|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | |  |  |  | | МИНОБРНАУКИ РОССИИ | | | | Федеральное государственное бюджетное образовательное учреждение  высшего образования  **«МИРЭА** – **Российский технологический университет»**  **РТУ МИРЭА** | | | | |
| Институт информационных технологий (ИТ) | |
| Кафедра прикладной математики | |

|  |  |
| --- | --- |
| **ОТЧЁТ ПО ПРАКТИЧЕСКОЙ РАБОТЕ № 6** | |
| **по дисциплине «Технологии и инструментарий анализа больших данных»** | |
|  | |
| Выполнил студент группы ИКБО-20-21 | Фомичев Р.А. |
| Проверил ассистент кафедры ПМ ИИТ | Тетерин Н.Н. |

**Задание 1**

Найти данные для кластеризации. Данные в группе не должны повторяться. Если признаки в данных имеют очень сильно разные масштабы, то необходимо данные предварительно нормализовать. Код представлен на рисунке 1.



Рисунок 1 – Код программы

**Задание 2**

Провести кластеризацию данных с помощью алгоритма k-means. Использовать «правило локтя» и коэффициент силуэта для поиска оптимального количества кластеров. Код представлен на рисунке 2.

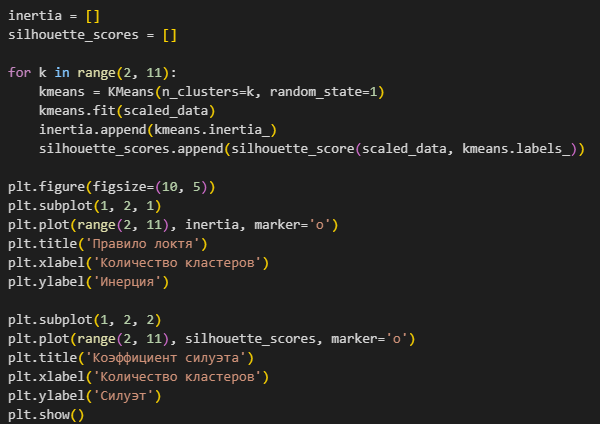


Рисунок 2 – Код программы

Результат работы программы представлен на рисунке 3.

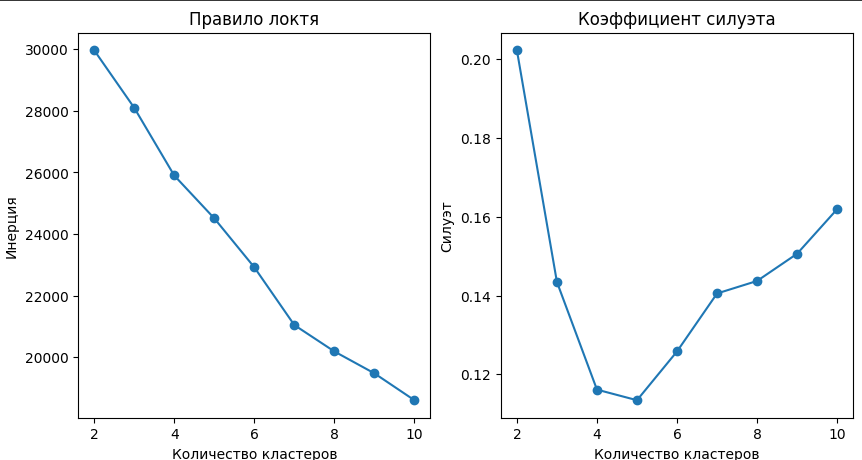


Рисунок 3 – Результат работы программы

Код представлен на рисунке 4.

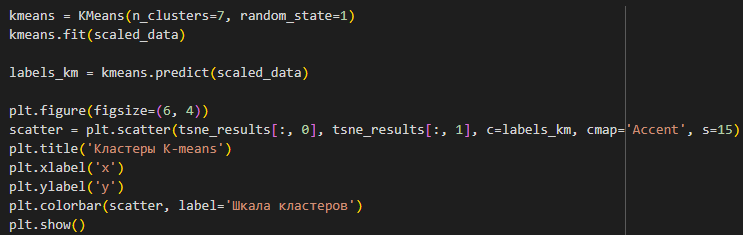


Рисунок 4 – Код программы

Результат работы представлен на рисунке 5.

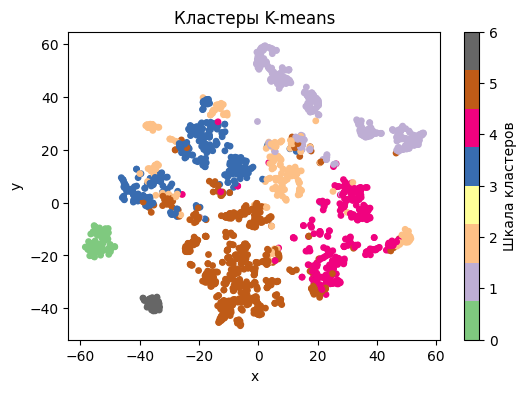


Рисунок 5 – Результат работы программы

**Задание 3**

Провести кластеризацию данных с помощью алгоритма иерархической кластеризации. Код представлен на рисунке 6.

