DETECTION OF CRIME AGAINST WOMEN USING MACHINE LEARNING ALGORITHMS

1 K. Siva Rama krishna, M.tech, Assistant professor & Hod department of CSE, lingayas institute of management and technology, Andhrapradesh, India

2 P. Tarun, department of CSE, lingayas institute of management and technology, Andhrapradesh, India

3 G. Satish Kumar, department of CSE, lingayas institute of management and technology, Andhrapradesh, India

4 M. ANANDAMAI, department of CSE, lingayas institute of management and technology, Andhrapradesh, India

ABSTRACT: Preventing crime is essential because it is a troubling aspect of our culture. A methodical approach to identifying and analyzing criminal behavior patterns and trends is crime analysis. In order to determine the most effective strategies for the management of criminal activity, it is essential to look into the underlying causes, the various variables, and the connections that exist between the various crimes that take place. Using grouping techniques, this initiative aims to distinguish between various offenses based on instances and consistency. In criminal investigation, inquiry, and trend detection, data mining is utilized. A clustering procedure is used in this undertaking to look at the criminal information; The K-Means calculation is utilized to bunch the saved information. After categorizing and grouping the data, we can predict a crime. This suggested approach can differentiate between areas with a higher rate of crime and locations with a high likelihood of crime.

Keywords – Machine learning, K-means algorithm, Clustering techniques.

1. INTRODUCTION

The National Commission for Women (NCW) reports that in 2020, India received 23,722 reports of crimes committed against women, the most in the previous six years. Women are still regarded as slaves in some parts of India, and it poses a significant

threat to civilization. Women's sexism goes back as far as human history. Women in India face a variety of challenges, including marriage, prostitution, acid attacks, abortion, hijacking, and abduction, not only in rural areas but also in urban areas. An inadequate legal justice system, lax legal regulations, crime forecast, and criminal identity are the primary causes of crime against women. Government officials must assess and forecast the rate of crime and take the necessary precautions to stop an increase in violence against women.

Fig.1: Example figure

As per the National Commission for Women (NCW), 23,722 reports of offenses against ladies were submitted in India in 2020, the most in the past six years. In some parts of India, women are still considered to be slaves, and it poses a significant threat to civilization. Women's sexism goes back as far as human history. In India, inheritance,

prostitution, acid attacks, abortion, hijacking, and theft are just a few of the issues that women face. Wrongdoing against ladies is basically brought about by an incapable legitimate equity framework, remiss standards of regulation, wrongdoing figure, and unlawful character. Government officials must assess and forecast the rate of crime and take the necessary precautions to stop an increase in violence against women.

2. LITERATURE REVIEW

Survey on the Crime analysis and crime prediction using several Data Mining techniques:

Data mining is the most common way of breaking down and surveying large previous datasets to create new data that might be helpful to the association. New data is expected to be extracted from existing databases. There have been numerous attempts at data mining analysis and forecasting. In the field of forensics, however, only a few attempts have been made. Few have made endeavors to look at the material delivered by these strategies. Police divisions and other comparative law enforcement associations hold various enormous datasets of data that can be

utilized to anticipate or examine unlawful developments and the cooperation of crime in the public eye. Measurements on crime can likewise be utilized to anticipate criminals. The primary objective of this research is to examine both controlled and autonomous methods of crime detection. This report gives an outline of criminal examination and gauge utilizing different Data Mining strategies.

Crime analysis and prediction using data mining:

Crime investigation and counteraction is a deliberate way to deal with distinguishing and dissecting crime examples and patterns. By predicting locations with a high likelihood of crime occurring, our algorithm is able to anticipate and visualize crime-prone areas. As computerized systems become more common, crime data specialists can help law enforcement agents solve crimes more quickly. We can pull already unseen, significant data from disorderly information utilizing the possibility of information mining. We have an arrangement that makes an data mining system that can assist with tackling crimes all the more rapidly by consolidating

software engineering and law enforcement. We are concentrating on the daily factors that contribute to crime rather than the causes of crime, such as the criminal history of the perpetrator or political animosity.

