|  |
| --- |
| МИНОБРНАУКИ РОССИИ |
| Федеральное государственное бюджетное образовательное учреждение высшего образования **«МИРЭА − Российский технологический университет»**  **РТУ МИРЭА** |

**Институт информационных технологий (ИИТ)**

**Кафедра прикладной математики (ПМ)**

**ОТЧЕТ ПО ПРАКТИЧЕСКОЙ РАБОТЕ**

по дисциплине «Технологии и инструментарий анализа больших данных»

**Практическая работа № 6**

|  |  |  |  |
| --- | --- | --- | --- |
| Выполнил студент группы ИНБО-01-17 | ИВБО-06-20, Чурилов А.К. | (подпись) | |
| Принял | Парамонов А.А. | (подпись) | |
| Отчет представлен | «23» ноября 2023 г. | |  | |

Москва 2023 г.

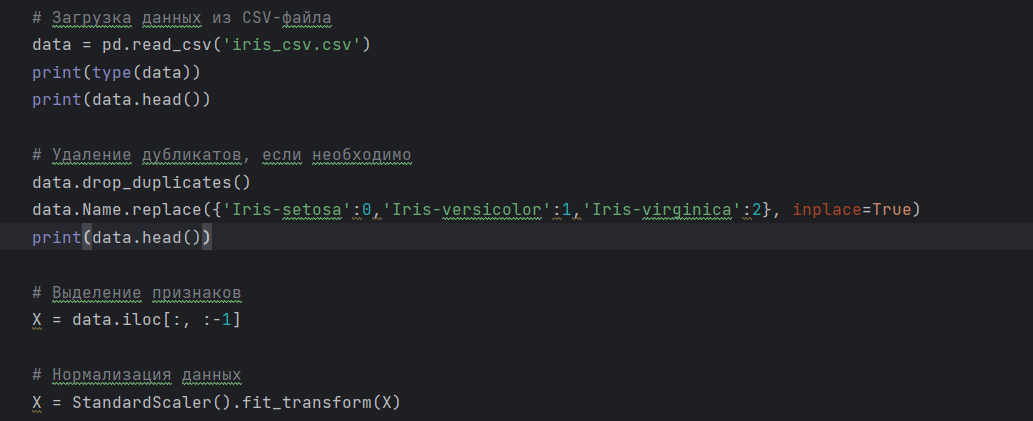
**Задание**

1. Найти данные для кластеризации. Данные в группе не должны повторяться! Внимание, если признаки в данных имеют очень сильно разные масштабы, то необходимо данные предварительно нормализовать.
2. Провести кластеризацию данных с помощью алгоритма k-means. Использовать «правило локтя» и коэффициент силуэта для поиска оптимального количества кластеров.
3. Провести кластеризацию данных с помощью алгоритма иерархической кластеризации.
4. Провести кластеризацию данных с помощью алгоритма DBSCAN.
5. Визуализировать кластеризованные данные с помощью t-SNE или UMAP если данные многомерные. Если данные трехмерные, то можно использовать трехмерный точечный график.
6. Оформить отчет о проделанной работе. Сделать выводы.

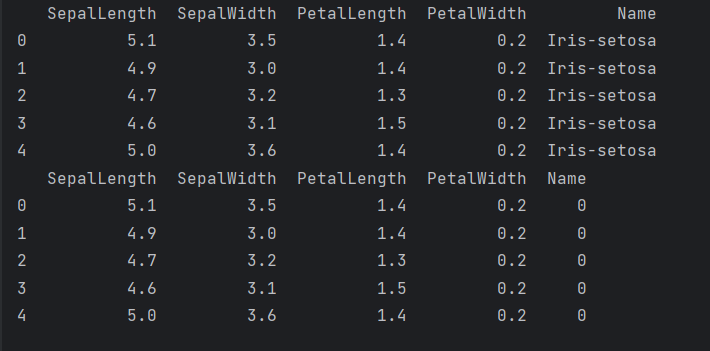
**Ход работы**

**Задание 1**

Найти данные для кластеризации. Данные в группе не должны повторяться! Внимание, если признаки в данных имеют очень сильно разные масштабы, то необходимо данные предварительно нормализовать.

