© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 10 Issue: 04 | October - December 2019
Paper is available at http://www.jrps.in | Email : info@jrps.in

Analysis Cyber Crime Data Using K-mean Technique

1. **Mr. SAGAR DAROKAR**
(Research Scholor, Dept Of Information Technology,
Dr. C.V. Raman University Kota, Bilaspur (C.G.), India)
2. **DR. NEELAM SAHU**
(Associate Professor, Dept Of Information Technology,
Dr. C.V. Raman University Kota, Bilaspur (C.G.), India)

Abstract

"Data mining is the process of analyzing data from different perspectives and summarizing the results as useful information." Data Mining is the procedure which includes evaluating and examining large pre-existing database in order to generate new information which may be essential to the organization .The extraction of new information is predicated using the existing database.

ISSN : 2278-6848 03
9 772278 684800
© International Journal for Research Publication and Seminar

Keywords: Cyber Crime, Types of Cyber Crime, Kmean Clustering Algorithm, Python, Cyber Crime Dataset, Result Analysis.

1. Introduction

Cyber Crime is technology based crime committed by technocrats. This paper deals with Variants of cyber crime held in Chhattisgarh between 2005to 2018. Under this, the Age wise Clustering of arrested people has been displayed on basis of cybercrime in Chhattisgarh .data mining K-Mean algorithm is used for clustering.

2. Methodology

Cluster analysis or clustering is the process of partitioning a set of data objects into subsets. Each subset is a cluster, such that objects in a cluster resemble one another, yet dissimilar to objects in other clusters. In this context K-Mean methods may generate different clustering's on the Cyber Crime dataset. The partitioning is not performed by humans, but by the clustering algorithm K-Mean clustering algorithms were used for formation of clusters on cyber crime database. The data was collected from the National crime record bureau (2005 to

2018) data set converted into iris dataset using in python. The data set contains the various instances and the 4 attributes. The attributes are year, Crime type (act according), People arrested, Crime type. The algorithm is used in following manner:

2.1 K-Mean Technique:

The k-mean algorithm takes the input parameter, k, and a Partitions of n objects into k clusters so that the resulting intracluster similarity is high but the intracluster similarity is low. Cluster similarity is measured in regard to the mean value of the objects in a cluster, which can be viewed as the cluster‘s centeroid or center of gravity.

Algorithm k-means. The k-means algorithm for partitioning, where each cluster’s center is represented by the mean value of the objects in the cluster.

Input:

K: the number of clusters

D: a data set containing n objects.

Output: A set of k clusters

Method:

1. Arbitrarily choose k objects from D as the initial cluster centers.
2. Repeat
3. Re assign each object to the cluster to which the object is the most similar, based on the mean value of the objects in the cluster.
4. Update the cluster means, i.e. calculate the mean value of the objects for each cluster.
5. Until no change.

3. Technology & Dataset

K-Mean Clustering is one of the popular clustering algorithms. The goal of this algorithm is to find groups (clusters) in the given data. **K-Means Clustering** is one of the popular clustering algorithms. The goal of this algorithm is to find groups (clusters) in the given data.We implement K-Means algorithm using Python packages: pandas, NumPy, scikit-learn, Seaborn and Matplotlib.

1) Pandas: Pandas is used to working with “relational” or “labeled” data both easy and intuitive. It aims to be the fundamental high-level building block for doing practical, real world data analysis in Python.

© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 10 Issue: 04 | October - December 2019
Paper is available at http://www.jrps.in | Email : info@jrps.in

2) Numpy: NumPy is used for N-dimensional array object and sophisticated (broadcasting) functions.
3) Scikit-learn:Scikit-learn provideK-Mean algorithms via a consistent interface in Python.
4) Seaborn:Seaborn is use for data visualization and a high-level interface for drawing attractive and informative statistical graphics.
5) Matplotlib:Matplotlib is used for 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms.

The data was collected from the National crime record bureau (2005 to 2018) data set converted into iris dataset using in python. The data set contains the various instances and the 4 attributes. The attributes are year, Crime type (act according IT ACT-0,IPC ACT-1), No.of crime (0,1) Act wise,People arrested(0,1) Act wise.

4. Result

In every model, the accuracy and the cost analysis plays an important role in the acceptance of that model for the application. Result Show elbow method according elbow method we 3 cluster are used result show three cluster symbol and numeric value. This operation performs in IT/IPC ACT Data Set.

