering, Safety Percentage.*

1. Introduction

Criminality is a problem that affects both developing and developed countries. Crimes and criminal acts can have an impact on both the public and private sectors. When people migrate or relocate, public safety is an important factor in creating secure environments. Actually, various kinds of consequences might have various results. consequences are carried out for an assortment of reasons, like extraordinary inspirations, human instinct and conduct, basic circumstances, and neediness. Joblessness, orientation disparity, thick populace, kid work, and ignorance are factors that add to destitution.

Moreover, an expansion in crimes can be credited to an assortment of elements, including joblessness, orientation disparity, high populace thickness, kid work, and lack of education. In developing and swarmed urban communities, higher crime percentages are related with an assortment of conditions, including business structures and metropolitan lodging regions. A few crime analysts and researchers have as of late endeavoured to control wrongdoing through exploration and estimating utilizing different demonstrating and measurable methodologies. Since crime percentages keep on increasing, there might be a requirement for a few critical examinations to help policymakers and concerned divisions in tending to difficulties and issues in the space of wrongdoing expectation and control techniques. A human's range of abilities neglects to monitor criminal records when dealt with manually. As an outcome, an original technique for dissecting crimes related information is required.

It is unsurprising that citizens in a developing country like India hear about crimes on a frequent basis. We must be constantly aware of our surroundings as cities progressively urbanise. To avoid the worst-case situation, we'll track crime rates using the fuzzy c-means method. It will make an educated prediction as to the type of crime that will take place, as well as when, where, and how it will take place. This data will illustrate crime tendencies in a specific area, which could be useful in criminal investigations. It will also provide us with information on the most serious crimes that have occurred in a certain location. In this paper, the fuzzy c-means machine learning technique will be applied.

Because there is so much data on crime, crime detection and identifying criminals is one of the police department's top concerns. Technology that allows for quicker case resolution is in high demand. The premise behind this study is that if we can sift through a vast amount of data and uncover patterns that can be utilized to configure what is needed, crimes can be anticipated easily. Because of recent breakthroughs in machine learning, this task is now feasible. Date, time, and location (longitude, latitude) will be entered as inputs, with an output that will inform us which crimes are most likely to occur in that area.

In the field of traditional crime evaluation, an effective computer for predicting crimes is required because traditional tactics cannot be employed when crime data is highly dimensional and complex queries must be processed. As a result, a device for detecting and evaluating criminal offences was desired for accurately determining crime patterns. This paper offers a few approaches that can be used to predict which types of crimes are more likely to occur in which areas and at what times.

Data processing techniques are employed to make this endeavour easier. Law enforcement agencies have been able to integrate enormous databases with specific information on serious crimes such as murder, rape, and arson thanks to improved computer accessibility and data

breakthroughs. In recent years, a large number of crimes have been reported all across the world. When a criminal threatens to use force against a victim, this is considered a primary offence. It encompasses crimes where a violent act is the motive, such as murder or rape, as well as crimes where violence is used as a kind of compulsion, such as robbery. Depending on the jurisdiction, offences can be committed with or without weapons, and violence crimes range from murder to harassment.

It is difficult to detect illegal activity accurately, but it is necessary for preventing criminal behaviour. The precise assessment of the crime rate, types, and hot places from historical patterns presents several computational challenges and opportunities. Despite the fact that machine learning-based crime detection is now the industry norm, only a few studies have carefully examined different machine learning algorithms.

Crime detection is significant because it aids in the prediction of crime rates and the assessment of future crime rates. Officials can use this information to take action and try to reduce crime rates. You'll need a lot of information to figure out crime trends.

The main objective of this work is to reduce the crimes by calculating the safety percentage of the location using the crime data in that area. This can be achieved by using machine learning techniques. So, in this paper, we proposed to use a clustering model called Fuzzy-c-means clustering. to identify the crimes in a particular location and calculate the safety percentage of the location.

2. Related Work

The CrimeTracer model is introduced by Martin Ester et al, (2014) [1] In this model some processes are implemented. The stochastic process in which the starting state is known and the future state is computed using a transition probability matrix that calculates the likelihood of migrating from one graph node to another. The random walk method converges to a stationary distribution that assigns an important value to each node in the graph under particular conditions.