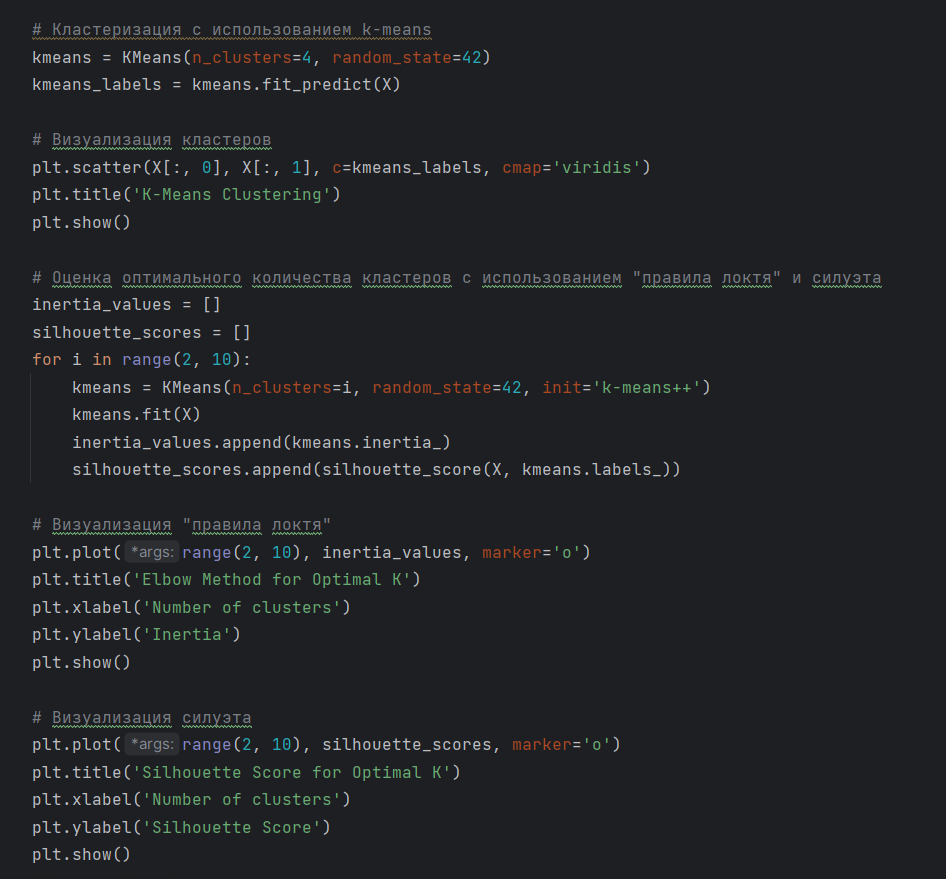
****

**Рисунок 1 – Задание 1**

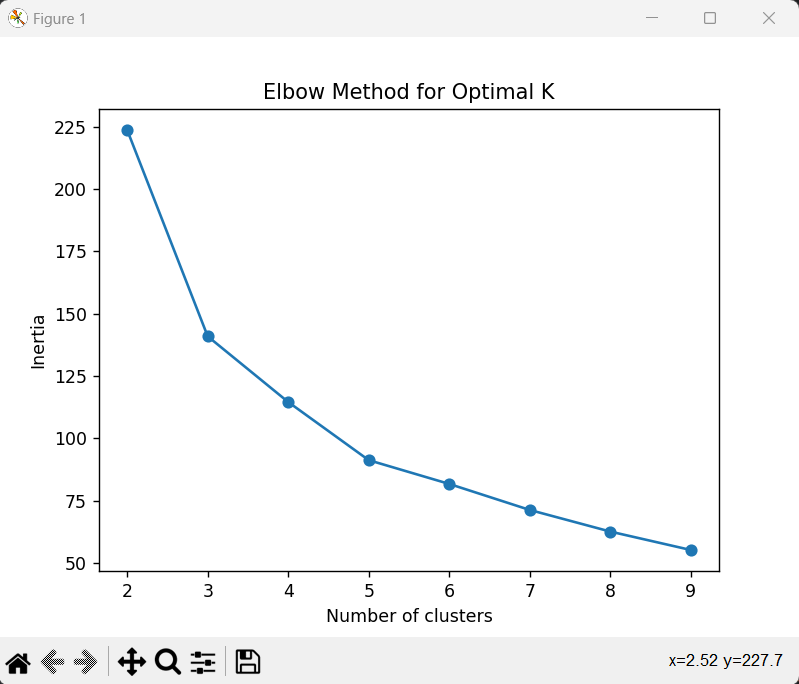
****

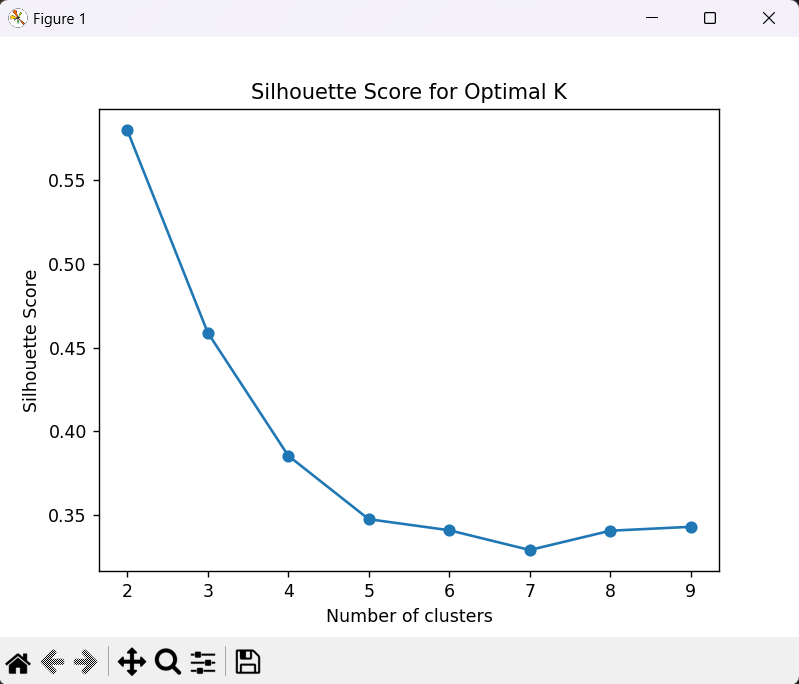
**Задание 2**

Провести кластеризацию данных с помощью алгоритма k-means. Использовать «правило локтя» и коэффициент силуэта для поиска оптимального количества кластеров.

