

ASSIGNMENT 8

24/04/24

NAME : SHRESTH SONKAR

REGNO : 20214272

GROUP : CS6D

TOPIC : DATA MINING LAB

CODE : CS-16202

```

import numpy as np
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
from sklearn.preprocessing import LabelEncoder
import pandas as pd

def euclidean_distance(point1, point2):
    return np.sqrt(np.sum((np.array(point1) -
np.array(point2))**2))

def distance_matrix(dataset):
    n = len(dataset)
    dist_matrix = np.zeros((n, n))

    for i in range(n):
        for j in range(i+1, n):
            dist_matrix[i, j] =
euclidean_distance(dataset[i], dataset[j])
            dist_matrix[j, i] = dist_matrix[i, j]

    return dist_matrix

def kmeans_clustering(k, dataset):
    kmeans = KMeans(n_clusters=k, random_state=0)
    cluster_labels = kmeans.fit_predict(dataset)
    return cluster_labels

def visualize_clusters(dataset, cluster_labels,
class_labels):
    unique_labels = np.unique(cluster_labels)
    colors = plt.cm.rainbow(np.linspace(0, 1,
len(unique_labels)))

    for label, color in zip(unique_labels, colors):
        cluster_data = np.array([dataset[i] for i in
range(len(dataset)) if cluster_labels[i] == label])
        plt.scatter(cluster_data[:, 0], cluster_data[:,
1], color=color, label=f'Cluster {label}')

    plt.xlabel('Feature 1')
    plt.ylabel('Feature 2')
    plt.title('K-means Clustering Result')
    plt.legend()
    plt.show()

```

```

if __name__ == "__main__":
    filename = '/content/drive/MyDrive/Colab Notebooks/
Buy_Computer.csv'
    df = pd.read_csv(filename)

    label_encoder = LabelEncoder()
    for column in df.columns:
        if df[column].dtype == 'object':
            df[column] =
label_encoder.fit_transform(df[column])

    dataset = df.drop(columns=['Buy_Computer']).values

    dist_matrix = distance_matrix(dataset)
    print("Distance Matrix:")
    print(dist_matrix)
    print()

    k = 2 # Number of clusters
    cluster_labels = kmeans_clustering(k, dataset)
    print("Cluster Labels:")
    print(cluster_labels)
    print()

    visualize_clusters(dataset, cluster_labels,
df['Buy_Computer'].tolist())

```