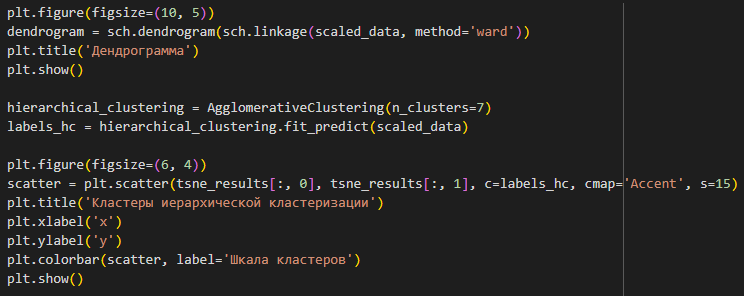


Рисунок 6 – Код программы

Результат работы программы представлен на рисунке 7.

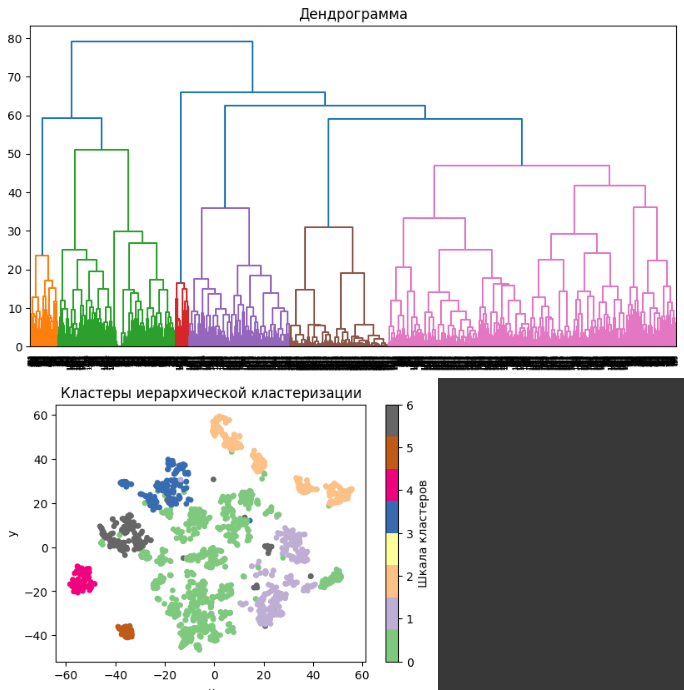


Рисунок 7 – Результат работы программы

**Задание 4**

Провести кластеризацию данных с помощью алгоритма DBSCAN. Код программы представлен на рисунке 8.

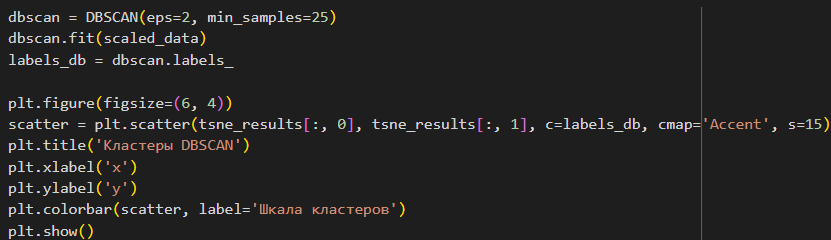


Рисунок 8 – Код программы

Результат работы программы представлен на рисунке 9.

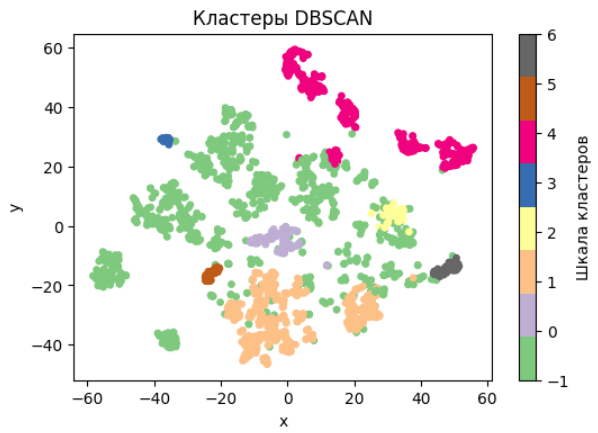


Рисунок 9 – Результат работы программы

**Задание 5**

Визуализировать кластеризованные данные с помощью t-SNE или UMAP, если необходимо. Если данные трехмерные, то можно использовать трехмерный точечный график. Код представлен на рисунке 10.

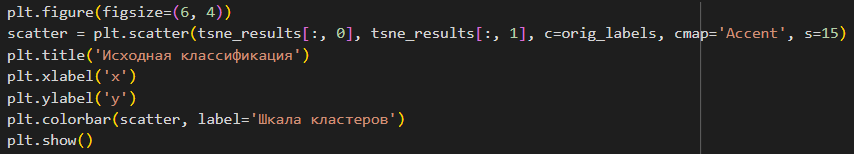


Рисунок 10 – Код программы

Результат работы программы представлен на рисунке 11.

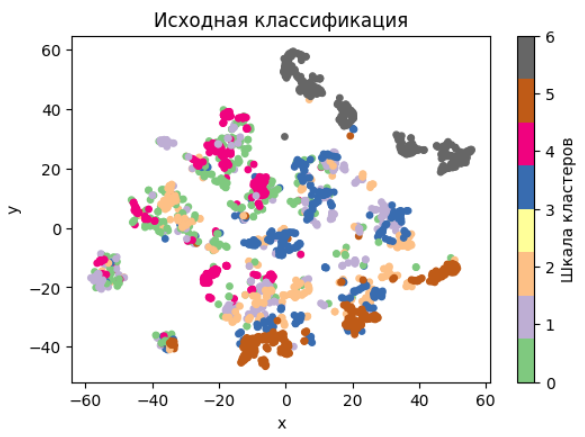


Рисунок 11 – Результат работы программы