IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, November 2022
Impact Factor: 6.252

Analysis of Cybercrime using Machine Learning

Prof. Abhijit Khadke[1], Gauri G. Pande[2], Gaurav P. Rathod[3], Sudarshan M. Kawale[4], Siddhesh Y. Hande[5]
Professor, Department of Computer Engineering[1]
Students, Department of Computer Engineering[2,3,4,5]
SKN Sinhgad Institute of Technology and Science, Kusgaon (BK), Pune, Maharashtra, India

Abstract: *In today's world security is an aspect which is given higher priority by all political and government worldwide and aiming to reduce crime incidence. Cybercrime is criminal activity that either targets or uses a computer, a computer network or a networked device. Cyber Crime is technology-based crime committed by technocrats. This paper deals with Variants of cyber-crime. As data mining is the appropriate field to apply on high volume cybercrime dataset and knowledge gained from data mining approaches will be useful. Data mining k-Means algorithm is used for clustering. In k-means clustering, we are given a set of n data points in d-dimensional space Rd and an integer k and the problem is to determine a set of k points in Rd, called centres, so as to minimize the mean squared distance from each data point to its nearest centre.*

Keywords: Cybercrime, k-Means algorithm, clustering

I. INTRODUCTION

Cybercrime always involves some degree of infringement on the privacy of others or damage to computer-based property such as files, web pages or software. In the world of cyber age, cybercrime is spreading its roots considerably. e exploration emphasizes to attend cybersecurity from the standpoint of access control, particularly by detecting the cyber stoner and reporting felonious conduct to the cybercrime investigators so that they can probe and take legal conduct against the culprits. To resolve any case concerning the cybercrime, there are no data available beforehand, and hence, there occurs a need of a machine literacy model, in which data can be classified precisely through analysis, and by considering the features, the vaticination of the classes can be carried out. the ultimate thing is to enhance the security performance of the network so that it can be shielded from bushwhackers. With the help of cluster computing ways and real- time dataset, the performance evaluation of several cybercriminal discovery styles is anatomized. e assessment of the classifier's performance is also performed.

II. LITERATURE REVIEW

In present scenario criminals are becoming technologically sophisticated in committing crime and one challenge faced by intelligence and law enforcement agencies is difficulty in analyzing large volume of data involved in crime and terrorist activities therefore agencies need to know technique to catch criminal and remain ahead in the eternal race between the criminals and the law enforcement [1].

Cybercrime always involves some degree of infringement on the privacy of others or damage to computer-based property such as files, web pages or software. This paper is completely focused on cyber-crime analysis. This paper focuses on crime analysis by implementing clustering algorithm on crime dataset using and here we do crime analysis. From the clustered results it is easy to identify crime trend over years and can be used to design precaution methods for future[2].

In this paper k means clustering technique of data mining used to extract useful information from the high-volume crime dataset and to interpret the data which assist police in identify and analyse crime patterns to reduce further occurrences of similar incidence and provide information to reduce the crime [3].

This paper presents a k-Mean clustering using python. It is taking cyber-crime dataset and classification of peoples arrested in that year by cluster. It also helpful for other prescribe dataset. The clustering has been made through the k-mean algorithm which is based on cyber-crime dataset. We can do k-means algorithm using python [4].

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, November 2022
Impact Factor: 6.252

III. METHODOLOGY

3.1 Data pre-processing

In the data pre-processing phase the cyber crime data that is collected from the various forms are cleaned. The data collected from the website over the internet may contain noisy data and some missing values in the data. In order to reduce those noisy data here we apply the data pre-processing technique. After applying the pre-processing wewill extract the features from the cleaned data.

Figure: Data preprocessing

3.2 Algorithms

The main purpose of this algorithm is to differentiate clusters. Complexity of implementing algorithm is very low. The process is continued till all clusters were reached at last the clusters are grouped according to the number of clusters mentioned at beginning of the process.

Example: - There is a situation to find number of groups in a class who plays different games. Like some people play cricket, some play basketball, some play tennis, some may play football so there are different players in class so to classify them and group them clustering algorithm is used. which is reliable in finding distance and group them.

3.3 Partitioning

In this method, several partitions are separated and then evaluated based on given condition. Given a data-frame of n elements, it constructs k partitions of the data. Each element must belong to exactly one group. Each group must contain at least one object. The main objective of partition clustering algorithm is to divide the data points into K partitions. Each partition represents one cluster. The technique of partition depends upon certain objective functions like Logical operators by using matrix technique.

3.4 K-Mean Algorithm

K-means clustering is one of the method of cluster analysis which aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean.

Apply K-Mean Algorithm for k clusters containing n data sets.

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 centers of the clusters.
3. Next, the algorithm considers each instance and assigns it to the cluster which is closest.
4. The cluster centroids are recalculated either after the whole cycle of re-assignment or each instance assignment.
5. This process is iterated.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, November 2022
Impact Factor: 6.252

