

EX NO: EXPLORATORY ANALYSIS : DEVELOP K-MEANS AND MST BASED CLUSTERING TECHNIQUES, DEVELOP
DATE: THE METHODOLOGY FOR ASSESSMENT OF CLUSTERS FOR GIVEN DATASET

AIM:

BACKGROUND THEORY:

K-MEANS CLUSTERING:

The algorithm operates by doing several iterations of the same basic process. Each training instance is assigned to the closest cluster in relation to the specified distance function that is applied to the instance and cluster center.

MST:

The Minimum Spanning Tree (MST) is a subset of a connected, undirected graph that connects all the vertices together with the minimum possible total edge weight. In other words, it is a tree-shaped structure that contains all nodes of the graph and has the least sum of edge weights.

PROCEDURE:

K-Means Clustering

1. Load Data:

- o Drag the "File" widget to the canvas.
- o Select your dataset file (e.g., dataset.csv).

2. Preprocess Data:

- o Drag the "Preprocess" widget to the canvas. And Connect the "File" to the "Preprocess".
- o Configure preprocessing steps such as imputing missing values, normalizing, or standardize it.

3. K-Means Clustering:

- o Drag the "K-Means" widget to the canvas.
- o Connect the "Preprocess" widget to the "K-Means" widget.
- o Set the number of clusters (k) in the "K-Means" widget.
- o Optionally, connect to the "Data Table" widget to inspect clustered data.

4. Inspect Clusters:

- o Drag the "Scatter Plot" widget to the canvas. And connect "K-Means" to "Scatter Plot".

5. Cluster Evaluation:

- o Drag the “Silhouette Plot” widget to the canvas.
- o Connect the “K-Means” widget to the “Silhouette Plot” widget to evaluate silhouette scores.
- o For the Elbow Method, manually adjust k and observe the within-cluster sum of squares.

Hierarchical Clustering (MST-like)

1. Load Data:

- o Follow the same steps as in K-Means to load and preprocess the data.

2. Hierarchical Clustering:

- o Drag the “Hierarchical Clustering” widget to the canvas.
- o Connect the “Preprocess” widget to the “Hierarchical Clustering” widget.
- o Configure linkage method, such as single linkage.

3. Inspect Dendrogram:

- o Drag the “Dendrogram” widget to the canvas.
- o Connect the “Hierarchical Clustering” widget to the “Dendrogram” widget.

4. Cluster Evaluation:

- o Drag the “Silhouette Plot” widget to the canvas.
- o Connect the “Hierarchical Clustering” to the “Silhouette Plot” to evaluate clusters.

OUTPUT:

