REPORT

ASSIGNMENT-5 ES19BTECH11003

1)

a)

The clusters used for the k-Means algorithm is 2 and also the graph is plotted.

b)

DBSCAN algorithm is implemented and also finally there are 2 clusters formed and also so as to check the accuracy I have also compared my DBSCAN labels output with the inbuilt DBSCAN function from sklearn library.

c)

K-means Clustering	**DBScan Clustering**
This approach causes issues in the realm of anomaly detection since anomalous data points are assigned to the same cluster as "regular" data points.	The DBScan method, on the other hand, finds high-density areas that are separated from one another by low-density areas.
The number of clusters specified in K-means clustering is important.	The number of clusters need not be specified
The K-means clustering algorithm is unaffected by varying data point densities.	DBScan clustering is ineffective for sparse datasets or data points with varying densities.

d)

Pros of DBSCAN

- There is no need to define the number of clusters ahead of time.
- It works nicely with clusters of arbitrary shapes.
- DBSCAN is resistant to outliers and can detect them.

Cons of DBSCAN

- DBSCAN is not well adapted to define clusters when in-cluster densities are extremely diverse. Cluster characteristics are determined by a combination of the eps-minPts parameters. The algorithm can't generalize well to clusters with varied densities because we only give it one eps-minPts combination.

- In some circumstances, calculating an adequate neighborhood distance (eps) is difficult and necessitates domain knowledge.

Pros of K-Means

- It is straightforward, adaptable, and efficient. In comparison to Neural Networks, the simplicity of k-means makes it easy to explain the results.

- If there are any issues, the flexibility of k-means allows for simple adjustments.

- The efficiency of k-means indicates that the algorithm is effective at segmenting a dataset.

- When the centroids are recomputed, an instance can change cluster (move to another cluster).

- The clustering results are simple to understand.

Cons of K-Means

- When performing the analysis, the k-means algorithm will choose several locations at random from which to form clusters. This can be good or bad depending on where the algorithm starts. The center of the clusters is then recalculated until an adequate "center" for the number of clusters requested is found.

- The order of the data affects the final results.

- It does not allow for the development of the optimal set of clusters, and the number of clusters must be determined prior to the analysis.

2)
a)
We cannot attempt taking no_of_iterations as 100 as the minimum no of iterations for TSNE is 250.
Fixing perplexity to 30, as we repeat the algorithm with the number of iterations = 250,1000 and 2000 we can clearly observe that the points are being scattered more and more as the number of iterations increases which we can clearly observe if we see the graphs for iterations=250 and iterations=1000.

b)
The TSNE algorithm discovers patterns in data by identifying observed clusters based on data point similarity with multiple features. However, it is a dimensionality reduction algorithm, not a clustering algorithm. **Because it maps multidimensional data to a lower-dimensional space, the input features are no longer discernible**. As a result, you cannot draw any conclusions based solely on the result of TSNE. As a result, it is primarily a data exploration technique.