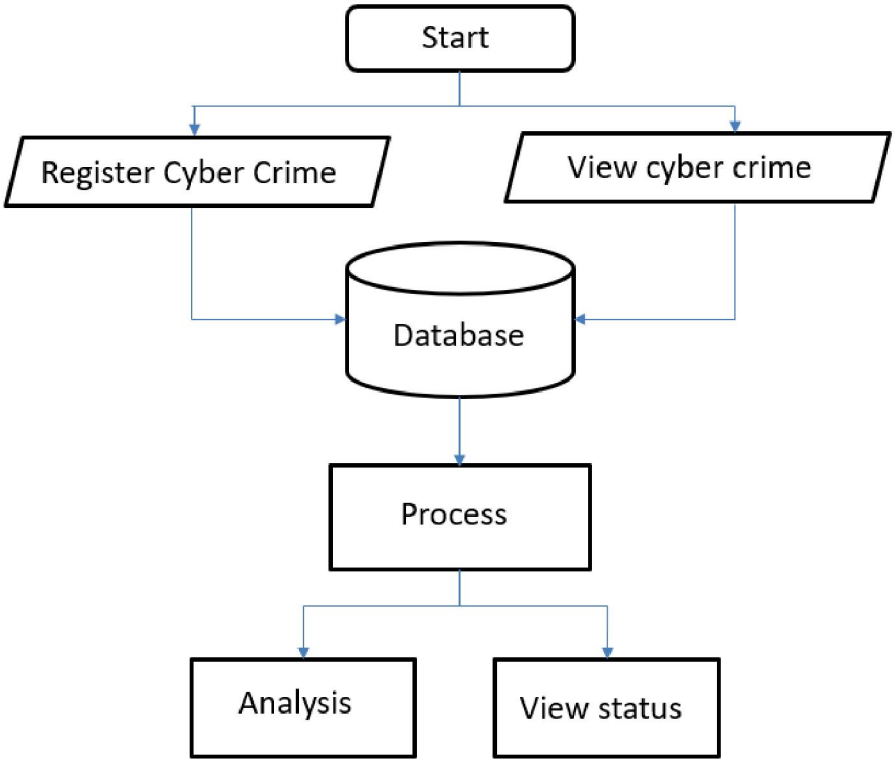

Figure: Model

IV. CONCLUSION

In the present world, cybercrime offenses are passing at an intimidating rate. As the use of the Internetis adding numerous malefactors, make use of this as a means of communication in order to commit a crime. The frame developed in our work is essential to the creation of a model that can support analytics regarding the identification, discovery and bracket of the integrated cybercrime offenses(structured and unshaped). The main focus of our work is to find the attacks that take advantage of the security vulnerabilities and dissect these attacks by making use of machine literacy ways. The end is that the developed frame will give the essential broad knowledge of cybercrime offenses in the society, enable them to consider the trouble geography of similar attacks and avoid the manifestation of the cybercrime offenses. From the results, it's apparent that the developed frame reduces the time consumption and homemade reporting process. It helps to identify the number of filing cases by incident wise and area-wise. This report is useful to prognosticate the cases and to take preventative way against cybercrime cases on certain hot- spot places linked.

V. ACKNOWLEDGEMENT

The authors would like to acknowledge the support and guidance provided by management and guides of SKN Sinhgad Institute of Technology and Science, Lonavala for providing the necessary support and guidance in carrying out this work.

REFERENCES

[1]. Arora T, Sharma M, Khatri SK. 2019. Detection of cyber crime on social media using random forest algorithm. In: 2nd International Conference on Power Energy, Environment and Intelligent Control (PEEIC). Piscataway: IEEE, 47–51.

[2]. Bharati A, Sarvanaguru RAK. 2018. Crime prediction and analysis using machine learning. International Research Journal of Engineering and Technology 5(9):1037–1042

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, November 2022
Impact Factor: 6.252

[3]. Kim S, Joshi P, Kalsi PS, Taheri P. 2018. Crime analysis through machine learning. In: 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON). Piscataway: IEEE, 415–420.

[4]. J. Agarwal, R. Nagpal, and R. Sehgal, "Crime analysis using k-means clustering", International Journal of Computer Applications, Vol. 83 – No4, December 2013.

[5]. Rasoul Kiani, Silamak Mahdavi and Amin Keshavarzi, "Analysis and Prediction of Crimes by Clustering and Classification", IJARAI, Vol. 4, Issue 8, pp. 1-7, 2015.

[6]. Lin YL, Chen TY, Yu LC. 2017. Using machine learning to assist crime prevention. In: 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI). Piscataway: IEEE, 1029–1030.

[7]. Vineeth KRS, Pandey A, Pradhan T. 2016. A novel approach for intelligent crime pattern discovery and prediction. In: International Conference on Advanced Communication Control and Computing Technologies (ICACCCT). Piscataway: IEEE, 531–538

[8]. Aishwarya Upadhyay, Arunesh, Akshay Chaudhari, Sarita Ghale, Prof. S. S. Pawar "Detection and Prevention measures for Cyberbullying and Online Grooming" International Conference on Inventive Systems and Control (ICISC-2017)

[9]. Chavan, V.S.; Shylaja, S.S. Machine learning approach for detection of cyber-aggressive comments by peers on social media network. In Proceedings of the 2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Kochi, India, 10–13 August 2015; Institute of Electrical and Electronics Engineers (IEEE): New York, NY, USA, 2015; pp. 2354–2358

[10]. Malathi. A, Dr. S. Santhosh Baboo. "An Enhanced Algorithm to Predict a Future Crime." International Journal of Computer Applications, 2011: 1-6.

[11]. Hassan M, Rahman MZ. 2017. Crime news analysis: location and story detection. In: 20th International Conference of Computer and Information Technology (ICCIT). Piscataway: IEEE, 1–6.

[12]. Ghankutkar S, Sarkar N, Gajbhiye P, Yadav S, Kalbande D, Bakereywala N. 2019. Modelling machine learning for analysing crime news. In: 2019 International Conference on Advances in Computing, Communication and Control (ICAC3). 1–5.

[13]. Feng M, Zheng J, Han Y, Ren J, Liu Q. 2018. Big data analytics and mining for crime data analysis, visualization and prediction. In: International Conference on Brain Inspired Cognitive Systems. Cham: Springer, 605–614.