Predicting Crime Using Time and Location Data:

In order to respond more effectively to illegal behavior, it is essential to understand crime trends. This example is reviewed using criminal records from the Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) system. Chicago's different areas are remembered for this assortment. Foreseeing which sort of crime is probably going to happen at a specific overall setting in Chicago is the essential target of this task. Finally, each algorithm's precision is evaluated using a distinct algorithm like Random Forest or Decision Tree as well as a variety of ensemble techniques like Extra Trees, Bagging, and AdaBoost.

3. METHODOLOGY

Numerous research articles on network restructuring have been published in recent years. In this context, optimizing the

geographic coverage of sensor nodes, job distribution, resource utilization, and visual data transfer are all relevant. These developments aim to develop energy-efficient strategies for extending the lifespan of VSNs. In this article, we look at some notable optimisation techniques from the existing literature for resource-constrained situations.

Drawbacks

● It takes longer to transfer content into the network.

● The level of accuracy is low.

Proposed system

Database-related machine learning techniques are readily available to users. However, there are two categories of learning algorithms: unstructured as well as controlled learning algorithms. We will use K-means clustering to create a few groups and names through autonomous learning. From labeled training data, supervised learning algorithms derive knowledge or "the correct answer." The algorithms are given a particular quality or set of qualities to use in forecasting. There are ways to get rid of any empty or endless numbers that

could affect the accuracy of the system that are included in the process of preparing the data.

Advantages

● Accuracy enhanced

● Crime forecast aims to lessen the number of crimes committed by predicting which kinds of crimes will occur in the future.
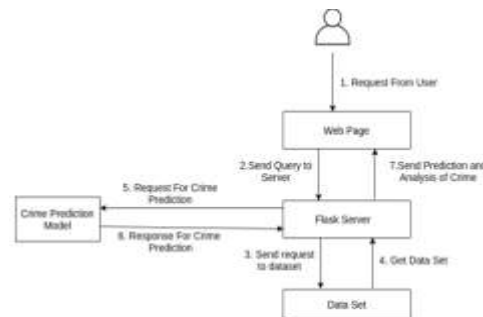

Fig.2: System architecture

MODULES:

1. DATA COLLECTION

2. DATA PRE-PROCESSING

3. FEATURE EXTRATION

4. EVALUATION MODEL

5. FLASK

DATA COLLECTION

The method involved with get-together data from different hotspots to foster ML calculations is known as information assortment. The data should be saved in a manner that is reasonable for the issue. The information assortment is changed into a conceivable construction that can be placed into ML calculations in this stage.

This study uses 15 characteristics of cervical cancer as its data. At this stage, you have to choose from all of the available materials to work with. Problems with ML start with data, and you should have a lot of it—examples or samples—for which you already have a solution. The goal response is already known for data that is labeled.

DATA PRE-PROCESSING

To organize your chosen data, format, clear, and sample it.

There are three typical steps in data pre-processing:

Formatting: It's possible that you won't be able to work with the data you chose. You could have data in a relational database and want to store it in a flat file, or you could

have data in a private file format and want to store it in a text file or a relational database.

Cleaning: The removal or restoration of lost data is known as data cleaning. There may be partial data examples that do not include the information you believe is necessary to resolve the issue. It might be necessary to get rid of these things. Moreover, a portion of the qualities might contain secret data, which should be sifted or erased totally from the information.

Sampling: It's possible that there are far more carefully selected statistics available than you need. Programming times can become significantly longer as a result of more data, as can processing and memory requirements. You can take a smaller indicative sample of the chosen data before looking at the entire dataset, which may make it easier to investigate and test solutions.

FEATURE EXTRATION

The accompanying step is to Feature extraction is a strategy for lessening the quantity of characteristics. Feature extraction alters the attributes, in contrast to feature selection, which ranks current

attributes according to their prognostic significance. Linear mixtures of the initial qualities make up the altered characteristics. Finally, we use the Classifier approach to instruct our models. The categorise function from the Natural Language Toolkit package is used on Python. We make use of the information that has been annotated that was gathered. The algorithms' evaluation will be based on the remaining annotated data. To recognize pre-handled information, some ML strategies were utilized. It was decided to use random forest classifications. When categorizing text, these techniques are frequently employed.

EVALUATION MODEL

A crucial step in the modeling process is model evaluation. It helps decide which model best reflects our data and how well it will perform in the future. Since it is not difficult to create overoptimistic and overfit models, it isn't suitable in that frame of mind to assess model achievement utilizing the information utilized for preparing. There are two techniques for assessing models in information science: pause and cross-endorsement. Both methods use a test collection that is not visible to the model to

assess the success of the model to avoid overfitting.

