|  |  |
| --- | --- |
| **Gerb-BMSTU_01** | **Министерство науки и высшего образования Российской Федерации**  Калужский филиал  федерального государственного бюджетного  образовательного учреждения высшего образования  ***«Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)»***  ***(КФ МГТУ им. Н.Э. Баумана)*** |

**ФАКУЛЬТЕТ** ***ИУК «Информатика и управление»***

**КАФЕДРА** \_\_***ИУК4 «Программное обеспечение ЭВМ, информационные технологии»***

**ЛАБОРАТОРНАЯ РАБОТА №6**

**«Кластерный анализ»**

**ДИСЦИПЛИНА: «Технологии анализа данных»**

|  |  |  |
| --- | --- | --- |
| Выполнил: студент гр. ИУК4-82Б | | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ ( Карельский М.К. )  (Подпись) |
| Проверил: | | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ ( Ерохин И.И. )  (Подпись) |
| Дата сдачи (защиты):  Результаты сдачи (защиты): | | |
|  | - Балльная оценка:  - Оценка: | |

Калуга, 2024

**Цель:** формирование практических навыков решения задач кластерного анализа.

**Задачи:** классификация объектов.

**Вариант 5**

Считать данные из файла iris\_df.csv в структуру DataFrame. Построить график показывающий выбор числа кластеров. Построить график кластеризации для Sepal Length, Petal width. С помощью k-means построить их центроиды. Провести иерархическую кластеризацию методом average. Labels = Sepal Length. Построить дендрограмму.

**Листинг:**

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.cluster import KMeans

from scipy.cluster.hierarchy import dendrogram, linkage

iris\_df = pd.read\_csv('iris\_df.csv')

sns.set(style='whitegrid')

sns.pairplot(iris\_df, hue='Species', markers=["o", "s", "D"])

plt.show()

X = iris\_df[['Petal Length', 'Petal Width']]

kmeans = KMeans(n\_clusters=3)

kmeans.fit(X)

centroids = kmeans.cluster\_centers\_

plt.figure(figsize=(8, 6))

plt.scatter(X['Petal Length'], X['Petal Width'], c=kmeans.labels\_, cmap='viridis', s=50, alpha=0.5)

plt.scatter(centroids[:, 0], centroids[:, 1], c='red', marker='X', s=200)

plt.xlabel('Petal Length')

plt.ylabel('Petal Width')

plt.title('K-means Clustering')

plt.show()

Z = linkage(X, method='single')

plt.figure(figsize=(10, 7))

dendrogram(Z, labels=iris\_df['Species'].values)

plt.title('Hierarchical Clustering Dendrogram')

plt.xlabel('Species')

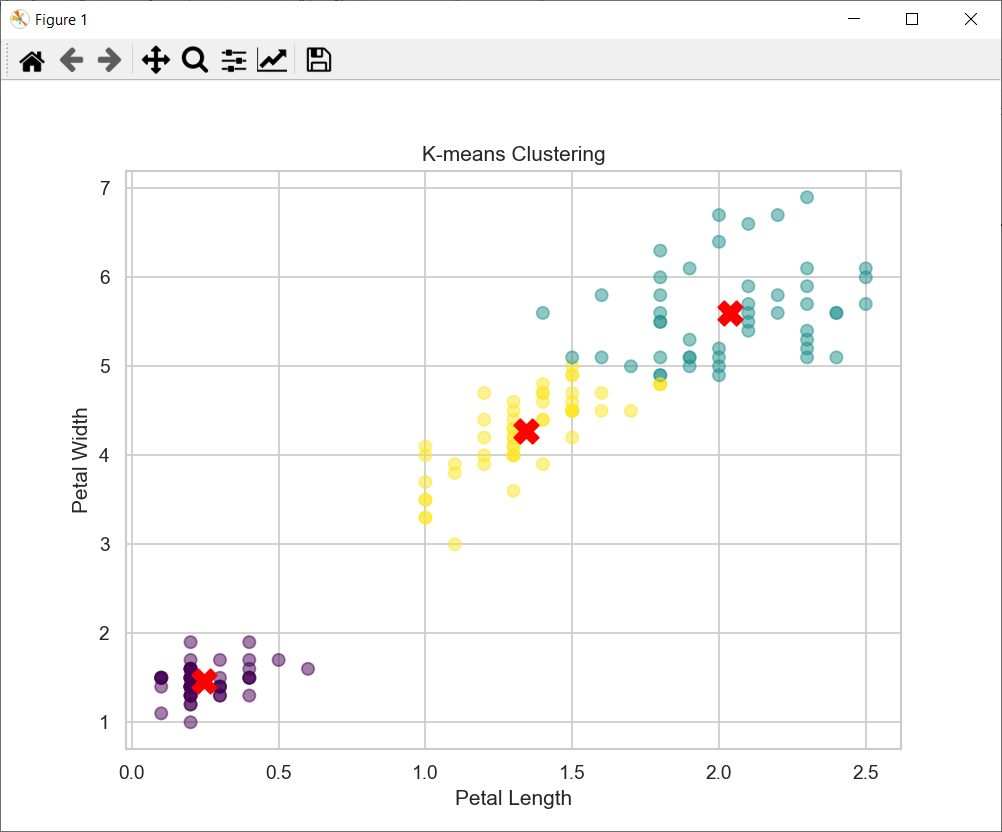
plt.ylabel('Distance')

plt.show()

