IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 6, June 2024

DOI: 10.17148/IJARCCE.2024.13623

CRIME RATE PREDICTION AND ANALYSIS USING K-MEANS ALGORITHM

Samudrala Dinesh[1], Ponnala Rohit[2], Masood Khan Patan[3], Mrs.S.T.Ramya[4]

BE (Information Technology) Matrusri Engineering College Hyderabad, India[1-4]

Assistant Professor, Matrusri Engineering College Hyderabad, India[5]

Abstract: Crime analysis and prediction is a systematic approach for identifying the crime. This system can predict region which have high probability for crime occurrences and visualize crime prone area. Using the concept of data mining we can extract previously unknown, useful information from an unstructured data. The extraction of new information is predicted using the existing datasets. Crimes are treacherous and common social problem faced worldwide. Crimes affect the quality of life, economic growth and reputation of nation. With the aim of securing the society from crimes, there is a need for advanced systems and new approaches for improving the crime analytics for protecting their communities. We propose a system which can analysis, detect, and predict various crime probability in given region. This paper explains various types of criminal analysis and crime prediction using several data mining techniques.

Keywords: Machine learning,Supervised learning, Unsupervised learning, Prediction, Classification, Model evaluation, Data Preprocessing, Future Engineering.

I. INTRODUCTION

Day by day crime data rate is increasing because the modern technologies and hi- tech methods are helps the criminals to achieving the illegal activities .according to Crime Record Bureau crimes like burglary, arson etc. have been increased while crimes like murder, sex, abuse, gang rap etc. have been increased. Crime data will be collected from various blogs, news and websites. The huge data is used as a record for creating a crime report database. The knowledge which is acquired from the data mining techniques will help in reducing crimes as it helps in finding the culprits faster and also the areas that are most affected by crime .Data mining helps in solving the crimes faster and this technique gives good results when applied on crime dataset, the information on obtained from the data mining techniques can help the police department. A particular approach has been found to be useful by the police, which is the identification of crime

_hot spots _which indicates areas with a high concentration of crime. Use of data mining techniques can produce important results from crime report datasets. The very step in study of crime is crime analysis. Crime analysis is exploring, inter relating and detecting relationship between the various crimes and characteristics the crime. This analysis helps in preparing statistics, queries and maps on demand. It also helps to see if a crime in a certain known pattern or a new pattern necessary. Crimes can be predicted as the criminal are active and operate in their comfort zones. Once successful they try to replicate the crime under similar circumstances. The occurrences of crime depended on several factors such as intelligence of criminals, security of a location; etc. The work has followed the steps that used in data analysis, in which the important phases are Data collection, data classification, pattern identification, prediction and visualization. The proposed framework uses different visualization techniques to show the trends of crimes and various ways that can predicts the crime using machine learning algorithm.

II. LITERATURE SURVEY

[1] Literature survey is the main advance in programming improvement measure. Prior to building up the instrument it is important to decide the time factor, economy and friends strength. When these things are fulfilled, at that point the subsequent stage is to figure out which working framework and language can be utilized for building up the device.When the developers begin assembling the apparatus the software engineers need parcel of outer help. This help can be gotten from senior developers, from book or from sites. The major part of the project development sector considers and fully survey all the required needs for developing the project.

[2] Before developing the tools and the associated designing it is necessary to determine and survey the time factor, resource requirement, man power, economy, and company strength. Prior to building the framework the above thought are considered for building up the proposed framework. The significant piece of the undertaking advancement area considers and completely survey all the necessary requirements for building up the venture.

[3] For each undertaking Literature survey is the main area in programming improvement measure. Prior to

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 6, June 2024

DOI: 10.17148/IJARCCE.2024.13623

building up the instruments and the related planning it is important to decide and survey the time factor, asset prerequisite, labor, economy, and friends strength. When these things are fulfilled and completely surveyed, at that point the following stage is to decide about the product details in the separate framework, for example, what kind of working framework the venture would require and what are largely the important programming are expected to continue with the subsequent stage like building up the apparatuses, and the related activities. Here we have taken the general surveys of different creators and noted down the fundamental central issues with respect to their work. In this venture literature survey assumes a prevailing part in get assets from different areas and all the connected points that are exceptionally valuable under this segment.

