Bahria University,

Karachi Campus

A logo with text on it

Description automatically generated

LAB EXPERIMENT NO.

09

LIST OF TASKS

|  |  |
| --- | --- |
| TASK NO | OBJECTIVE |
| **01** | Perform K-Mean Clustering in KNIME on IRIS dataset. |
| **02** | Perform K-Mean Clustering in python on IRIS dataset. |
|  |  |

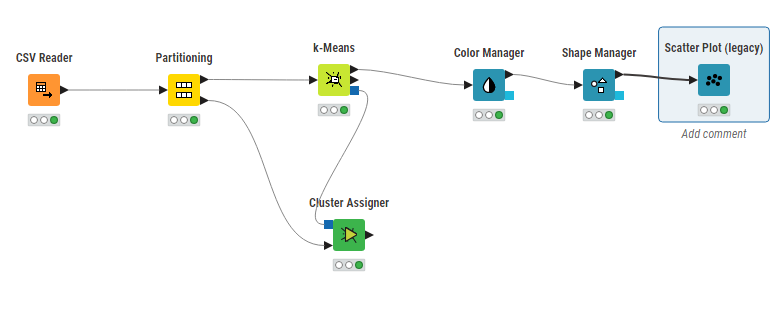
Submitted On:

01-05-2024

(Date: DD/MM/YYYY)

**Task No. 01:** Perform K-Mean Clustering in KNIME on IRIS dataset.

**Description**: You work as a data scientist at a botanic garden, and your team is interested in analyzing the Iris Setosa, versicolor, and virginica. Your goal is to use K-means clustering to group the flowers based on their measurements and determine if the measurements can be used to accurately identify the different types of iris flowers.

**Solution:**

A screen shot of a computer

Description automatically generated**Output:**

**Task No. 02:** Perform K-Mean Clustering in python on IRIS dataset.

**Description**: You work as a data scientist at a botanic garden, and your team is interested in analyzing the Iris dataset to identify patterns in the measurements of different types of iris flowers. The dataset contains measurements for the sepal length, sepal width, petal length, and petal width of three different types of iris flowers: setosa, versicolor, and virginica. Your goal is to use K-means clustering to group the flowers based on their measurements and determine if the measurements can be used to accurately identify the different types of iris flowers.

* 1. Load the Iris dataset into a Pandas dataframe.
* 2. Drop the 'species' column from the dataframe, as we will be performing unsupervised clustering.
* 3. Scale the data and use the elbow method to determine the optimal number of clusters for K-means clustering.
* 4. Perform K-means clustering on the scaled data using the optimal number of clusters.
* 5. Visualize the results of the clustering by creating a scatter plot of the data points, with each point colored based on its assigned cluster.
* 6. Analyze the results of the clustering to identify patterns in the measurements of the different types of iris flowers.

**Solution:**

import pandas as pd

from sklearn.datasets import load\_iris

# Load the dataset

iris = load\_iris()

data = iris.data

df = pd.DataFrame(data, columns=iris.feature\_names)

from sklearn.preprocessing import StandardScaler

from sklearn.cluster import KMeans

import matplotlib.pyplot as plt

scaler = StandardScaler()

scaled\_data = scaler.fit\_transform(df)

wcss = [] # Within-cluster sum of squares

for i in range(1, 11):

kmeans = KMeans(n\_clusters=i, init='k-means++', max\_iter=300, n\_init=10, random\_state=42)

kmeans.fit(scaled\_data)

wcss.append(kmeans.inertia\_)

plt.plot(range(1, 11), wcss)

plt.title('Elbow Method')

plt.xlabel('Number of clusters')

plt.ylabel('WCSS')

**A graph of a number of clusters

Description automatically generated**plt.show()

# Perform K-means clustering with the optimal number of clusters (k=3)

kmeans = KMeans(n\_clusters=3, init='k-means++', max\_iter=300, n\_init=10, random\_state=42)

kmeans.fit(scaled\_data)

clusters = kmeans.predict(scaled\_data)

# Add the cluster labels to the dataframe

df['Cluster'] = clusters

import seaborn as sns

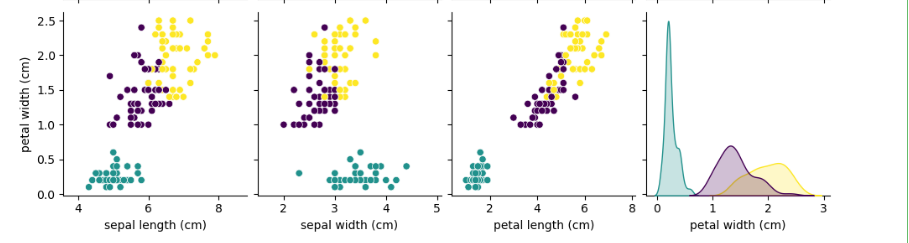
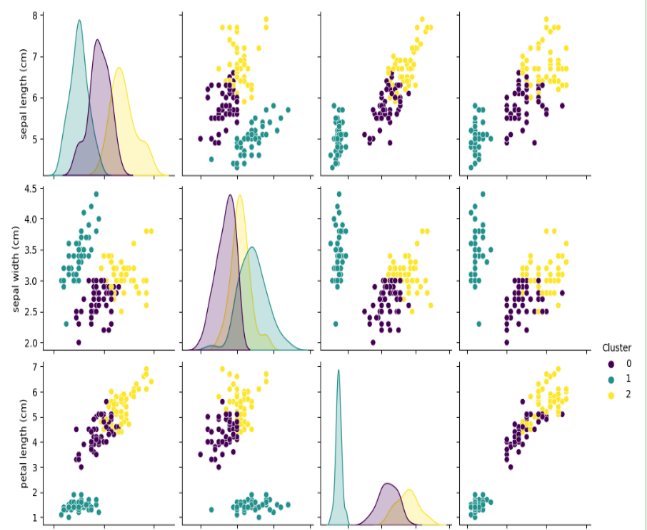
# Add the cluster labels to the original dataframe for visualization

df['Cluster'] = clusters

# Create a pairplot and color the points based on their cluster

sns.pairplot(df, hue='Cluster', palette='viridis')

plt.show()



**A graph of different colored dots

Description automatically generated with medium confidence**

print("Cluster centers:\n", kmeans.cluster\_centers\_)

print("Cluster distribution:\n", df['Cluster'].value\_counts())

mean\_values = pd.DataFrame(scaled\_data).groupby(clusters).mean()

print("Mean values per cluster:\n", mean\_values)

**Output:**