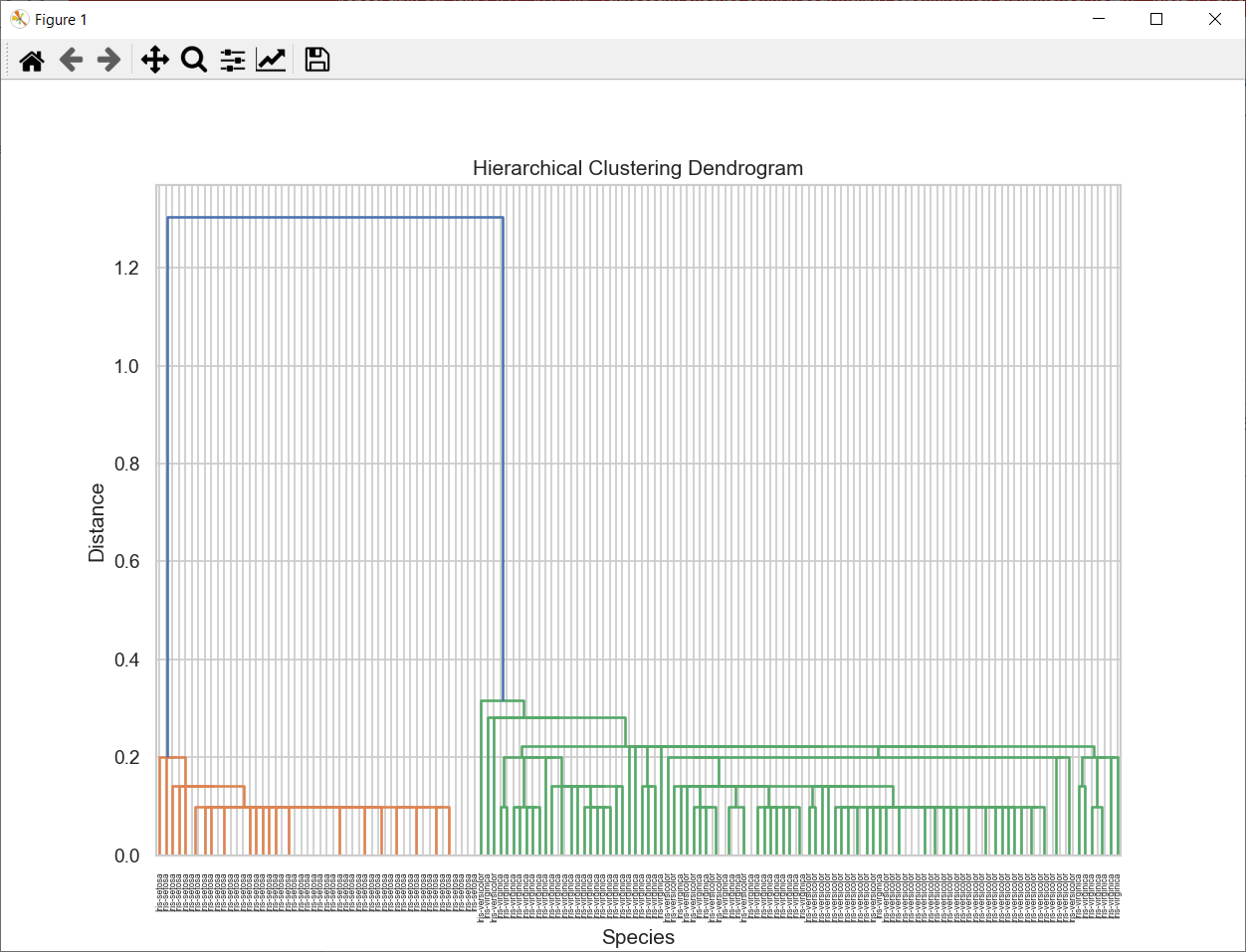
**Результат:**



**Рис. 1.** График выбора числа кластеров



**Рис. 2.** Кластеризация



**Рис. 3.** Дендограмма

**Вывод:** в ходе выполнения лабораторной работы были получены практические навыки решения задач кластерного анализа.