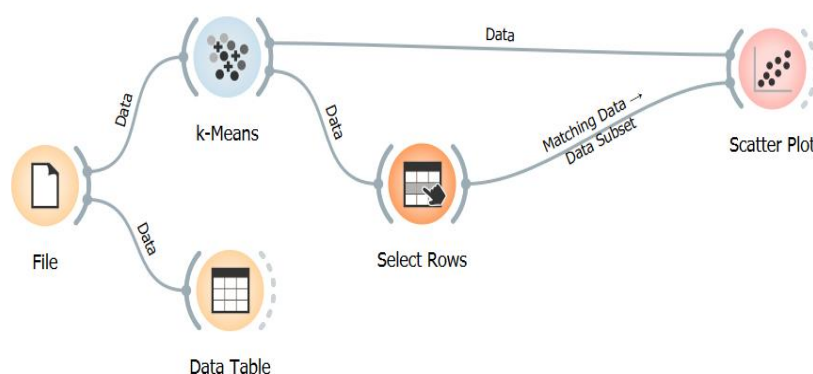


FIG: 3.1 : K-MEANS CLUSTERING

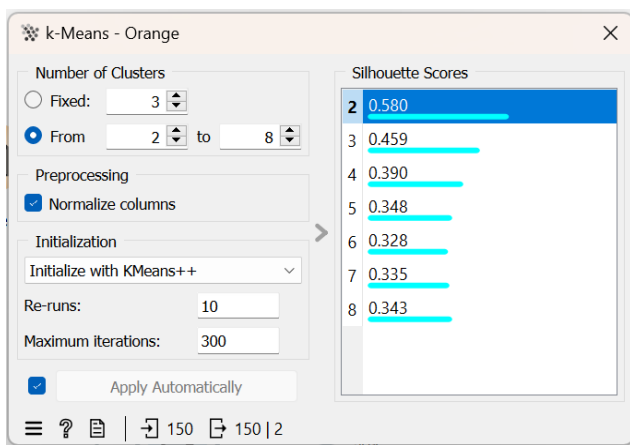


FIG: 3.1.1 : NO OF CLUSTERS AND SILHOUETTE SCORES

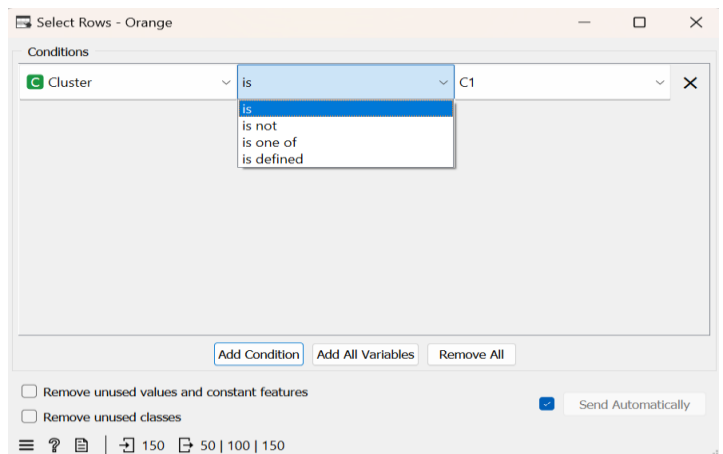


FIG: 3.1.2 : SELECTING ROWS



FIG : 3.1.3 : SCATTER PLOT

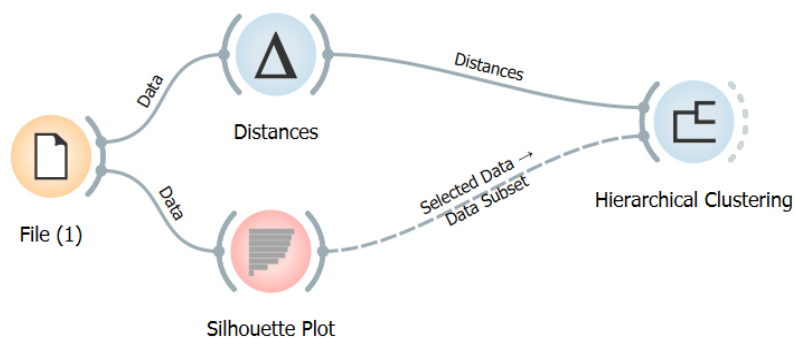