****

**Рисунок 2 – Задание 2**

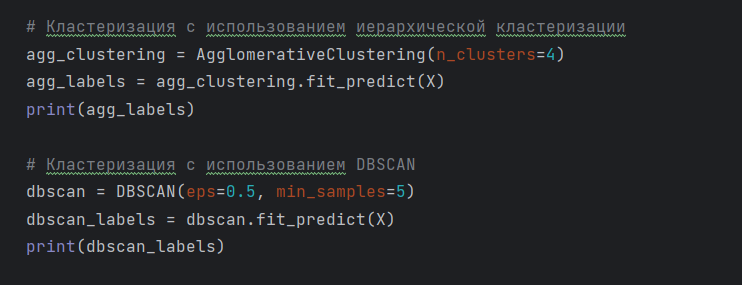




**Рисунок 3 – Результат задания 2**

**Задание 3**

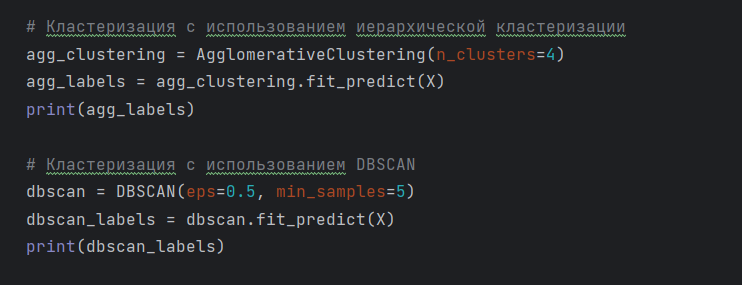
Провести кластеризацию данных с помощью алгоритма иерархической кластеризации.

