Cluster Analysis

Sindhu Bommali

agenda

HIERARCHICAL CLUSTERING WITH JACCARD DISTANCE

VISUALIZATION CLUSTERING AS DENDROGRAM

VISUALIZATION OF CLUSTERS IN BINARY RISK TAGS

HEATMAP OF CLUSTER CENTROIDS

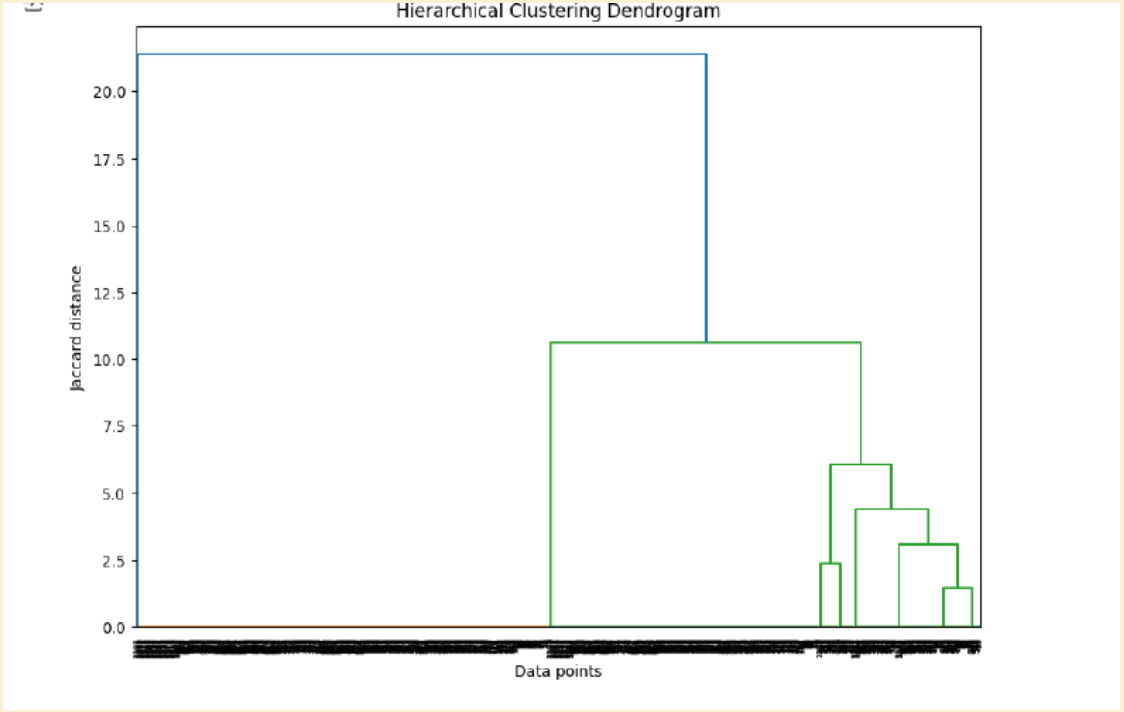
Hierarchical Clustering with Jaccard Distance

```
linkage_matrix
array([[ 90.        , 140.        ,   0.        ,  28.        ],
       [ 94.        , 141.        ,   0.        ,  29.        ],
       [ 98.        , 142.        ,   0.        ,  30.        ],
       [ 31.        ,  61.        ,   0.        ,   2.        ],
       [ 85.        , 144.        ,   0.        ,   3.        ],
       [ 23.        , 119.        ,   0.        ,  13.        ],
       [ 24.        , 146.        ,   0.        ,  14.        ],
       [ 26.        , 147.        ,   0.        ,  15.        ],
       [ 29.        , 148.        ,   0.        ,  16.        ],
       [ 30.        , 149.        ,   0.        ,  17.        ],
       [ 32.        , 150.        ,   0.        ,  18.        ],
```

- Using the complete linkage method, we perform hierarchical clustering.
- Complete linkage considers the maximum distance between observations in two clusters.

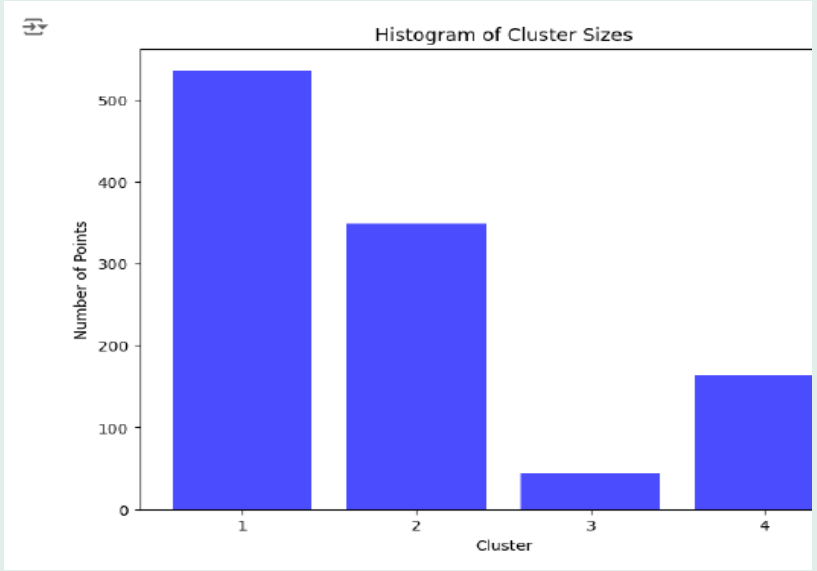
Visualization clustering as dendrogram

- The dendrogram shows how each cluster is linked, with the y-axis representing the distance (dissimilarity) at which clusters are merged.
- The height of the merge point indicates the similarity between clusters: lower heights suggest greater similarity.



Visualization Of clusters in Binary Risk Tags

	Is_honeypot				exploitation			
t	mean	std	median	count	mean	std	median	count
5	0.00000	0.000000	0.0	535	0.000000	0.000000	0.0	535
9	0.00000	0.000000	0.0	349	1.000000	0.000000	1.0	349
5	1.00000	0.000000	1.0	45	0.555556	0.502519	1.0	45
4	0.29878	0.459125	0.0	164	0.573171	0.496132	1.0	164



HEATMAP OF CLUSTER CENTROIDS