FIG : 3.2 : HIERARCHICAL CLUSTERING

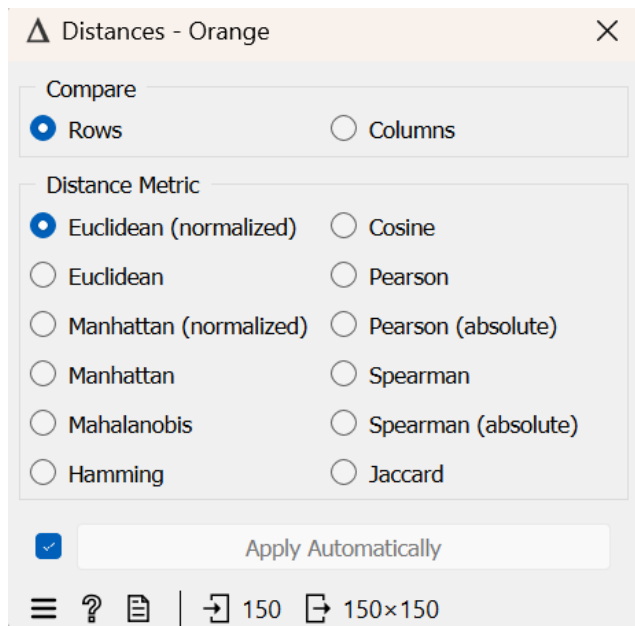


FIG: 3.2.1 : CALCULATING DISTANCE

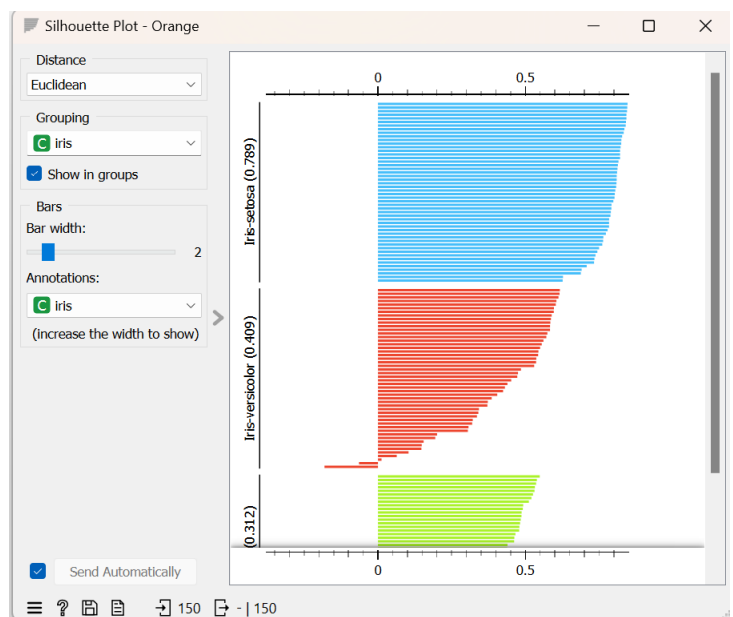


FIG: 3.2.2 : SILHOUETTE PLOT

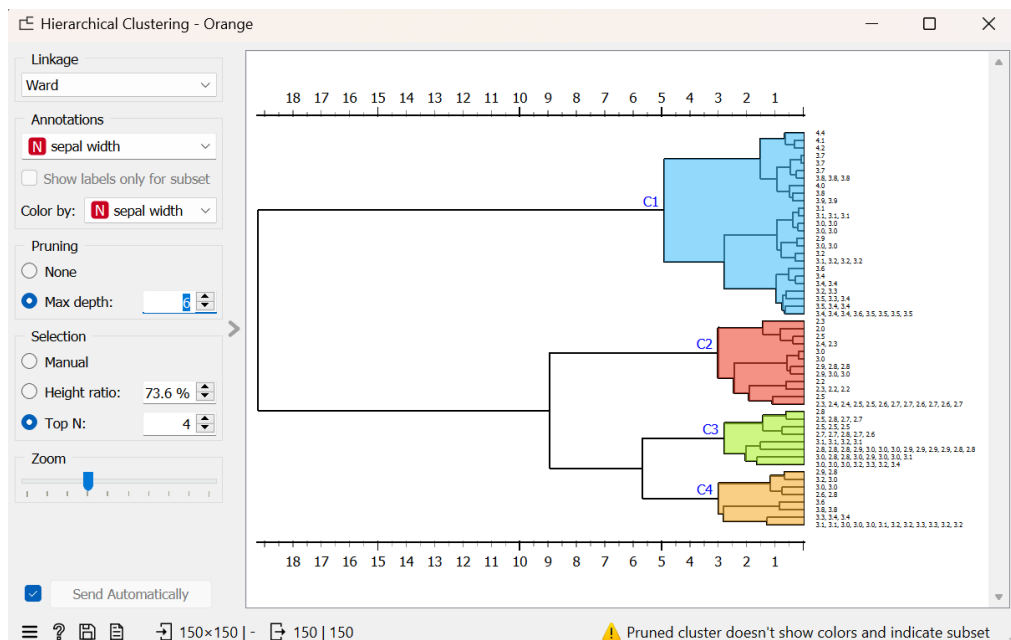


FIG : 3.2.3 : HIERARCHICAL CLUSTER PLOTING

RESULT: