Name : Devkumar Biswas

Class: BE(AI&DS)

Div: B

Subject : IR(CL-II Lab)

Roll no. : BEAD21267

**Assignment No. – 4**

**Problem Statement** : Implement Agglomerative hierarchical clustering algorithm using appropriate dataset.

Code:-

pip install --upgrade scikit-learn

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

from sklearn.datasets import load\_iris

from sklearn.cluster import AgglomerativeClustering

from sklearn.metrics import silhouette\_score

from scipy.cluster.hierarchy import dendrogram, linkage

# Step 1: Load the Iris dataset

iris = load\_iris()

X = iris.data

y = iris.target

# Step 2: Perform Agglomerative Hierarchical Clustering

# Remove 'affinity' parameter

agg\_clustering = AgglomerativeClustering(n\_clusters=3, linkage='ward')

labels = agg\_clustering.fit\_predict(X)

# Step 3: Calculate the silhouette score

silhouette\_avg = silhouette\_score(X, labels)

print(f'Silhouette Score: {silhouette\_avg:.2f}')

# Step 4: Plot the results

plt.figure(figsize=(10, 5))

# Scatter plot of clusters

plt.subplot(1, 2, 1)

plt.scatter(X[:, 0], X[:, 1], c=labels, cmap='rainbow', edgecolor='k', s=100)

plt.title('Agglomerative Clustering on Iris Dataset')

plt.xlabel('Sepal Length')

plt.ylabel('Sepal Width')

# Dendrogram

plt.subplot(1, 2, 2)

linked = linkage(X, 'ward')

dendrogram(linked, orientation='top', labels=iris.target\_names[y], distance\_sort='descending', show\_leaf\_counts=True)

plt.title('Dendrogram of Agglomerative Clustering')

plt.xlabel('Iris Species')

plt.ylabel('Distance')

plt.tight\_layout()

plt.show()

OUTPUT:-

