|  |
| --- |
|  |
| МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ |
| Федеральное государственное бюджетное образовательное учреждение высшего образования  **"МИРЭА - Российский технологический университет"**  **РТУ МИРЭА** |
| Институт информационных технологий (ИТ) |
| Кафедра прикладной математики (ПМ) |

|  |  |
| --- | --- |
| **ОТЧЕТ**  **ПО ПРАКТИЧЕСКОЙ РАБОТЕ №10** | |
| **по дисциплине** |  |
| **«Технологии и инструментарий анализа больших данных»** | |
| Выполнил студент группы ИКБО-20-19 | Московка А.А. |
| Принял семинарист кафедры Прикладной Математики | Десятников А.А. |

|  |  |  |
| --- | --- | --- |
| Практическая работы выполнены | « » 2022г. |  |
| «Зачтено» | « » 2022 г. |  |

Москва 2022

**Постановка задачи**

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. Сравнить скорость работы алгоритмов. Результаты изобразить в виде таблицы.
6. Визуализировать кластеризованные данные с помощью t-SNE или UMAP если данные многомерные. Если данные трехмерные, то можно использовать трехмерный точечный график.
7. Оформить отчет о проделанной работе. Сделать выводы.

**Программный код с комментариями и результаты его выполнения**

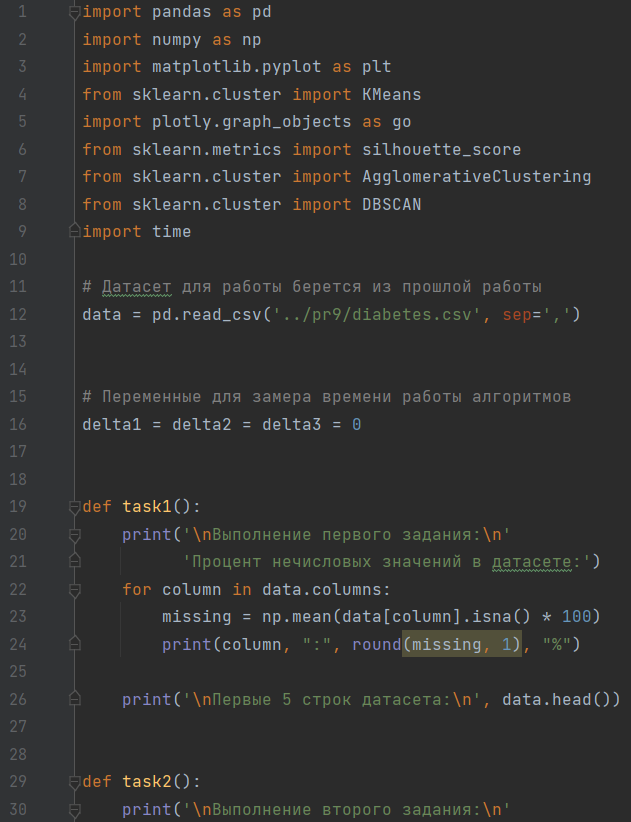


Рисунок 1 – Скриншот программного кода (Часть 1)



Рисунок 2 – Скриншот программного кода (Часть 2)

