Report - Instagram User Dynamics using Single Linkage Agglomerative (Bottom-Up) Clustering Technique

Meduri Harshith Chowdary
21CS10042
IUHC-AS

Moto

This project aimed to analyze Instagram user dynamics using clustering techniques. The dataset consisted of attributes such as the number of posts, followers, followings, likes on previous posts, self-presenting posts, and gender of users. The goal was to cluster users into optimal groups and evaluate the clustering results

Analysis (also see ***report_logs.txt***)

- **Dataset Reading:** The dataset was successfully read in 0.00389 seconds.

- **K-means Clustering:** K-means clustering was performed initially with K=3. The clustering process took 0.287 seconds. The silhouette coefficient was calculated for various values of K (3, 4, 5, 6), with the highest silhouette score achieved at K=3, indicating well-separated clusters.

- **Optimal K Determination**: Finding the optimal value of K was time-consuming, taking approximately 10.92 seconds. The optimal K value was determined to be 3 based on the highest silhouette coefficient.

- **Hierarchical Clustering:** Hierarchical clustering was performed with the optimal K value of 3, which took significantly longer, approximately 122.55 seconds, due to the nature of hierarchical clustering.

- **Jaccard Similarity:** Jaccard similarity was calculated to compare the clustering results of K-means and hierarchical clustering. The process was quick, taking only 0.0002 seconds. The Jaccard similarity scores for each mapping revealed that mapping 2 had the highest similarity score of 0.46, indicating significant overlap between the clusters obtained from K-means and hierarchical clustering.

Conclusion

K-means clustering with an optimal K value of 3 yielded well-separated clusters, taking Euclid distance as the metric improves the result 2x. With which optimal K value would still be 3, but Jaccard Similarity of 0.864 is achieved in mapping 0

Euclid is better because this project deals with numeric data, where values matter.