Each categorization algorithm's performance is measured using its mean performance. The final result will be depicted. Data that can be categorized is represented by graphs.

The expression "accuracy" alludes to the extent of test information figures that are precise. By partitioning the absolute number of assessments by the quantity of exact figures, it is not difficult to compute.

4. IMPLEMENTATION

All learning takes place at the center of machine learning. The way computers learn is similar to how humans learn. People learn from experience. We are better able to anticipate with more information. In contrast, our chances of success are lower when we face a known situation than when we face an uncertain one. Machines are shown in a similar way. In order to make an accurate prediction, the computer looks at a sample. The computer is able to predict the outcome when given a case that is comparable. However, just like a human, the computer has trouble predicting outcomes if fed a previously unknown case. Machine

learning's primary objective is to learn and guess. The computer learns first and foremost by observing patterns. This finding was made conceivable by insights. The careful selection of which data to provide the computer with is one of the crucial responsibilities of the data scientist. An assortment of qualities used to resolve an issue is known as a component vector. A piece of information that is utilized to take care of an issue is known as a component vector. The computer reduces reality and turns this finding into a model using sophisticated algorithms. Consequently, a model is created by characterizing and summarizing the data during the learning step.

K-MEANS:

K-means clustering is a sign handling determined vector quantization technique that partitions n perceptions into k groups, with the nearest mean (cluster centers or cluster centroid) filling in as the layout for each group. Thusly, Voronoi cells are utilized to isolate the information region. K-means clustering limits inside bunch deviations (squared Euclidean distances), yet not ordinary Euclidean distances, the more troublesome Weber issue: the mean works on squared messes up, yet the numerical center cutoff points Euclidean distances. K-medians and k-medoids, for example, can be utilized to track down better Euclidean responses. Despite the fact that the issue is troublesome mathematically (NP-hard), effective heuristic methodologies rapidly show up at a nearby ideal. Like the assumption boost strategy for blends of Gaussian dispersions, both k-means and Gaussian mixture modeling utilize gradual correction. The data are described by both using cluster centers; However, in contrast to the Gaussian mixture model, which allows for a variety of cluster types, k-means clustering identifies groups with similar geographic breadth. The k-nearest neighbor classifier, a notable regulated ML approach for classification that is much of the time mistook for k-means because of the terminology, is connected with the unattended k-means strategy. The k-nearest neighbor algorithm is used to group new data into the existing groups in the k-means cluster centers. This is known as the Rocchio strategy or nearest centroid arrangement.
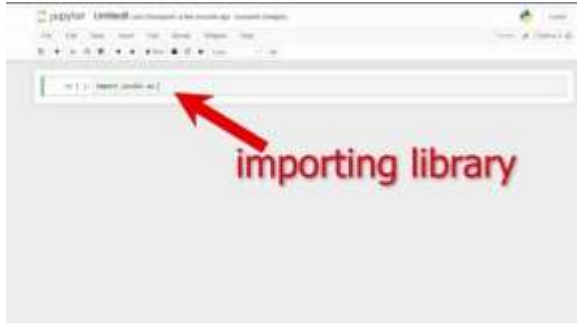
5. EXPERIMENTAL RESULTS

Fig.3: Import packages
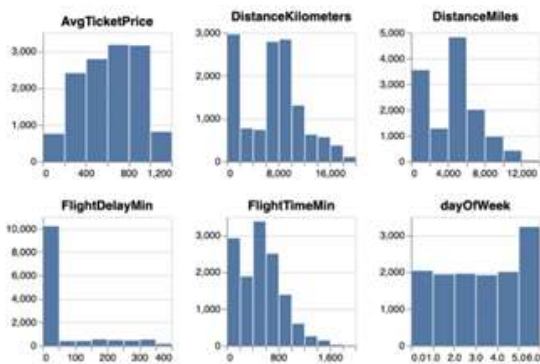

Fig.4: Data collection

Fig.5: Feature extraction

Fig.6: Web page

6. CONCLUSION

The execution was carried out using the Python programming language. We are able to determine which states have more or fewer criminals by looking at the numbers in each region. It is very gainful for authorities to be more educated. The process of organizing similar data elements into distinct categories is known as clustering. There are no parallels between the elements in one cluster, which are as similar as possible. While requiring less adjusts, the advanced k-means calculation yields an effective outcome with further developed accuracy of the last group. The most effective ways to utilize assets to forestall crime and an assortment of crime the executives procedures can be created involving crime exploration's discoveries later on.