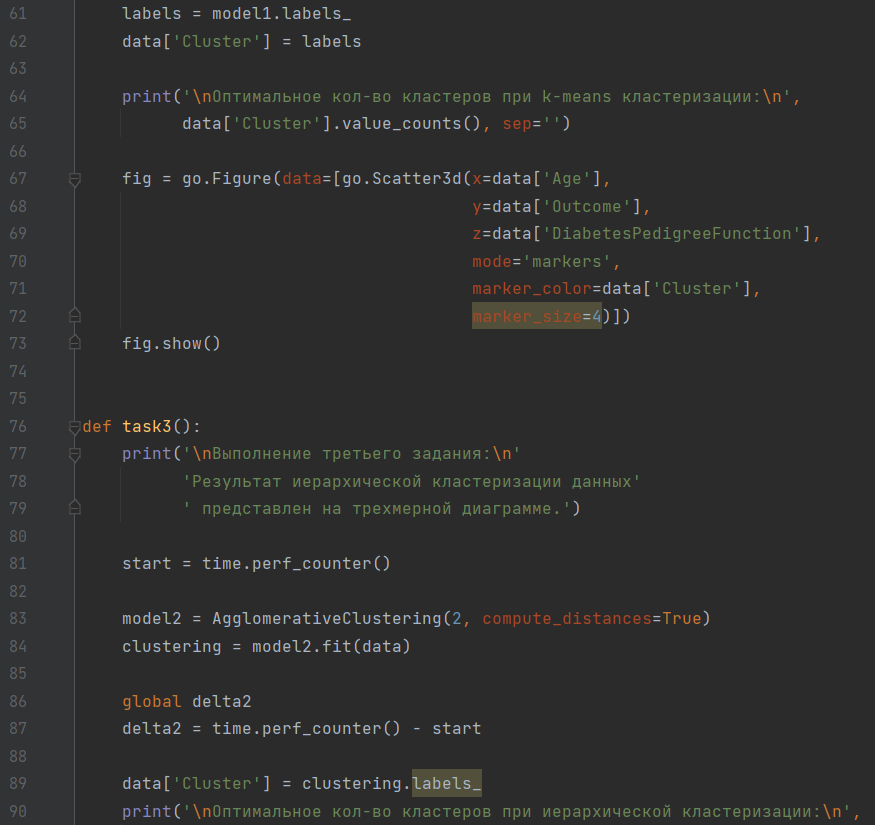


Рисунок 3 – Скриншот программного кода (Часть 3)

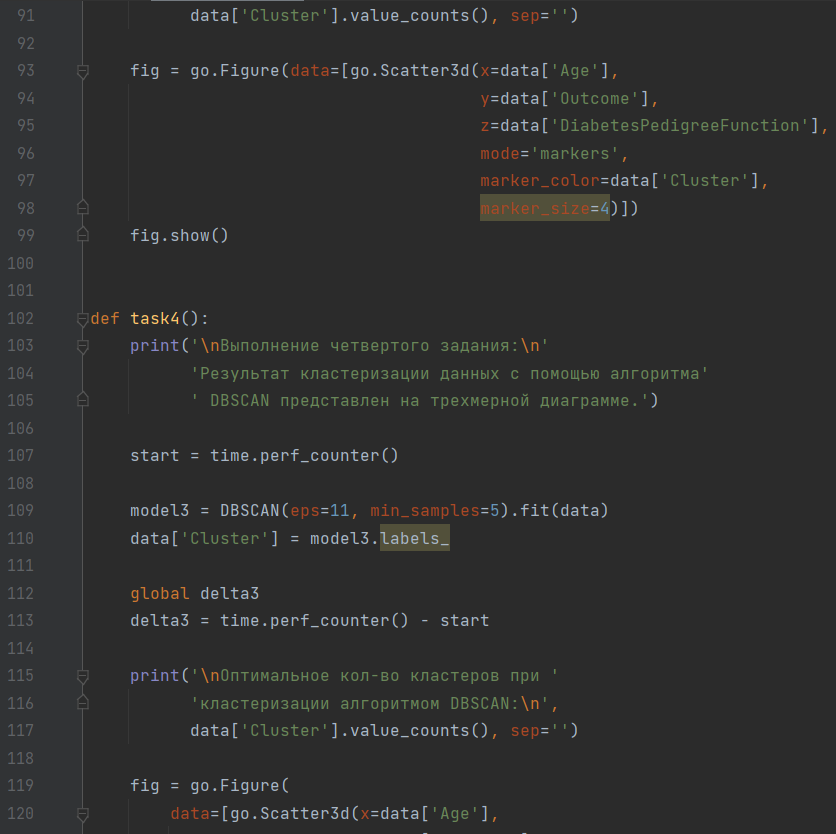


Рисунок 5 – Скриншот программного кода (Часть 4)

