MANTECH Publications

Journal of Data Mining and Knowledge Engineering
Volume 7, Issue 1, January-June 2022

Crime Analyzer

Shailesh Sahu[1], Gouri Kalbande[2], Prajakta Matey[3], Sanskruti Bachchuwar[4]

Professor[1], Students[2, 3, 4]

Department of Computer Science Engineering

Cummins College of Engineering for Women, Hingna, Nagpur, Maharashtra

Corresponding Author's Email: - *prajaktamatey711@gmail.com[3]*

Abstract

Crime is an act harmful not only to some individual but also to a community, society, or the state. Such acts are forbidden and punishable by law. Crimes are a global problem as they can affect economic growth and quality of life. It can also harm the nation under both social and economic conditions. Crime control is an inescapable step that is compulsory for the welfare and sustainable development of a nation. We know very well that in the digital world, it is not an easy task to expose the criminals and the vulnerable areas that are continuously getting affected by their wrongdoings. The police departments of every nation are also continuously working in a paced manner to overcome the crimes, criminals and their techniques. The difficulty in investigating a large amount of data regarding crimes and criminals has become a major challenge for police department officials. An approach is needed that can classify, systematically investigate and forecast the crimes that help to reduce the crime rate. There are various methodologies and paradigms which will help police officials to discover and eliminate crimes from society. Data mining empowers us with several practical and convenient ways to assess large and distinct sets of information. It helps to uncover hidden information from the large database of criminal records for investigating, controlling and preventing crime for organizations and users. Various researchers and data analysts gave their valuable time and knowledge to the field of data mining. The paper focuses primarily on presenting a short and snappy overview of various research papers focused on the techniques of data mining that have been applied in crime analysis. The

Page 1-4 © MANTECH PUBLICATIONS 2022. All Rights Reserved

dataset used in this paper is taken from the Kaggle. The analysis part identifies the number of unsolved crimes. The parameters used are analysis of the various factors like the year, month, location, gender, type of crime used in the unsolved crimes.

Keywords: *Data mining, Crime analysis, Criminal investigation, Machine Learning, KNN, Clustering*

INTRODUCTION

Crime analysis has been performed from different perspectives and in diverse dimensions by academicians. It has remained the area of interest for the researcher in different fields such as sociologists, criminologists, computer scientists, physicists, mathematicians and others. In the last few decades, there has been greater development in computer and Internet-related technology. The advancement in computer technology has contributed to crime analysis tasks to a greater extent. For instance, criminal network analysis is an essential type of crime analysis, which analyzes globally organized crimes.

There are four major phases of the framework, including (a) network creation; (b) network partitioning; (c) structural analysis; and (d) network visualization. Chen et al. (Chen et al., 2003), in the article, discussed four case studies using data mining techniques in the project COPLINK. These include: (a) entity extraction from police narrative reports; (b) detection of criminal identity deception; (c) authorship identification in cybercrimes, and (d) criminal network analysis. The advancement of Internet sources has privileged the news channels to the tremendous amount of news. It will just be a waste of cyber-space if the volume of information is not utilized properly (Nizamani, 2014). Despite these computational models, in the literature, there exist a number of mathematical models for the analysis of criminal activities. For instance, the articles (Lindelauf et al., 2008; Maeno&Ohsawa, 2009; Memon et al., 2007) exploited graph theory for the study of covert networks (Fellman et al., 2015; McGahan, 2015). The mathematical approach of game theory (Sandler et al., 2003) is also widely used by researchers for crime analysis. Researchers from the field of Physics have also contributed to the field of crime analysis. Galam (2002) modelled the role

Page 1-4 © MANTECH PUBLICATIONS 2022. All Rights Reserved

of passive supporters in the global threat of terrorism using the concept of percolation.

METHODOLOGY

Data Collection

Firstly, data collected generally refers to the crime and the criminal offender, often without recording the full details or perspective of victims. Such official data may lack information regarding the relationship of victim and offender, for instance. We collected data sets from Kaggle and later preprocessed the datasets.

Clustering

Clustering is an unsupervised task without having a priori knowledge by discovering groups of similar documents. There are two types of categories in clustering algorithms; they are the partitioned algorithm and the hierarchical algorithm. K-Means algorithm and the link clustering come under these two categories. K-Means and hierarchical clustering have many comparisons. In hierarchical clustering, the size of data increases as the computational expansive, K-Means is faster. It updates the centroid clusters with each iteration and reallocates each document by its nearest centroid. By this, we can say that it is an iterative algorithm.

3K-Means

K-means clustering is one of the methods of cluster analysis that aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean. K means algorithm complexity is O(tcn), where n instances, c clusters, and t is iterations and relatively efficient. It often terminates at a local optimum. Its disadvantage is applicable only when the mean is defined and the need to specify c, the number of clusters, in advance. It is unable to handle noisy data and outliers and not suitable to discover clusters with non-convex shapes. K-Means clustering investigation plans to partition n perceptions into k bunch, during which each perception includes a place with the bunch with the nearest centroid.

Algorithm Illustration Process

1. Initially, the number of clusters must be known, let it be k

2. The initial step is to choose a set of K instances as centres of the clusters.

3. Next, the algorithm considers each instance and assigns it to the cluster which is closest.

Page 1-4 © MANTECH PUBLICATIONS 2022. All Rights Reserved

4. The cluster centroids are recalculated either after the whole cycle of re-assignment or each instance assignment.

5. This process is iterated.

Data Visualization

and increased crime clearance. A data visualization platform enables agencies to quickly pinpoint areas that require immediate and long-term attention. Graphs, charts, and maps can clearly demonstrate which crimes have the biggest community impact and assist in strategic resource deployment strategies. Consequently, command staff can prioritize more effectively, which leads to improved community perception.

CONCLUSION

Using the concept of Machine Learning, we have built a model using training data set that have undergone data cleaning and data transformation. The model analyses the type of crime, crime rate, age, gender of the victim and Data visualization helps in the analysis of the data set and prediction of crimes. It also gives a statistical output representation of crimes with respect to different aspects in graphs, pie charts, etc.

REFERENCES

1. Ashish Sharma, Dinesh Bhuriya, Upendra Singh. "Survey of Stock Market Prediction Using Machine Learning Approach", ICECA 2017.
2. Python: The Complete Reference - Martin C. Brown Shiju Sathyadevan, M.S. Devan and S. Surya Gangadharan, “Crime Analysis and Prediction using Data Mining”, Proceedings of IEEE 1st International Conference on Networks and Soft Computing, pp. 406-412, 2014.
3. Yu-Yueh Huang, Cheng-Te Li and Shyh-Kang Jeng, “Mining Location-based Social Networks for Criminal Activity Prediction”, Proceedings of 24th IEEE International Conference on Wireless and Optical Communication, pp. 185-190, 2015.
4. De Bruin ,J.S.,Cocx,T.K,Kosters, W.A., Laros, J. and Kok, J.N(2006) Data mining approaches to criminal carrer analysis ,”in Proceedings of the Sixth International Conference on Data Mining (ICDM”06) ,Pp. 171-177.

Page 1-4 © MANTECH PUBLICATIONS 2022. All Rights Reserved