7. FUTURE SCOPE

While requiring less adjusts, the enhanced k-means calculation yields an effective outcome with further developed accuracy of the last cluster. The best ways to use resources to prevent crime and a variety of crime management strategies can be

developed using crime research's findings in the future.

REFERENCES

1. ShijuSathyadevan M.S, Surya Gangadharan: Crime Analysis and Prediction Using Data Mining,inNetworksSoft Computing(ICNSC),(2014) First International Conference. https://ieeexplore.ieee.org/document/6906719.

2. H. Benjamin Fredrick David1,A.Suruliandi: Survey on crime analysis and prediction using data mining techniques.Department of Computer Science and Engineering,ManonmaniamSundaranar University, India. Ictact journal on soft computing, april(2017). https://www.researchgate.net/publication/3222541877_SURVEY_ON_CRIME_ANALAYSIS_AND_PREDICTION_USING_ DATA_MINING_TECHNIQUES.

3. JesiaQuader Yuki, Md.MahfilQuaderSakib, ZaishaZamal, Khan Mohammad Habibullah, Amit Kumar Das: Predicting Crime Using Time and Location Data(2019). https://www.researchgate.net/publication/335854157_Predicting _Crime_Using_Time_and_Location_Data.

4. Peng Chen,Justin Kurland, Modus Operandi: Time,Place,A Simple Apriori Algorithm Experiment for Crime Pattern Detection(2018).9th International Conference on IISA. https://www.researchgate.net/publication/330877862_Time_Place_and_Modus_Operandi_A_Simple_Apriori_Algorithm_Experiment_for_Crime_Pattern_Detection.

5. JyotiAgarwal,RenukaNagpal, RajniSehgal: Crime Analysis using K-Means Clustering(2013). https://www.researchgate.net/publication/269667894_Crime_Analysis_using_K-Means_Clustering

6. Malathi. A &Dr. S. SanthoshBaboo: An Enhanced Algorithm to Predict a Future Crime using Data Mining. International Journal of Computer Applications (0975 – 8887) Volume 21– No.1, May 2011

7. Khushabu A. Bokde,Tiksha P. Kakade, Dnyaneshwari S. Tumsare, ChetanG.Wadhai: Crime Analysis Using K-

Means Clustering(2018).https://www.ijert.org/crime-analysis-using-k-means-clustering.

8. Mrs.S. Sujatha, Mrs. A. Shanthi Sona: New Fast K-Means Clustering Algorithm using Modified Centroid Selection Method(2013).

9. Keerthi A, Remya M S, Nitha L: Detection of Credit Card Frauds Using Hidden Markov Model With Improved K-Means Clustering Algorithm (2015). https://www.ijarcs/article/download/1239/1227. [7] InformedHealth.org, "Cervical cancer: Overview," 14 Dec 2012. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK279259/. [Accessed Jan 2019].

[8] Cancer Report. February 2017 [cited WHO World Health Organization 25.09.2017]; Available from: http://www.who.int/mediacentre/factsheets/fs297/en/.

[9] Fernandes, K., J.S. Cardoso, and J. Fernandes, Transfer Learning with Partial Observability Applied to Cervical Cancer Screening, in Pattern Recognition and Image Analysis: 8th Iberian Conference, IbPRIA 2017, Faro, Portugal, June 20-23, 2017, Proceedings, L.A. Alexandre, J. Salvador Sánchez, and J.M.F. Rodrigues, Editors. 2017, Springer International Publishing: Cham. p. 243-250.

[10] Sen, T. and S. Das, An Approach to Pancreatic Cancer Detection using Artificial Neural Network, in Proc. of the Second Intl. Conf. on Advances in Computer, Electronics and Electrical Engineering (CEEE2013). 2013. p. 56-60.

[11] Fernandes, K., J.S. Cardoso, and J. Fernandes, Transfer Learning with Partial Observability Applied to Cervical Cancer Screening, in Iberian Conference on Pattern Recognition and Image Analysis. 2017: Faro, Portugal. p. 61-69.