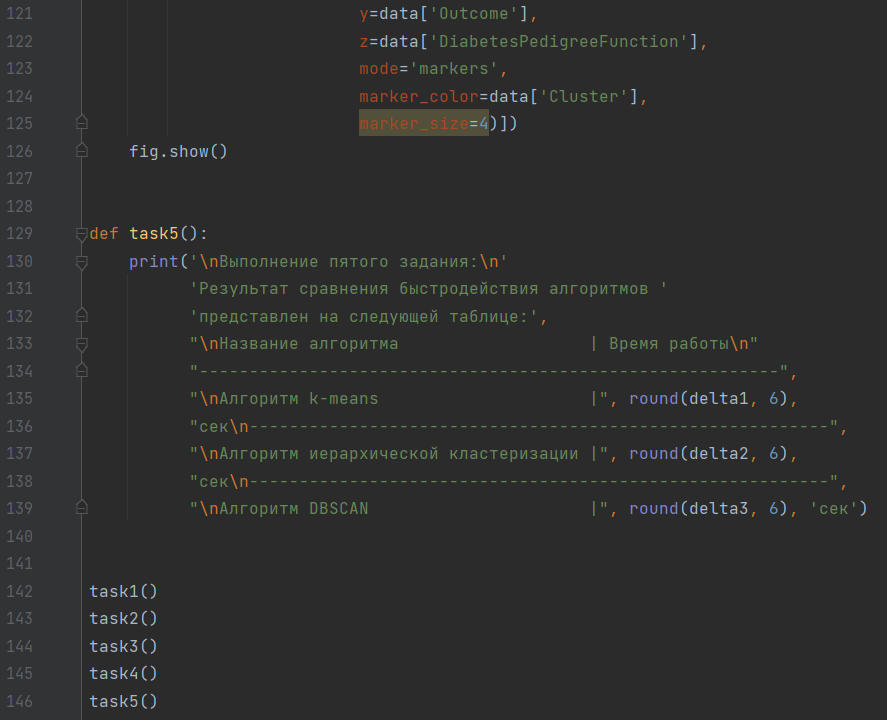


Рисунок 6 – Скриншот программного кода (Часть 5)

На следующих рисунках представлены скриншоты результатов выполнения программы со всеми заданиями с пояснениями, а также диаграммы (Рисунок 7-14).

