Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 4, January-June: 2024

PREDICTING AND ANALYZING CRIME RATES WITH K-MEANS CLUSTERING ALGORITHM

I. Usha Komali[1] , G. Pushpa[2] , K. Anil[3]

[1] Assistant Professor, Department of Computer Applications, Aurora's PG College (MBA), Uppal, Hyderabad
Email: kommaanil48@gmail.com

[2]Assoc. Professor, Department of Computer Applications, Aurora's PG College (MBA), Uppal, Hyderabad
Email: gunde.latha@gmail.com

[3]Assistant Professor, Department of Computer Applications, Aurora's PG College (MBA), Uppal, Hyderabad
Email: kanil25@gmail.Com

ABSTRACT

The crime rate in India is rising daily. The influences of social media, recent technology advancements, and contemporary methods enable criminals to carry out their crimes in the current environment. A systematic approach that categorizes and looks at crime patterns is used for both crime analysis and crime prediction. Numerous clustering techniques are available for crime analysis and pattern prediction; however, they do not disclose all necessary information. The K means algorithm among them offers a more accurate means of forecasting the outcome. Predicting the age groups with greater propensity for crime and the areas with higher crime rates was the primary goal of the planned study project. In order to reduce the time complexity and boost efficiency in the outcome, we provide an improved K means method.

KEYWORDS: Clustering, k-means Algorithm, Crime

1. INTRODUCTION

The crime rate is significantly rising daily. Since crime is neither random nor systematic, it cannot be predicted. Modern tools and high-tech techniques also assist criminals in carrying out their crimes. As per the Crime Records Bureau, there has been a decline in certain crimes like as burglary and arson, but a rise in other crimes like murder. Even while we are unable to foresee every potential victim of crime, we are able to identify the locations where it is most likely to occur. Although the anticipated outcomes cannot be guaranteed to be 100% accurate, they do demonstrate that our application, by supplying security in high-crime regions, contributes to a somewhat lower crime rate. Therefore, we must gather and assess crime data in order to create such a potent crime analytics platform.

2. LITERATURE SURVEY AND RELATED WORK

There are various papers which contributed to the study of sentimental classification of

citations. Based on the study of the papers, this project was proposed.

Paper-1Summary: ProposedbySutapatThirprungsri

Thepurposeofthisstudyistoexaminethepossibilityofusingclusteringtechnologyforcontinuous auditing. Automating fraud filtering can be of great value to preventive continuousaudits. In this paper, cluster-based outliers help auditors focus their efforts when evaluating grouplife insurance claims. Claims with similar characteristics have been grouped together and thoseclusterswithsmallpopulationhavebeenflaggedforfurtherinvestigations.Somedominantcha racteristics of those clusters are, for example, having large beneficiary payment, having hugeinterest amount and having been submitted long time before getting paid. This study examines theapplication of cluster analysis in accounting domain. The results provide a guideline and evidenceforthe potential application ofthis technique in the field ofaudit.

Paper-2Summary:ProposedbyK. ZakhirHussain

Crime analysis, a part of criminology, is a task that includes exploring and detecting crimes andtheir relationships with criminals. The high volume of crime datasets and also the complexity ofrelationships between these kinds of data have made criminology an appropriate field for applyingdata mining techniques. Identifying crime characteristics is the first step for developing furtheranalysis. The knowledge that is gained from data mining approaches is a very useful tool whichcan help and support in identifying violent criminal behaviour. The idea here is to try to captureyears of human experience into computer models via data mining and by designing a simulationmodel.

3.EXISTING SYSTEM

Crime analysis tool is developed using various distinct data mining methods. It supports the police officers for investigating crimes. Implementing a clustering algorithm on crime datasets enables analysis of crimes . It makes identification and analysis of various criminality trends over the years through their conclusion. The random initial starting points produced by K-means which gives results in the form of cluster that helps in reaching the local optima [8]. So to overcome this problem, the partitioned data along with the data axis with the highest variance for assigning the initial centroid for K- Means clustering was applied. So it is observed that the proposed technique uses a lesser number of iteration thereby reducing the clustering time. Using merge sort, K- means algorithm can be improved for clustering the Hidden Markov Model (HMM).