III. METHODOLOGY

Apply the k-means algorithm using scikit-learn. Initialize the algorithm with a suitable number of clusters (k) and run it to partition the data into k clusters. Analyze the resulting clusters to interpret the patterns. Evaluate the characteristics of each cluster, including common types of crimes,areas

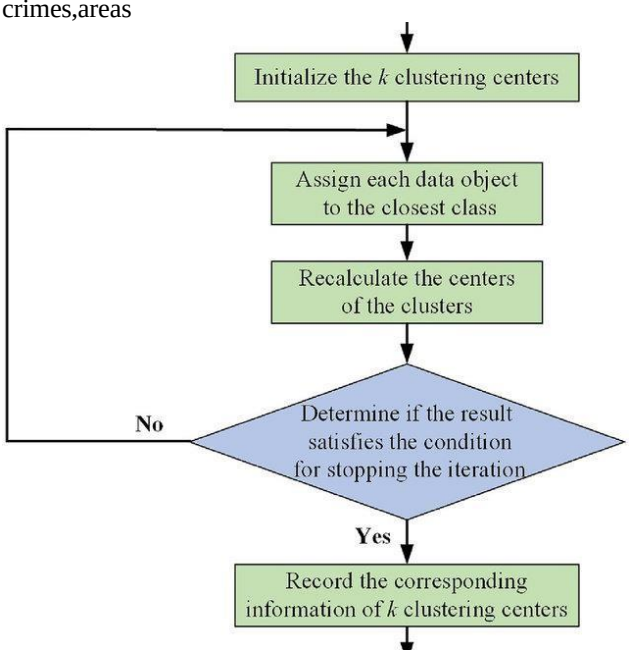


Fig1: Flow chart for methodology

A. Collection of data sets

In this stage two dataset are taken as input (multiple symptoms-based disease prediction and second dataset text file) with question and answers for chat bot application.

B. Understanding Features of data sets

Features in early life style disease dataset all these features are taken as input each disease is classified based on type of feature status 1 or 0 and disease name is used as label. Chatbot: For chat bot application text file data set is taken as input which has features as question and label as answer.

Split data into training dataset and testing dataset.

Data set is split in to two parts using test train split function (80 and 20) as test and train datasets. Train features are called as train x and labels as train y. These values are used to train algorithm and test data is used to check accuracy of each disease dataset**.**

C. Pre-processing the data

In this stage disease data set is processed to extract features and labels and chat bot text file dataset is pre- processed using NLP technique.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 6, June 2024

DOI: 10.17148/IJARCCE.2024.13623

D. **Apply ML decision tree algorithm to dataset to predict disease.**

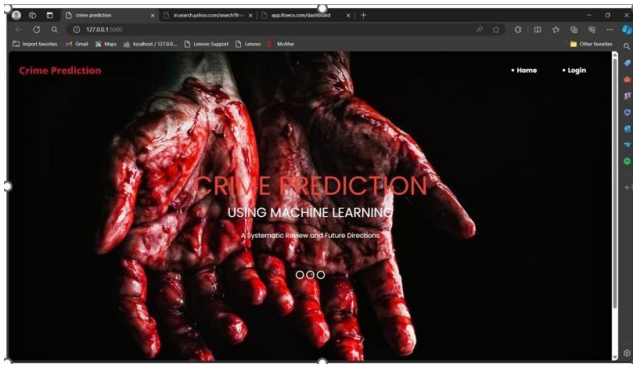
In this stage pre-processed dataset is taken as input of disease dataset and trained features and labels are given as input to fit function to train model. From web application user can select symptoms and get predicted disease.