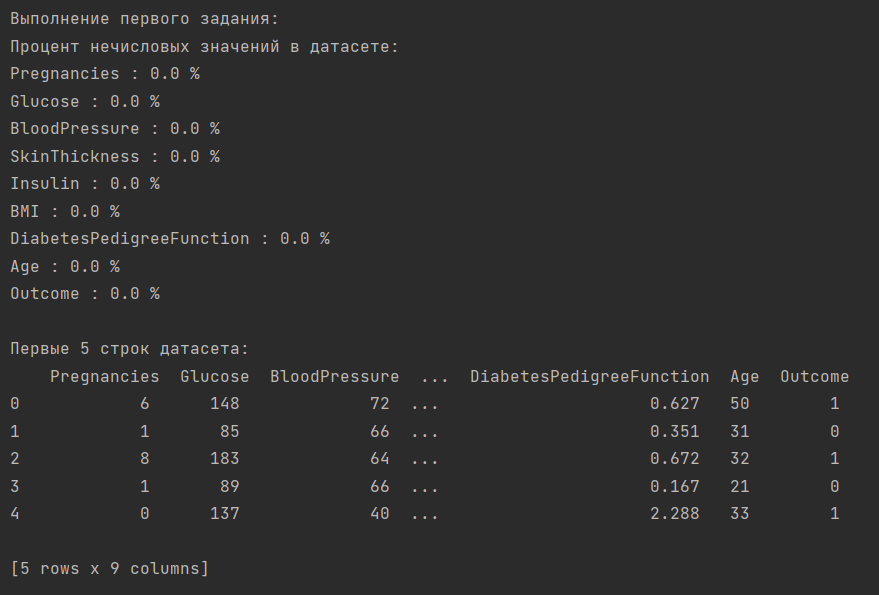


Рисунок 7 – Скриншот результата выполнения первого заданий

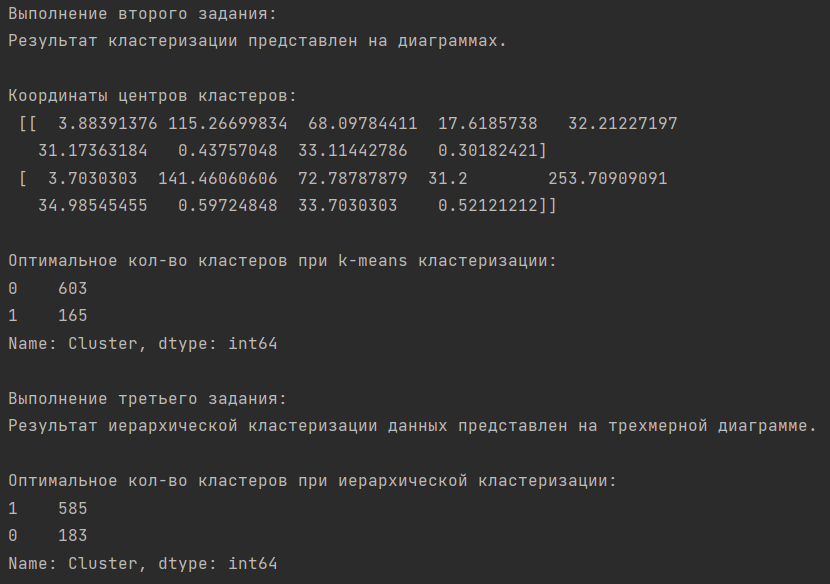


Рисунок 8 – Скриншот результата выполнения второго и третьего заданий

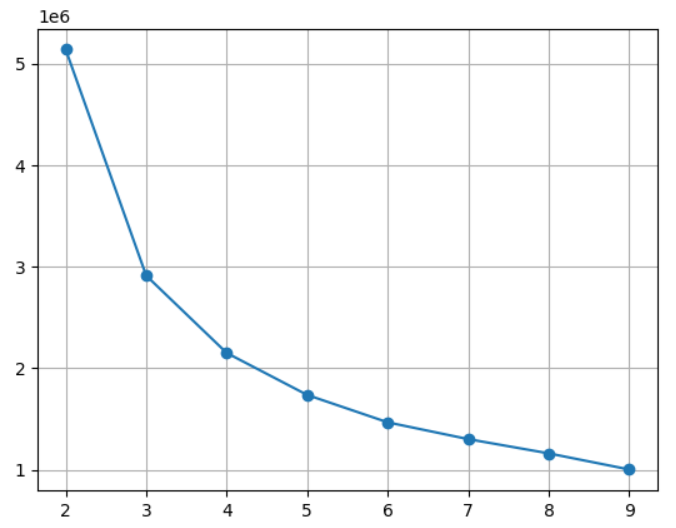


Рисунок 9 – Скриншот результата выполнения второго задания (диаграмма 1)

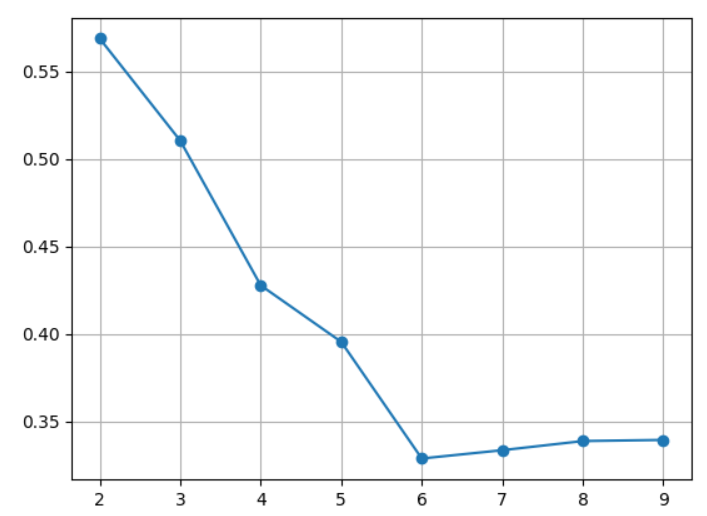


Рисунок 10 – Скриншот результата выполнения второго задания (диаграмма 2)