EXISTING SYSTEM DISADVANTAGES:

1. Less Accurancy
2. Low Efficieny

4. PROPOSED SYSTEM

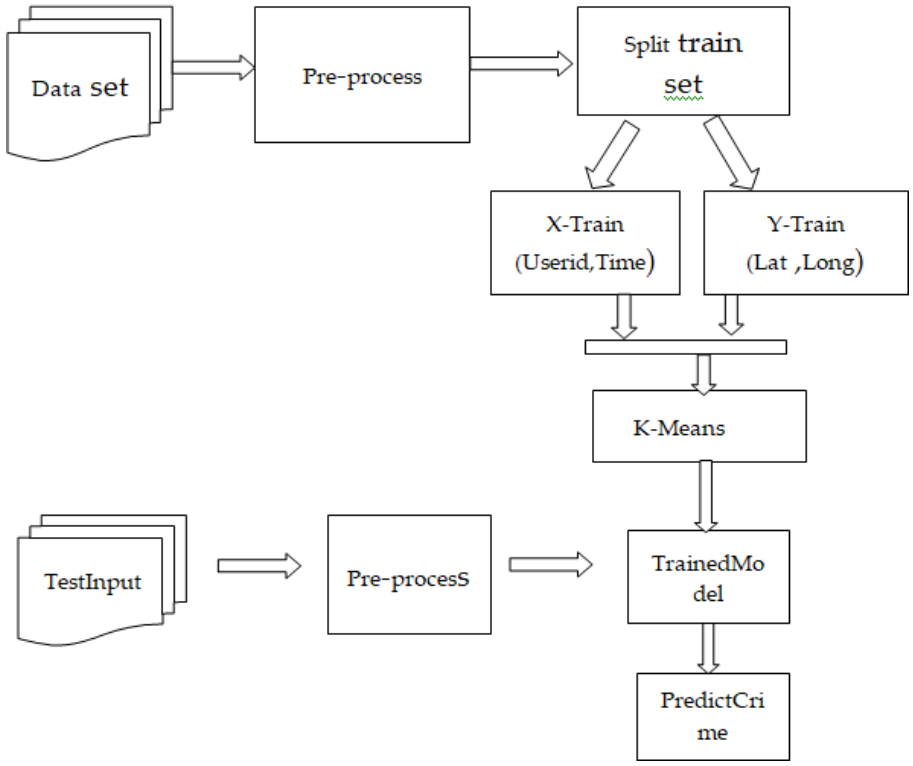
We are working on Spyder for implementation. Here we use a Spyder 3.7 version. Spyder is an integrated development environment for systematic programming in Python. Here we implemented different packages like matplotlib, numpy, sklearn, pandas, etc. Which helps to plot elbow graph and data frame table using a K-means clustering algorithm? Dataset is collected from Kaggle datasets and import datasets into Spyder in CSV format. We perform

normalization for finding the accurate number of clusters (k) using the elbow method. The elbow method performs k- means clustering on the obtained dataset for a range of values of k (2-15) and calculates the SSE. A line chart of the SSE is plotted for each value of k

PROPOSED SYSTEM ADVANTAGES:

1.high accuracy
2.high efficienc

SYSTEM ARCHITECTURE



METHODOLOGIES

Modules

1. Load Dataset:

Load data set using pandas read_csv() method. Here we will read the excel sheet data and store into a variable.

2. Split Data Set:

Split the data set to two types. One is train data test and another one is test data set. Here we will remove missing values from the dataset.

3.Train data set:

Train data set will train our data set using fit method. 80% of data from dataset we use for training the algorithm.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 4, January-June: 2024

4.Test data set:

Test data set will test the data set using algorithm. 20% of data from dataset we use for testing the algorithm.

5.Predict data set:

Predict () method will predict the results. In this step we will predict the ranking of the google play store app.

5. RESULTS AND DISCUSSION

FIG 2:- Next we view the number of total arrests for these crimes and the number of urban population in each state.

