

# LAB 10

## IRIS DATASET

It has features Id, SepalLengthCm, SepalWidthCm, PetalLengthCm and PetalWidthCm and the feature Species. There are three groups

- i) setosa
- ii) versicolor
- iii) virginica

This dataset will be used to train classification model. The 3 classes are in string format and we need to encode them to train the model.

## IMPORT LIBRARIES AND TOOLS

```
import numpy as np
import pandas as pd
from sklearn.datasets import load_iris
from sklearn.cluster import AgglomerativeClustering
import matplotlib.pyplot as plt
import seaborn as sns
```

## LOAD THE IRIS DATASET

```
iris = load_iris()
X = iris.data # Features
y = iris.target # not used in clustering
```

## Agglomerative Clustering

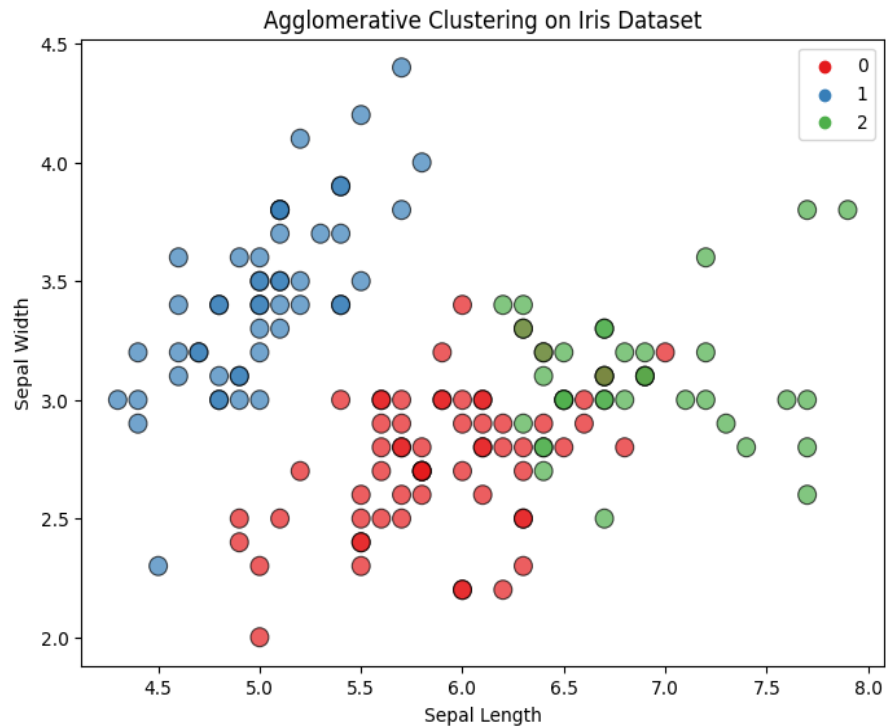
Agglomerative Clustering is a hierarchical clustering technique that builds the hierarchy from the bottom up by merging the closest pair of clusters at each step. It is widely used for unsupervised learning tasks like grouping similar data points together.

```
# Apply Agglomerative Clustering with 3 clusters (since there are 3 species
in the Iris dataset)
agg_clust = AgglomerativeClustering(n_clusters=3, affinity='euclidean',
linkage='ward')
y_agg = agg_clust.fit_predict(X)
```

```
c:\Users\kirti\AppData\Local\Programs\Python\Python311\Lib\site-
packages\sklearn\cluster\_agglomerative.py:983: FutureWarning: Attribute
`affinity` was deprecated in version 1.2 and will be removed in 1.4. Use
`metric` instead
warnings.warn(
```

```
# Visualize the Agglomerative Clustering results
plt.figure(figsize=(8, 6))
```

```
sns.scatterplot(x=X[:, 0], y=X[:, 1], hue=y_agg, palette='Set1', s=100,
edgecolor='black', alpha=0.7)
plt.title('Agglomerative Clustering on Iris Dataset')
plt.xlabel('Sepal Length')
plt.ylabel('Sepal Width')
plt.show()
```



## CONCLUSION

We have successfully implemented agglomerative clustering on iris dataset