Date __/__/____

Assignment-2

* Title:- Consider a suitable data for clustering of data instances in different groups, apply different clustering techniques. Visualize the clusters using suitable tool.

* Objectives:- Understand various clustering types and how to implement the same using suitable tool (python). Using python to create K-means clustering models and heirarchical clustering models.

* Outcomes:- will be able to demostrate different clustering algorithms. Student will be able to demostrate and visualize the effectiveness of K-means clustering and heirarchical clustering using graphical capabilities of python.

* S/W and H/W Requirements:-
64-bit OS, python 3.8, Jupyter Notebook, machine with 64-bit processor.

* Theory:- K-Means clustering is a type of unsupervised learning, which is used when you have unlabeled data. The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable K. The algorithm works iteratively to assign each data point to one of K-groups based on provided features.
(i) Centroids of the K-clusters, which are used to label data.
(ii) Labels for training data.

Date ___/___/_____

To calculate the optimal no. of clusters, which are used in elbow method is used. A graph of within cluster sum of squares (WCSS) against no. of clusters is plotted to find out elbow point.

$$WCSS = \sum_{ck}^{cn} \left(\sum_{di \in c_i} \text{distance}(di, c_k)^2 \right)$$

Steps to Implement K-means.

(i) Choose the no. of clusters K.
(ii) Select K random points (centroids).
(iii) Assign each data point to the closest centroid. This forms K clusters.
(iv) Label the data points according to the clusters.
(v) Model is ready.

* Heirarchical clustering :-

Given a set of N items to be clustered and an N X N distance matrix.

Following are the steps for clustering.

(i) Start by assigning each item to it's own cluster.
(ii) Find the closest pair of clusters and merge them into a single cluster.
(iii) Compute the distances between the new cluster and each of the old clusters.
(iv) Repeat steps 2 & 3 until all items are clustered into a single cluster.

With the help of the distance matrix, we can plot a dendogram which displays heirarchical relationship among data items.

Date __/__/____

Saathi Notebooks

* Conclusion: Thus, we studied different clustering techniques and implemented them using python.