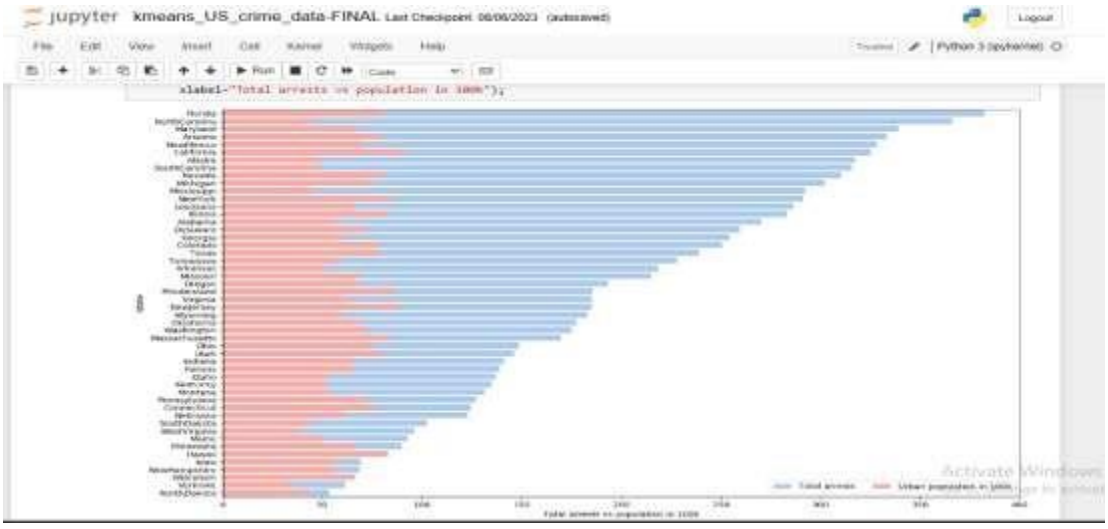


FIG 3 :-Next, we choose the optimal number of clusters using the elbow method by plotting the above table:

The kinks appear to be smoothening out after four clusters indicating that the optimal number of clusters is 4.Next, we divide the data into the chosen number of clusters.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 4, January-June: 2024

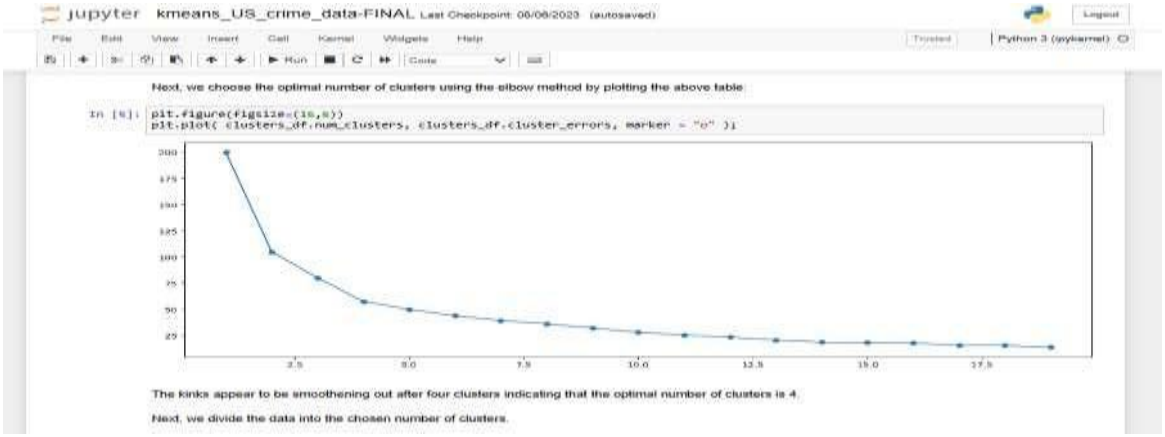


FIG 4 :-Analyzing the data pairwise - UrbanPop&TotalWe start by looking at the two main variables until digging into separate crime types.

