**Clustering Analysis Report:**

The clustering analysis is an essential step in understanding the underlying patterns and structures within the text data. This report provides an overview of the clustering process, evaluation metrics, and visualization of clustering results. Data Cleaning part is the same with classification model.

**1. Clustering Algorithm: KMeans**

- The KMeans algorithm is employed to cluster the text data into a specified number of clusters.

- In the provided code, `KMeans` is initialized with 4 clusters. You can adjust the number of clusters based on the nature of your data and desired clustering granularity.

**2. Clustering Evaluation Metrics:**

**- Silhouette Score:**

- The silhouette score measures how similar an object is to its cluster compared to other clusters.

- A higher silhouette score indicates better-defined clusters.

**- Sum of Squared Errors (SSE):**

- SSE measures the sum of squared distances between each data point and its assigned centroid.

- Lower SSE values indicate tighter clusters.

**3. Training Set Evaluation:**

- After fitting the KMeans model on the training data (`X\_train\_tfidf`), evaluation metrics are computed:

- Silhouette Score: Indicates the quality of clustering in the training set.

- SSE: Measures the compactness of clusters in the training set.

**4. Test Set Evaluation:**

- The trained KMeans model is used to predict clusters for the test data (`X\_test\_tfidf`).

- Silhouette score is calculated to assess the clustering performance on unseen data.

**5. Visualization of Clustering Results:**

- Dimensionality reduction using TruncatedSVD (`svd`) is performed to visualize clustering results in 2D space.

- The scatter plot displays the clusters identified by KMeans on the training set.

- Each data point is represented based on its two principal components, with colors indicating the assigned cluster.

**6. Interpretation:**

- Interpretation of the clustering results involves analyzing the distribution and separation of clusters in the visualization.

- A well-separated cluster indicates distinct patterns or topics within the text data.

- Visual inspection can provide insights into the effectiveness of the clustering algorithm in capturing underlying structures.