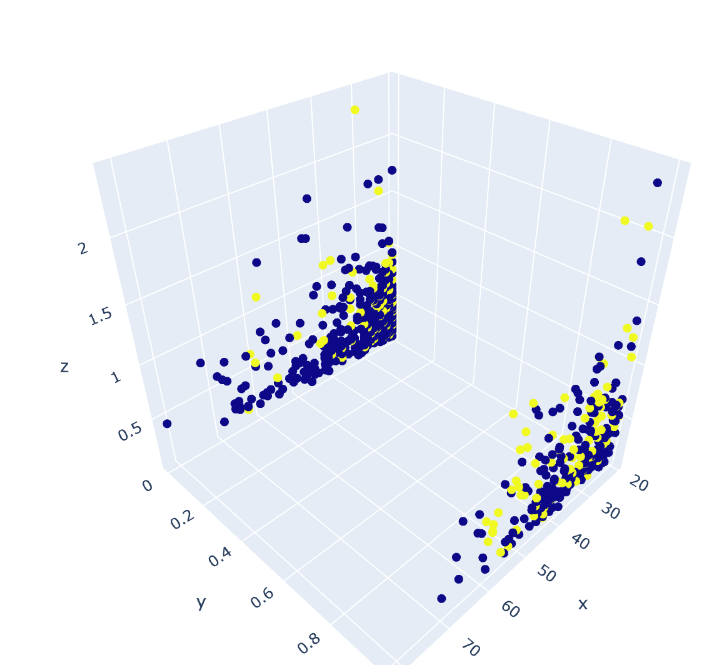


Рисунок 11 – Скриншот результата выполнения второго задания (трехмерная диаграмма)