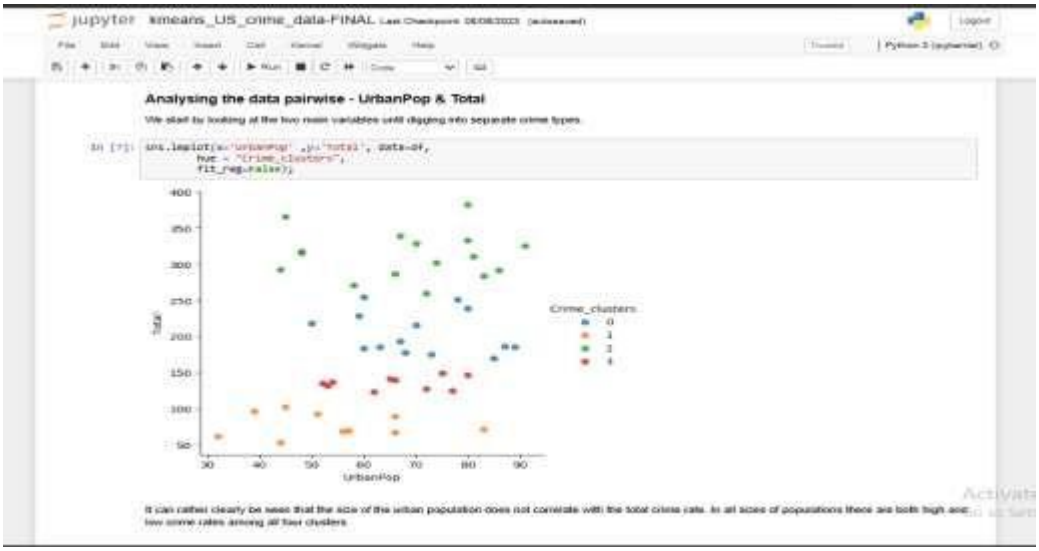


FIG 5 :- Analyzing the data pairwise - Murder &AssaultNext up, these two:

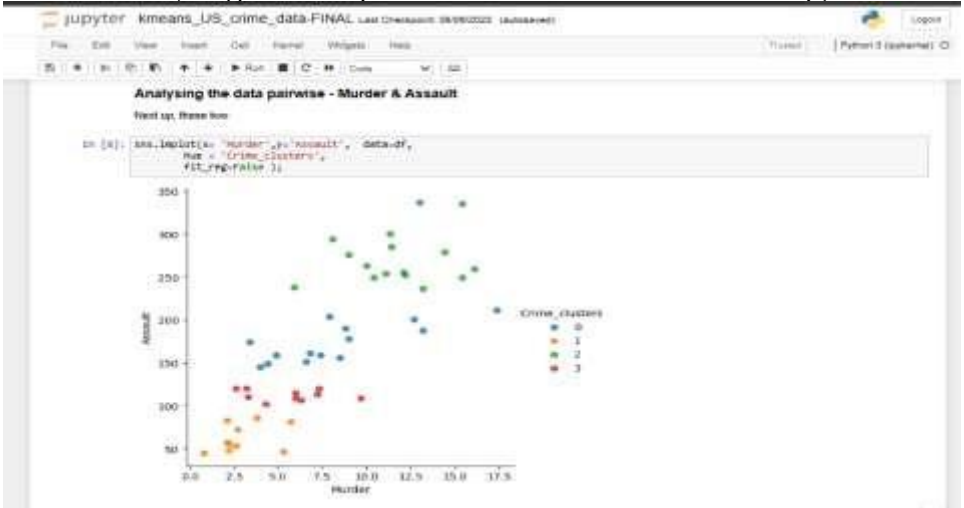


FIG 6 :- ontrary to murders vs. assaults, there is much more spread among the clusters when comparing murders vs. rapes. Some correlation is visible, though; low murder rates in a state seem to indicate lower number of rapes, as well. For the higher rate