Matthew S. Gerber (2014) [2] used a Decision Support System for crime rate detection. The DSS is an information system that assists and participates in human decision-making. DSS can help decision-makers use data, model, and solve unstructured situations in general. As a result, they finalized that there are certain DSS that can be utilized to assist security authorities in making sense of crime data and resolving the challenge of crime prevention decision-making.

J. Fitterer et al. used BNEs [3] for patterning and detecting crime. Many police departments are employing geographic information systems (GIS) as a crucial tool in intelligence-led policing, and spatial forecasts of crime are being used to minimize crime.

The most frequent technique for detecting crime is to utilize statistical detection algorithms to obtain the frequency and parameters of crime. There has been a lot of research done in this area. ARIMA and STL [4] are two of the most common regression models.The ARIMA model outperforms the other two techniques in terms of fitting and detecting accuracy.

Maria R et al, (2014) [5] prepared a review document on the statistical physics of crime, Controlling the spread of crime in urban areas remains a significant concern. Crimes may be repeated and proliferate if left uncontrolled, according to empirical research.

Clustering techniques were employed by S.Sivaranjani et al., (2016) [6] to detect crimes. Crime is a form of human rights violation that is frequently prosecuted and punished by the law. Criminology is the study of crime, and it is an interdisciplinary science that collects and examines data on crime and criminal behavior.

In 2011, a spatial-transient model of crime recognition in light of criminal individual qualities was created. A few articles, for example, ARIMA-LSSVM model for crime time series conjecture, notice blended models. To overcome the problem, paper [7] provides a strategy based on the TPML-WMA algorithm. It differs from other approaches in the following ways: First, the data set's geographical locations are divided into administrative districts. Second, Xinlei Wei et al, [7] (2016) built the Vector Motion model and utilized change likelihood networks to show the impact across regions, as well as planned a calculation to become familiar with the frameworks, because of the uniqueness of the information collection and examination challenges.

Qiang Zhang [9] et al.,(2016) performed hotspot detection based on Mixed Spatial-Temporal Characteristics of crime. The model's core idea is to convert the problem of Hot Spots detection into a multi-class classification problem by encoding area-specific criminal episodes into multiple degrees of heat levels. Different sorts of security-affecting features were merged into mixed spatial-temporal characteristics, which were employed as input variables for detection in the new model.

Sunil Yadav [11] et al.,(2017) used Isotropic triggering for crime rate detection. The aim is to describe the criteria that will be used to segment the entire database; once this is done, each dataset will naturally fall into one or more categories. Existing datasets may be easily understood with the help of classification, and it also aids in detecting how fresh individual datasets will behave based on the categorization criteria.

Cory Schnell et al., [13] used three degrees of geographic aggregation community regions, neighborhood clusters, and street segments. These geographical aggregation levels were chosen for their usefulness in understanding Chicago's geography as well as their previous application in social science research. They used different machine learning algorithms to implement their idea.

Nelson Baloian [15] et al., (2017) proposed methods for treating data based on data mining, also known as detective analytics, that can discover new criminal patterns. Near repeat approaches have been applied to the country of Merseyside in North West EnglandFinally, technologies based on geospatial techniques can be used to add new layers of data to crime maps.

George Mohler et al.,[16] provided an approach for predicting crime. For this Hawkes process was used. Using this they also created a model which enables the communication between any two organizations for transferring information like tips for the safety of the public, crime information, etc, This method is also used in other social media applications like Twitter.

T. Sarvani et al.,[18] used a feature selection process to detect crime patterns. Further applied Naive Bayes classifier for identifying the type of crime that is frequently happening. Their paper discusses both predicting crimes and the type of crime happening.The main question addressed in [18] is whether it is possible to reliably estimate chosen crimes in local locations, such as police precincts, a month in advance.

B.Sivanagleela et al,[19] used fuzzy clustering for analysing and predicting crimes. Using the crime data available, preprocessing and clustering were performed, as they result in detecting crime rates, leading to a decrease in crimes. Clustering raw data generates a membership function from the data and builds the fuzzy inference system fuzzy c, which deals with the data's uncertainty. Crimes can be predicted from the above results, This is clearly shown in their paper.

Chao Huang et al., [21] presented the recurrent neural networks model for anticipating crime hotspots that use spatial and temporal information and compared machine learning algorithms. This work is conducted by dividing the area into a number of square parts. Every square part consists of crime data that happened in that location in the past. This approach is similar to the divide and conquers technique.