****

**Рисунок 4 – Задание 3**

**Задание 4**

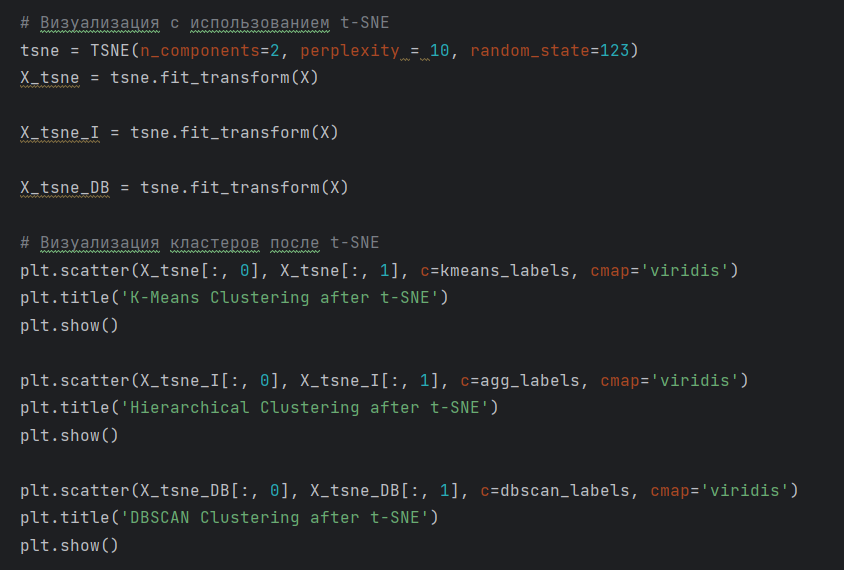
Провести кластеризацию данных с помощью алгоритма DBSCAN.