states, the differences are more scattered

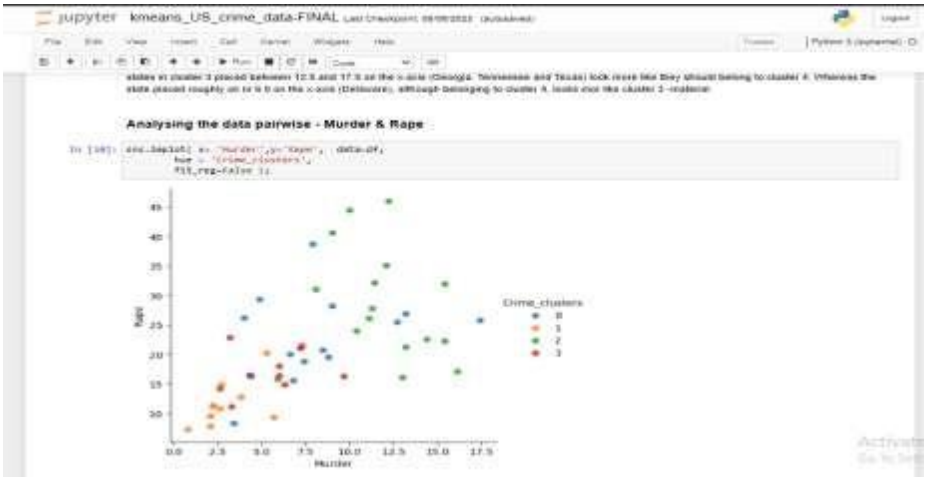


FIG 7:- This is interesting! The table quite well confirms the assumptions regarding variable correlations indicated also by the graphs. For example, murder and assault have the highest correlations,whereas the size of urba

6. CONCLUSION AND FUTURE SCOPE

This research has investigated class and prediction accuracy using entirely separate check sets. Based on the Bayes theorem, which demonstrated accuracy of above 90%, classification is carried out. We developed a model and trained a large number of news stories using this approach. We are feeding test data into the model to see improved outcomes during testing. Our approach uses an area's elemental properties and preprocesses them to provide the area's frequent patterns. The decision tree model is built using the pattern. By training on these recurring patterns, we create a model that corresponds to each location. Since patterns in crime vary throughout time, they cannot be static. By training, we mean that we are imparting knowledge to the system through specific inputs. On a given day, our software forecasts India's crime-ridden areas. If we focus on a specific state or region, the results will be more accurate. The fact that we are unable to foretell when a crime will be committed is yet another issue. Given that time plays a significant role in crime, we must forecast both the appropriate time and the areas that are likely to experience crime. The application's effectiveness in identifying frequent crime patterns and crime-prone locations for future prediction, as well as its speed of analysis, is demonstrated by the experimental findings. Based on the positive outcomes, we can apply other data mining techniques, such classification. Additionally, we are able to analyze a variety of datasets, including those from aid effectiveness studies, enterprise surveys, and datasets on poverty.

REFERENCES

1. DeBruin,J.S.,Cocx,T.K,Kosters,W.A.,Laros,J.andKok,J.N(2006)Dataminingapproachestocriminalcarreranalysis,”inProceedingsoftheSixthInternationalConferenceon DataMining(ICDM”06) ,Pp.171-177.
2. ManishandM.P.GuptaGupta1,B.Chandra11,200InformationSystem.
3. NazlenaMohamadAli1,MasnizahMohd2,HyowonLee3,AlanF.Smeaton3,FabioCrestani4 and ShahrulAzman Mohd Noah2 ,2010 VisualInteractiveMalaysiaCrimeNewsRetrieval System 7CrimeData Miningfor Indian Police.
4. Chung-

HsienYu,MaxW.Ward,MelissaMorabitoandWeiDing,"CrimeForecastingUsingDataMining Techniques",201111thIEEEInternationalConferenceonDataMiningWorkshops.

5. Tong Wang, Cynthia Rudin, Daniel Wagner, and Rich Sevieri. Detecting patterns of crimewithseriesfinder.InProceedingsoftheEuropeanConferenceonMachineLearningand Principles and Practice of KnowledgeDiscoveryinDatabases (ECMLPKDD 2013),2013.
6. Li Zhang, Yue Pan, and Tong Zhang. Focused named entity recognition using machinelearning.In Proceedingsofthe27th AnnualInternational.
7. Malathi. A and Dr. S. Santhosh Baboo. Article:an enhanced algorithm to predict a futurecrime using data mining. International Journal of Computer Applications, 21(1):1–6, May2011.Published byFoundationofComputerScience.
8. Eibe Frank and Remco R. Bouckaert. Naive bayes for text classification with unbalancedclasses. InProceedingsofthe10thEuropeanConferenceonPrinciple and Practice ofKnowledge Discovery in Databases, PKDD'06, pages 503–510, Berlin, Heidelberg, 2006.Springer-Verlag.
9. Wikipedia contributors.(9 July 2013), Stanford NLP. [Online].Available :http://www-nlp.stanford.edu/software/dcoref.shtml.Lastaccessed:24-Feb-2014,10:00AM.
10. Wikipediacontributors.(12May2014at19:05.),SeriesFinder.[Online].Available:http://en.wikipedia.org/wiki/Crime_analysis,Lastaccessed:12-Feb-2014, 12:00 PM.