L.Lochner, [22] investigated several data mining techniques used by law enforcement authorities to detect and prevent terrorism. The authors also looked into the limitations of data mining in fighting crime in San Francisco and came to the conclusion that data mining can only be utilized to help law enforcement authorities analyse crime.

In [23], N.V.Keerthana et al. used different machine learning algorithms to predict crimes. This paper not only predicts crimes but also predicts the type of crimes. One of the machine learning algorithms that achieve the above-mentioned goal is Naive Bayes. Of all the machine learning algorithms used in the previous work, Naive Bayes give more accuracy, remaining algorithms give accuracy but are not used for predicting the type of crime.

3. Proposed Work

Subsequent to concentrating on the different methodologies and strategies toward crime percentage discovery, a ground breaking thought that will assist us with picking the best area for safe living makes sense beneath. We use fuzzy c-means clustering to propose a model that gives a notion of whether a site is secure or not. With fuzzy, you can use any form of data, categorized or numerical. Fig 1 shows the architecture of crime rate detection for safety purposes. Safe living can be achieved through this.

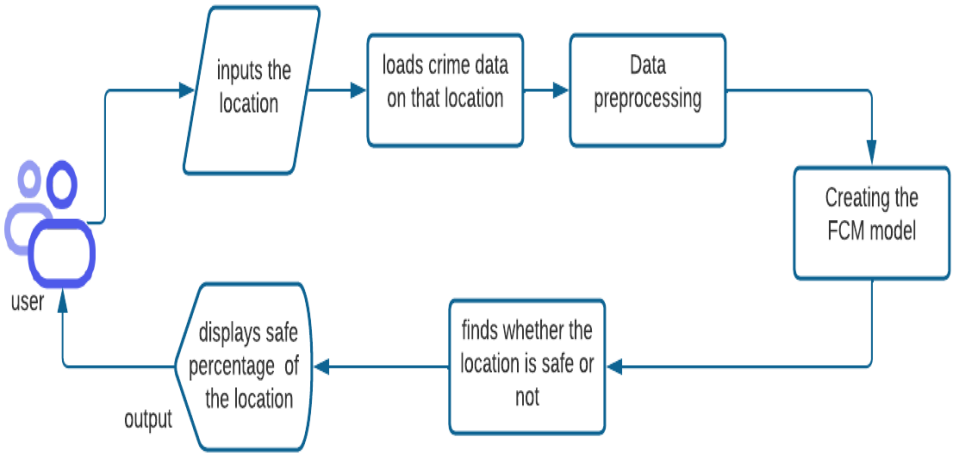

Fig.1. Block diagram of the proposed method

We can see from the architecture diagram above that when the user provides the location as input, our model assists us in determining whether or not the area is secure. It takes the location as input and loads the crime data of the location from the database. For better results, the data is further preprocessed. From this, the safe percentage of location is identified and displayed to the user. This project helps the users to find whether the location is safe or not.

From the above figure, we can see that the system takes the name of the location as input and loads the previous crime data on that location. Since the data is collected from various resources there may be inconsistencies in the data. So in the next step, we will use some preprocessing techniques to clean the data.

For this purpose, we will use feature scaling and remove a tuple method. By these two methods data preprocessing steps will be completed. The purpose of feature scaling is to balance all unbalanced rows and columns. In AI, highlight scaling is the last stage in information preprocessing. It is a procedure for normalizing the data set's free factors in a given reach. We set our factors in a similar reach and scale in highlight scaling with the goal that nobody variable offsets the other. Before performing data preprocessing exploratory data analysis should be done on the imported data set. It is the procedure that summarizes the main characteristics by showing them graphically. It gives a better understanding of data.

4. Methodology

We use fuzzy c-means clustering to propose a model that gives a notion of whether a site is secure or not. Unsupervised learning is used in the fuzzy C-means clustering algorithm. The word "fuzzy" here signifies "not sure," indicating that this is a soft clustering strategy. "C-means" stands for "c cluster centres" with the "K" in "K-means" replaced with a "C" to give it a new look. Fuzzy C-Means is created in view of the k-means clustering approach. The Fuzzy C-Means is a soft clustering technique.