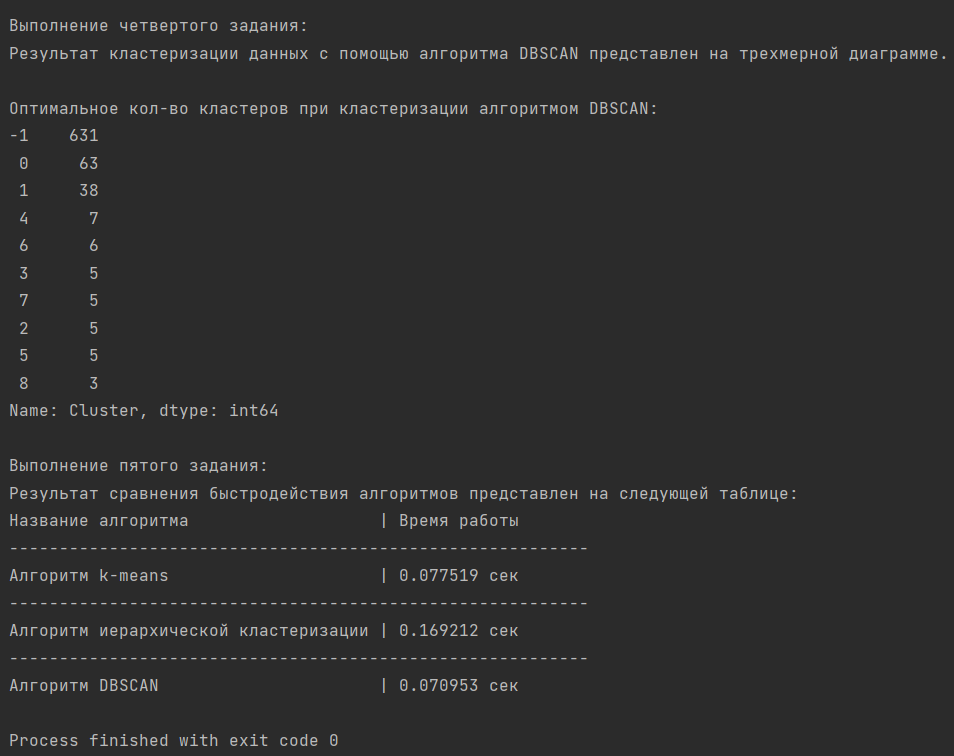


Рисунок 12 – Скриншот результата выполнения четвертого и пятого заданий

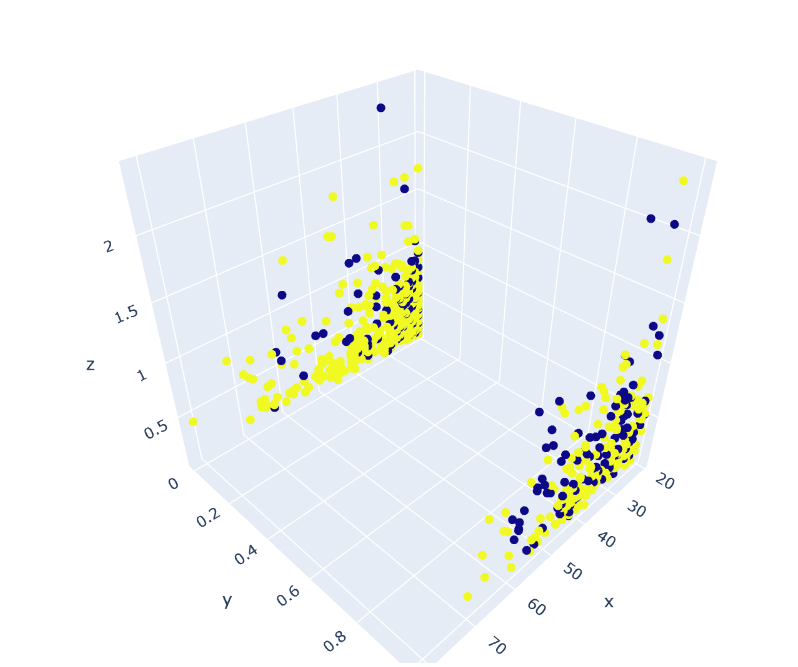


Рисунок 13 – Скриншот результата выполнения третьего задания (трехмерная диаграмма)

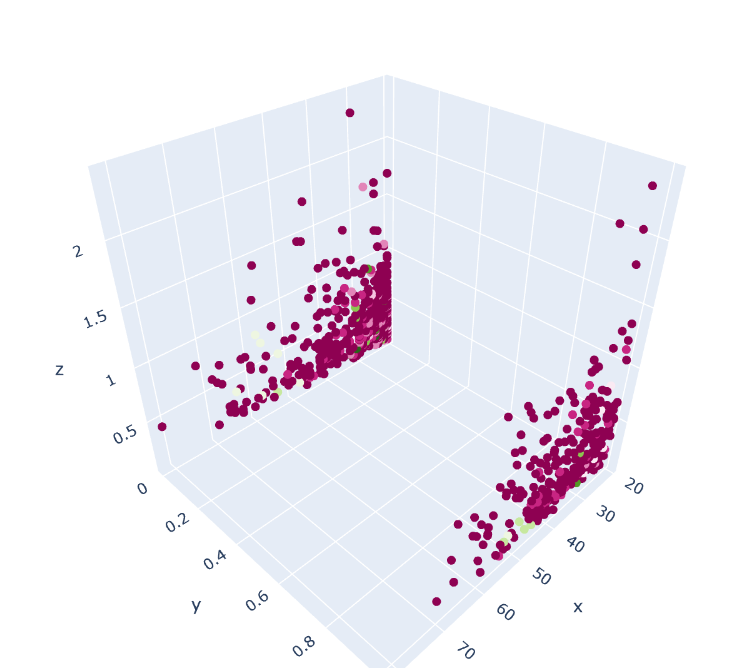


Рисунок 14 – Скриншот результата выполнения четвертого задания (трехмерная диаграмма)

**Вывод**

В результате выполнения данной практической работы были найдены данные для классификации (данные по диабету), проведена кластеризация данных с помощью алгоритма k-means, использовано «правило локтя» и коэффициент силуэта для поиска оптимального количества кластеров, проведена кластеризация данных с помощью алгоритма иерархической кластеризации, проведена кластеризация данных с помощью алгоритма DBSCAN, сравнена скорость работы алгоритмов, что было представлено в виде таблицы, визуализированы кластеризованные данные, наконец, оформлен отчет о проделанной работе.