****

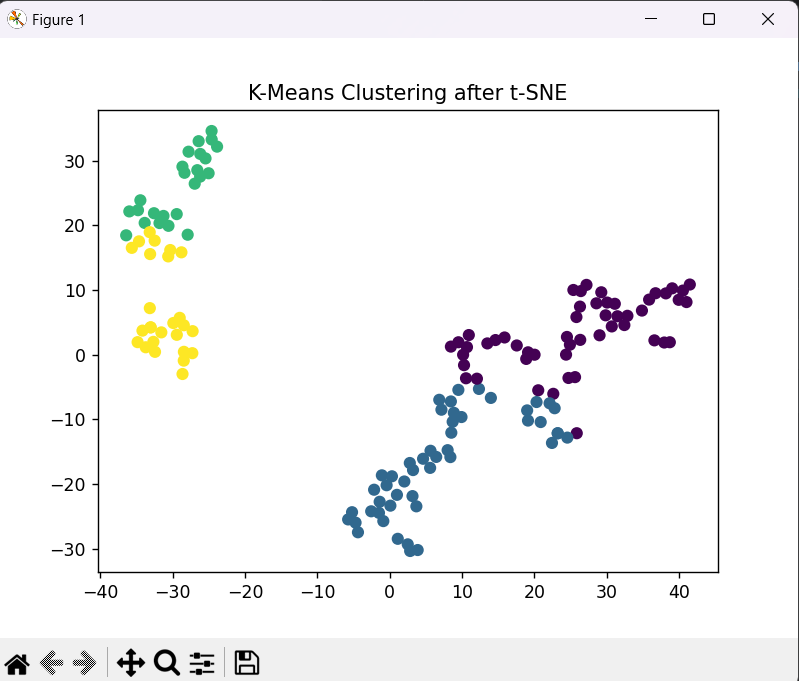
**Рисунок 5 – Задание 4**

**Задание 5**

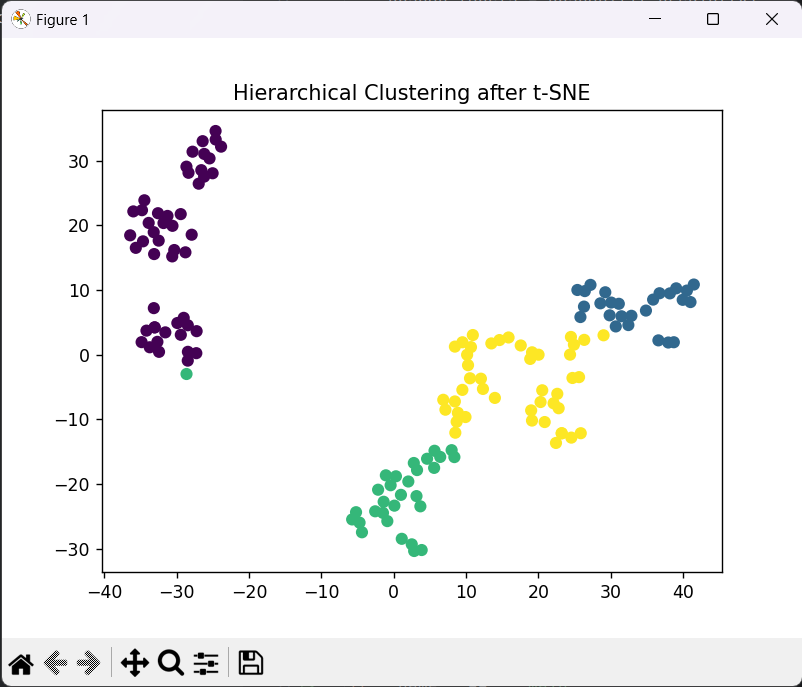
Визуализировать кластеризованные данные с помощью t-SNE или UMAP если данные многомерные. Если данные трехмерные, то можно использовать трехмерный точечный график.

****

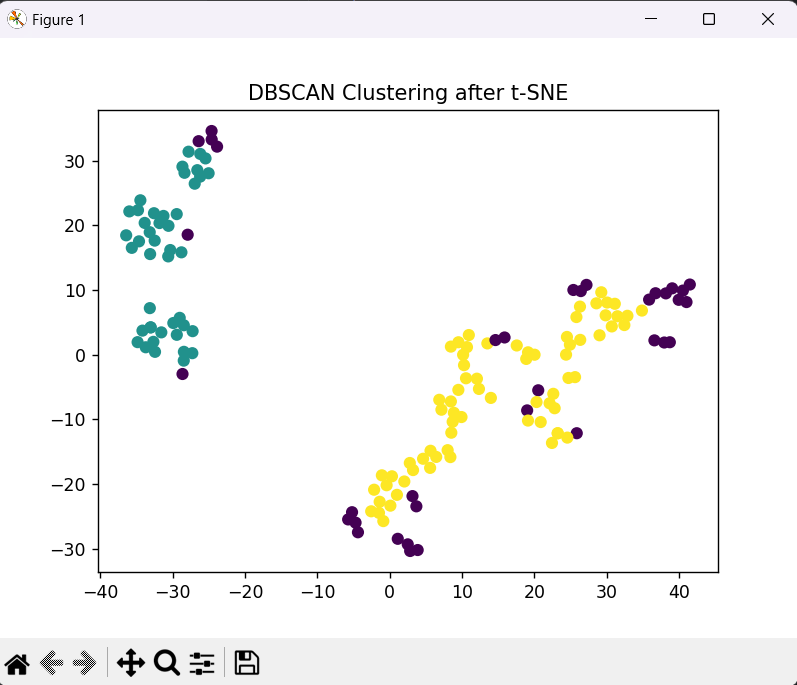
**Рисунок 7 – Задание 6**

****

**Рисунок 8 – Задание 6, k-means**

****

**Рисунок 9 – Задание 6, иерархическая кластеризация**

****

**Рисунок 10 – Задание 6, DBSCAN**

**Вывод**

В ходе выполнения данной практической работы мы познакомились с алгоритмами кластеризации, такими как: k-means (k-средних), иерархический кластерный анализ и алгоритм кластеризации, основанной на плотности (DBSCAN).