Fuzzy C-Means clustering is a soft clustering method in which every information point is designated a likelihood or probability score to have a place with that group. Each data point and each cluster are assigned to a fuzzy cluster depending on the membership degree; in fuzzy C, each object can have a membership value ranging from 0 to 1. With fuzzy, you can use any form of data, categorized or numerical. The degree of each cluster is determined by connecting the values of the cluster data to each data point in this clustering. All of the dataset’s membership values should be one.

Clustering raw data generates a membership function from the data and builds the fuzzy inference system fuzzy c, which deals with the data's uncertainty. Fuzzy c-means is a data clustering approach in which a dataset is partitioned into N clusters, with every information directly having a place in each cluster somewhat.

When compared to other algorithms, fuzzy c-means clustering is a better option.The information point can have a place with more than one group with a probability in the fuzzy c-means method. For overlapped data sets, fuzzy c-means clustering produces superior results.

Clustering, also known as cluster analysis, is the most common way of organizing data of interest into groups so that items in a single cluster are however comparable as conceivable while others may be essentially as different as could really be expected. Fuzzy clustering is a grouping strategy where every information point can be relegated to many clusters. Clustering

problems are used in a range of domains, including surface science, biology, health, psychology, economics, and a variety of others. After constructing 11 clusters for 11 different crime types, the frequency should be calculated. A new data frame will be created to determine the crime rate based on the neighbourhood or area. Insert all of the information into the Python dictionary after you've calculated the crime rates for each location. When a user enters a location, the output is the crime rate for that location. Dictionaries in Python are a form of hash table. A dictionary key can be any Python type, but numbers and texts are most frequent. Values, on the other hand, can be any Python object you choose.

A. Experimental Setup

To perform clustering techniques like Fuzzy C-Means clustering which is used for checking the performance in detecting the crimes, we used the following software tools. The model's performance was evaluated using Jupyter Notebook and the Anaconda IDE, with Python 3.8 as the scripting language. Visual Studio Code and Jupyter Notebook are the frontend technologies used to implement the project. We also used python libraries for implementing this.

B. Model Fitting

Our model, Fuzzy C Means, must first be installed. Import FCM from means. once we've installed fuzzy c means at work. Clustering is done with a Fuzzy c Means model. So, once the software is installed, we must assign a number to the data in order to convert it to clusters. After determining the number of clusters, we must fit our model and then locate centres and labels for clustering our data.

The clustering is formed by using labels and centres. The labels are determined by predicting the model whereas centres are determined directly. The clustering shows the count of the offenses which helps to know if the location is safe or not. If the offense rate in a location is low, then that location's safety is high. The clusters also show the offense rate in different colors. If we are taking two clusters then we found two different colors one is for high safety and another is for low safety. if there are two or more clusters then the clusters are divided by the percentage of offense rate to show in different colors. The above-mentioned percentage of safety is shown clearly in simple python code.

C. Data Set

We require past crime data of the areas to determine whether they are safe or not. The data set can be obtained by visiting the GitHub source. This data was used by the author Vashista to create crime data pattern analysis and visualization. This data set will be used to carry out our project. We used this information to carry out our project. Offenses that occurred in 2017 are included in this data collection. There are four criteria in total\, with 33,116 records. The parameters specify the offense reporting date and time, the location where the offense occurred, and the location's offense count. The data set includes the locations of Washington, United States of America. The coordinates are given in longitude and latitude format. Positive latitude values and negative longitude values are included in the data set.

The data set is to be preprocessed because it contains words instead of numbers in the offense column. Now we are replacing that word (ARSON) with the average value we get from the

remaining numbers. Now we are checking if all the values are in integers or not. Now all values are correct in their positions. Finally preprocessing is completed. With the preprocessed data we used we used the following software tools. The model's performance was evaluated using Jupyter Notebook and the Anaconda IDE, with Python 3.8 as the scripting language. Visual Studio Code and Jupyter Notebook are the frontend technologies used to implement the project. We also used python libraries for implementing this.

After preprocessing the data, a graph is plotted taking latitude on x-axis and longitude on y-axis. Based on the data set we have downloaded, the graph is plotted. Now in order to plot the clusters in the graph we have created the FCM model. Now using this FCM model the clusters are plotted on the graph with latitude on x-axis and longitude on y-axis which represents the crimes on the respective locations.