K-Mean Cluster Result for Cyber Crime (IT/IPC ACT) Dataset

Attribute: 3

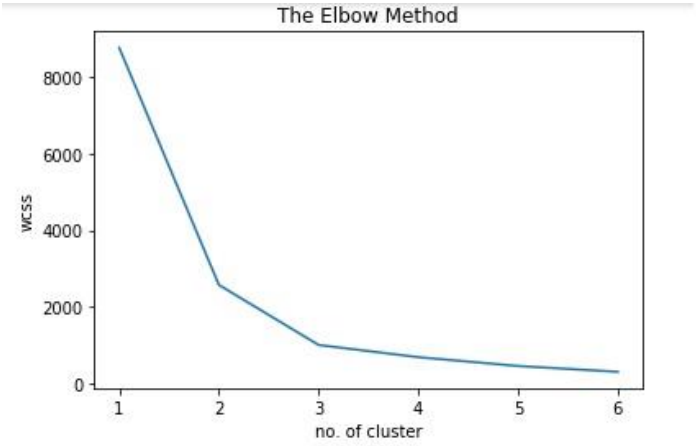
```
Data columns (total 3 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   year    15 non-null     int64
 1   IPCACT  15 non-null     int64
 2   ITACT   15 non-null     int64
dtypes: int64(3)
memory usage: 424.0 bytes
```

Test Mode: Evaluate data set

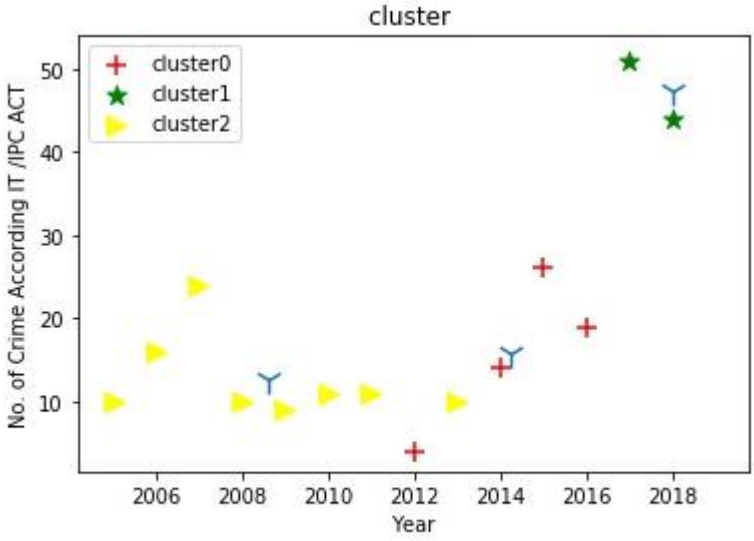
No. of iterations: 7

Result of Elbow Method

© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 10 Issue: 04 | October - December 2019
Paper is available at http://www.jrps.in | Email : info@jrps.in



No. of Cluster & Centroids :




```
Cluster0 (n=5):
Cluster1 (n=5):
Cluster2 (n=5):
            year   IPCACT   ITACT
Cluster
0         2012.0     10.0    13.2
1         2007.0     13.8     1.4
```

2 2017.0 37.4 41.2

Conclusion

This paper presents a K-Mean clustering using python. It is taking cyber crime dataset (Chhattisgarh Durg District) from 2005 to 2018) and classification of peoples arrested in that year by cluster. it also helpful for other prescribe dataset.

REFERENCES

1. Hemraj Saini, Yerra Shankar Rao, T.C.Panda, "Cyber-Crimes and theirImpacts: A Review", International Journalof Engineering Research andApplications(IJERA) ISSN: 2248-9622 www.ijera.com Vol. 2, Issue2, Mar-Apr 2012, pp.202-209
2. Varun Kumar, NishaRathee,"Knowledge Discoveryfrom Database using an Integration of clustering andClassification",IJACSA, vol 2 No.3,PP. 29-33,March2011.
3. Weka – Data Mining Machine Learning Software, http://www.cs.waikato.ac.nz/ml/.
4. Cheung Y.M. (2003). k-means: A New Generalisedk-means ClusteringAlgorithm. N-H Elsevier Pattern Recognition Letters 24, Vol 24(15), 2883–2893
5. Kanungo T., Mount D.M., Netanyahu N.S., Piatko C.D., Silverman R.and Wu A.Y. (2002). An Efficient k-means Clustering Algorithm: Analysisand Implementation. IEEE Transactions on Pattern Analysis and MachineIntelligence,Vol. 24, 881-892.
6. M. Inaba, N. Katoh, and H. Imai, "Applications of WeightedVoronoi Diagrams and Randomization to Variance-Basedk-clustering," Proc. 10th Ann. ACM Symp. Computational Geometry,pp. 332-339, June 1994.
7. Tapas Kanungo, Senior Member, IEEE, David M. Mount, Member, IEEE,An Efficient k-Means Clustering Algorithm:Analysis and ImplementationIEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 24, NO. 7, JULY 2002
8. D. Pollard, "A Centeral Limit Theorem for k-means Clustering,"Annals of Probability, vol. 10, pp. 919-926, 1982.
9. Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. From data mining to knowledge discovery: An overview. In Advances in Knowledge Discoveryand Data Mining, pp. 1 --34. AAAI Press, Menlo Park, CA, 1996.
10. Ester M., Kriegel H.-P., Sander J., Xu**X,** "ADensity-Based Algorithm for Discovering Clusters inLarge Spatial Databaseswith Noise", **In:** Proc. 2nd Int.Conf. on Knowledge Discovery and Data Mining,
11. Han Jiawei, Kamber M. Fan Ming, Meng Xiao-Fenget al. Translated. "Data Mining: Conceots and Techniques", Beijing: China Mechine Press, 200l(inChinese)