

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

Meduri Harshith Chowdary  
21CS10042  
IUHC-AS

## Moto

The aim of this project was 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 0.0039 seconds.
- **K-means Clustering:** K-means clustering was performed initially with K=3. The clustering process took 0.0040 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 9.35 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 124.06 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 0 had the highest similarity score of 0.846, 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, which were further validated by hierarchical clustering. The Jaccard similarity analysis confirmed a high degree of overlap between the clusters obtained from both methods, particularly in mapping 0.