5. RESULTS AND DISCUSSIONS

After the successful implementation of our project, the results are as follows.

Fig.2. The result after implementation with FCM

Here latitude and longitude of the location is given as input and the output will be the safety percentage of that location. In order to implement this, we have collected a dataset that contains the date and time of the crime, latitude, longitude, and offense that happened at that place. To implement this we used a fuzzy c-means clustering algorithm. On the data set we have collected, we plotted a graph before applying the fuzzy clustering algorithm and the graph is as follows.

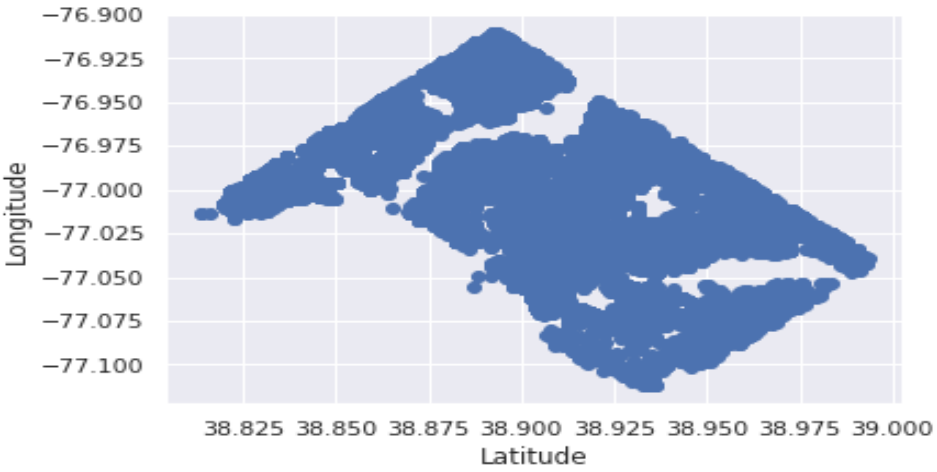

Fig.3. Plotting latitude and longitude before clustering

Fig.3 represents the locations in which we have collected the data set. The crimes that happened in the above locations are collected and used as our dataset, for implementing our project.

After successfully applying the FCM algorithm to our dataset the results are as follows, For implementing this firstly we performed data preprocessing on our dataset and then divided the data into clusters using the FCM clustering algorithm. Based on the clusters formed we are able to divide the areas into high crime, low crime, and medium crime areas. This is possible by taking the number of clusters as four. These four clusters represent four kinds of areas i.e., high crime, low crime, and medium crime. The above-represented graph is again plotted after clustering. That graph after clustering has four colors in it, representing four kinds of areas. Based on that we are able to classify the data into safe areas and non-safe areas. Using this we also calculated the safety percentage of the locations that we have taken in our data sets. Plotting of locations after completion of clustering is shown in the below figure.

