# **Machine Learning (CS564)**

# **Assignment 2 DBSCAN and Hierarchical clustering algorithms.**

**Experiment Overview:**

The experiment aimed to cluster a dataset of graduated students using two different clustering algorithms: DBSCAN (Density-Based Spatial Clustering of Applications with Noise) and Hierarchical Clustering. The dataset consisted of various attributes such as grad year, age, number of friends, and participation in different sports activities.

**DBSCAN Clustering:**

* **Cluster Counts:** DBSCAN formed a total of 183 clusters, with a significant number of points labeled as noise (-1). Some clusters contained a considerable number of points, while others had fewer points.
* **Silhouette Score:** The Silhouette Score for DBSCAN was approximately **0.1644**, indicating moderate clustering quality. This suggests that while DBSCAN successfully identified clusters, there might be room for improvement in the cluster separation.

**Hierarchical Clustering:**

Cluster Counts: Hierarchical Clustering formed a total of 3541 clusters, with varying numbers of points in each cluster. The number of clusters formed was significantly higher compared to DBSCAN.

Silhouette Score: The Silhouette Score for Hierarchical Clustering was approximately **0.4256**, indicating better clustering quality compared to DBSCAN. This suggests that Hierarchical Clustering was able to achieve better separation between clusters.

**Comparison:**

* Cluster Formation: DBSCAN formed fewer clusters compared to Hierarchical Clustering. DBSCAN's ability to automatically determine the number of clusters based on density resulted in a smaller number of clusters, whereas Hierarchical Clustering formed a larger number of clusters.
* Cluster Quality: While both algorithms successfully clustered the data, Hierarchical Clustering achieved better clustering quality as indicated by a higher Silhouette Score. This suggests that Hierarchical Clustering was more effective in separating data points into distinct clusters.

**Conclusion:**

Based on the experiment results, Hierarchical Clustering appears to be more suitable for clustering the given dataset of graduated students. However, further experimentation with different clustering algorithms and parameter tuning may be beneficial to explore other clustering possibilities.

Additionally, considering the high number of clusters formed by Hierarchical Clustering, further analysis may be required to understand the characteristics of each cluster and their significance.