E. **Accuracy Results**

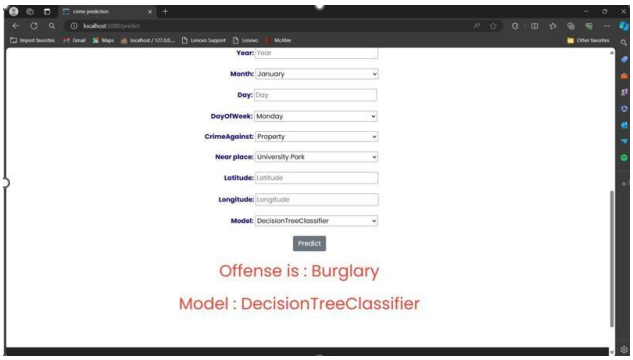
After training is done test set is given and input to algorithm to test accuracy of dataset. Overall, the accuracy results obtained from the testing phase serve as a critical validation of the effectiveness and utility of the developed AI healthcare bot system in facilitating patient consultation and diagnosis prediction.

Result Analysis

High incidence of petty theft and vandalism, primarily occurring in commercial areas. High rate of violent crimes such as assaults and robberies, concentrated in certain high-risk neighborhoods. Predominantly residential areas with moderate levels of burglary and domestic disturbances.



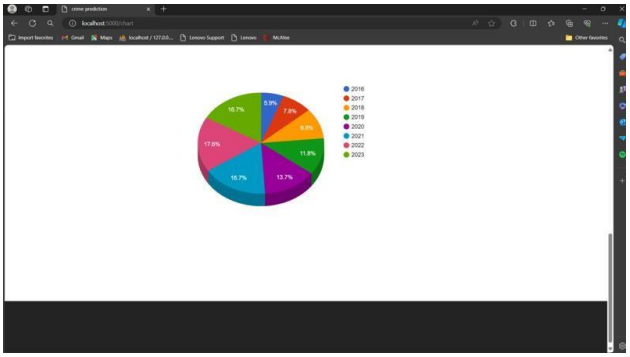
Login page



This figure shows the incident that will be happening on the particular date and particular timings.

Pie Chart

The below chart contains the crimes in various areas.



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 6, June 2024

DOI: 10.17148/IJARCCE.2024.13623

CONCLUSION

In this paper focused on building predictive models for crime frequencies per crime type per month. The crime rates in India are increasing day by day due to many factors such as increase in poverty, implementation, corruption, etc. The proposed model is very useful for both the investigating agencies and the police official in taking necessary steps to reduce crime. The project helps the crime analysis to analysis these crime networks by means of various interactive visualization. Future enhancement of this research work on training bots to predict the crime prone areas by using machine learning techniques**.** Since, machine learning is similar to data mining advanced concept of machine learning can be used for better prediction. The data privacy, reliability, accuracy can be improved for enhanced prediction. Crime analysis takes past crime data to predict future crime locations and time. Crime prediction for future crime is process that finds out crime rate change from one year to the next and projects those changes into the future.

REFERENCES

[1]. Ginger Saltos and Mihaela Coacea, An Exploration of Crime prediction Using Data Mining on Open Data, International journal of Information technology & Decision Making, 2017.
[2]. Shiju Sathyadevan, Devan M.S, Surya Gangadharan.S, Crime Analysis and Prediction Using Data Mining, First International Conference on networks & soft computing (IEEE) 2014.
[3]. Khushabu A.Bokde, Tisksha P.Kakade, Dnyaneshwari S. Tumasare, Chetan G.Wadhai B.E Student, Crime Detection Techniques Using Data Mining and K-Means, International Journal of Engineering Research & technology (IJERT) ,2018.H.Benjamin Fredrick David and A.Suruliandi,Survey on crime analysis and prediction using data mining techniques, ICTACT Journal on Soft computing, 2017.
[4]. Tushar Sonawanev, Shirin Shaikh, rahul Shinde, Asif Sayyad, Crime Pattern Analysis, Visualization And prediction Using Data Mining, Indian Journal of Computer Science and Engineering (IJCSE), 2015.
[5]. RajKumar.S, Sakkarai Pandi.M, Crime Analysis and prediction using data mining techniques, International Journal of recent trends in engineering & research,2019.
[6]. Sarpreet kaur, Dr. Williamjeet Singh, Systematic review of crime data mining, International Journal of Advanced Research in computer science, 2015.
[7]. Ayisheshim Almaw, Kalyani Kadam, Survey Paper on Crime Prediction using Ensemble Approach, International journal of Pure and Applied Mathematics,2018.