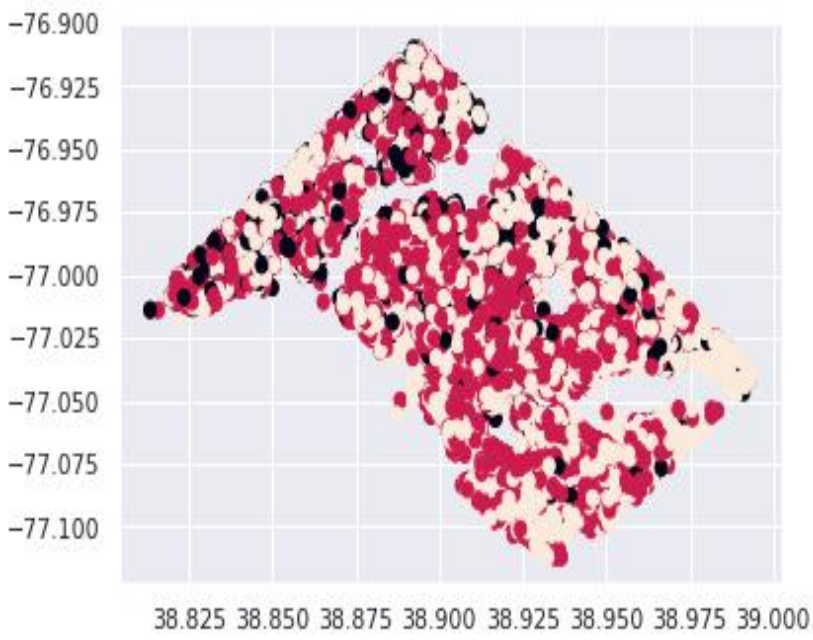

Fig.4. Plotting latitude and longitude after clustering

As explained above in order to implement our project we used the fuzzy c means clustering method. But fuzzy c means clustering is an unsupervised algorithm and also a clustering technique. There is no traditional approach for calculating the accuracy of a clustering algorithm. So in order to show the performance of fuzzy clustering we used a k-means clustering algorithm. In order to benchmark the fuzzy c-means clustering model, we have constructed a k-means clustering model. The results generated by k-means will be used to benchmark the fuzzy c-means model.

For implementing this we have taken the number of clusters to four for both fuzzy c-means and k-means. When compared with k-means clustering, fuzzy c-means clustering gives 38% more accurate and faster results. So if the accuracy of the k-means clustering is 58%, then the accuracy of fuzzy c-means will be 96%.

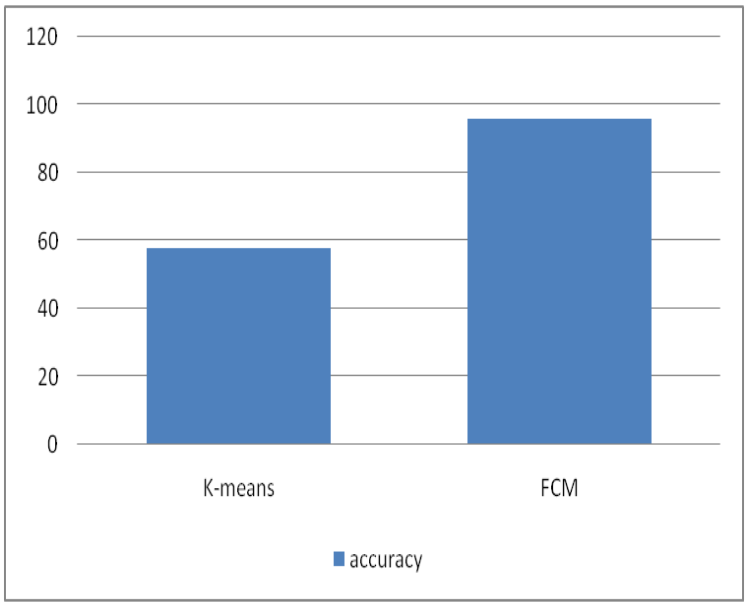

Fig.5. Comparisons graph

It is used in various applications. It is mainly used for the identification of crime-prone areas. It is used for security purposes. This project can be utilized by the police department and crime department It can also be used by the road-safety department. If it is developed and used as an application it will be helpful for the people who are new to a certain place. If a person needs to travel to a new place, he/she will choose the shortest path, but the person may not know whether the place is safe or not. Our project helps in finding the safety percentage of the location.

6. Conclusion and Future Enhancement

This project proposed a method for preventing criminal activity in society. This project assists in determining the safest travel area. When a person is new to a city, he or she may not be aware of whether or not a particular path is safe to take. If the individual wishes to travel from one location to another, he will take the quickest route accessible, but he may not know whether or not that route is safe to travel. As a result, our project is beneficial in this case. This project is set up in such a way that if we offer the location as an input, we will get the safety percentage of the place as an output. We employed one of the clustering approaches, the Fuzzy-

c-means clustering algorithm, to complete this assignment. We used this technique to partition the data into clusters and determine the location's safety percentage. When it comes to clustering algorithm accuracy, there is no standard method for calculating it. To benchmark the FCM, we employed k-means clustering. The FCM method will be compared and accuracy estimated based on the findings given by k-means. FCM is shown to be 38 percent more accurate than K-means after implementing this.

For the future enhancement of this idea, at present this project is implemented only in a particular city. In future, it can be extended such that it is implemented to a whole country. This project is implemented only by using python code, So it can only be used by people who have this code. In future this can be developed and can be converted as an app or web application so that this can be used by the public.

References

[1]. Martin Ester, C Graif, and Z. Li,, ‘‘Crime rate inference with big data,’’ in Proc. 22nd ACM SIGKDD Int. Discovery Data Mining, Aug. 2014, pp. 635–644

[2]. Matthew S. Gerber, C.Contreras, ‘‘A block-level analysis of medical marijuana dispensaries and crime in the city of Los Angeles,’’ pp. 1069–1095, Sep. 2014.

[3]. J. Fitterer, E. L. Piza A. and A. Braga, ‘‘The influence of community areas, neighborhood clusters, and street segments on the spatial variability of violent crime in Chicago BEN's’’ J. Quant. Criminal., vol. 33, pp. 469–496, Sep. 2014.

[4]. J. M. MacDonald and G. Ridgeway, ‘‘Effect of rail transit on crime: A study of Los Angeles from 1988 to 2014,’’ pp. 277–291, Jun. 2017.

[5]. Maria R, Nelson Baloian,” Crime Prediction using Patterns and Context”, IEEE 21st International Conference pp.1-9,2014.

[6]. S.Sivaranjani, Charlie Catlett, Eugenio Cesario,” A Data-driven Approach for Crime Predictions in Smart Cities using clustering.”, IEEE,2016.

[7]. Xinlei Wei, K. R. Majeed, and S. N. Christian, ‘‘Application of data analytics techniques in analyzing crimes using TPML-WMA algorithm,’’ pp. 1–7, 2016.

[8]. Agarwal Subhan "Crime Prediction Based on Statistical Model”, International conference of contemporary computing,2016.

[9]. Qiang Zhang, and D. W. Aye, ‘‘Research studies on human cognitive ability,’’ pp. 298–304, 2016.

[10]. R. A. Hirsch, and S. Abbott, ‘‘System backgrounds, psychosocial characteristics, and service access among dually involved youth: A Los Angeles case study,’’ Youth Violence Juvenile Justice, pp. 309–329, 2017.

[11]. Sunil Yadav, T. Cheng and G. Rosser ‘‘Improving the accuracy of crime prediction with isotropic triggering,’’ pp. 5–25, Mar. 2017.

[12]. L. Lochner, ‘‘Education and crime,’’ A Comprehensive Overview, S. Bradley and G. Green, pp. 109–117.

[13]. Cory Schnell, “An Empirical Analysis of Machine Learning Algorithms for Crime Prediction”, IEEE Access,2017.

[14]. T. Zhang, Z. Yuan‘‘Spatio-temporal pattern analysis and prediction for urban crime,’’ Aug. 2017, pp. 177–182

[15]. Nelson Baloian, ‘‘Gang violence predictability Using risk terrain modeling pp. 10–21, Sep. 2017
[16]. P.J. Brantingham and Mohler, ‘‘Privacy preserving, crowdsourced crime Hawkes processes,’’ pp. 14–19, Apr. 2018.
[17]. P. S. G. de Mattos Neto, J. F. L. de Oliveira, and D. S. de O. Santos Júnior ‘‘An intelligent hybridization of ARIMA with machine learning models for time series forecasting,”pp. 72–86, Jul. 2019.
[18]. M. BalaAnand, T. Sravani, N. Ramakrishiah, and T. Sravani ‘‘ Crime prediction and its type using Naïve Bayes classifier,’’ pp. 187–197, Sep. 2019.
[19]. B.Sivanaglela, S.Rajesh “Crime Analysis and Prediction using FCM”pp.595-599, Apr 2019.
[20]. WAJIHA SAFAT,” Empirical Analysis for Crime Prediction and Forecasting Using Machine Learning and Deep Learning Techniques” IEEE Access,2021.
[21]. Chao Huang, L. Liu, X. Zhang, and J. Ji, ‘‘Comparison of machine learning algorithms for predicting crime hotspots,’’ IEEE Access, pp. 181302–181310, 2020.
[22]. L.Lochner, M. Hüser, X. Lyu, T. Gumbsch, C. Esteban, C. Bock, M. Horn, M. Moor, B. Rieck, M. Zimmermann, D. Bodenham, K. Borgwardt, G. Rätsch, and T. M. Merz, ‘‘Early prediction of circulatory failure in the intensive care unit using machine learning,’’ Nature Med., vol. 26, no. 3, pp. 364–373, Mar. 2020.
[23]. N. Kanimozhi, N.V. Keerthana, G S Pavitra,” Crime type and occurrence prediction using machine learning algorithms”,pp.1-4,2021