|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | |  |  |  | | 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 № 7** | |
| **по дисциплине «Технологии и инструментарий анализа больших данных»** | |
|  | |
| Выполнил студент группы ИКБО-20-21 | Мухаметшин А. Р. |
| Проверил ассистент кафедры ПМ ИИТ | Тетерин Н.Н. |

**Задание 1**

Найти данные для задачи классификации или для задачи регрессии (данные не должны повторятся в группе).

Для работы будет использоваться Heart Disease Dataset (Классификация). Набор данных содержит информацию о пациентах с возможными сердечными заболеваниями. Признаки включают возраст, пол, артериальное давление и другие медицинские показатели. Цель — предсказать наличие сердечных заболеваний.

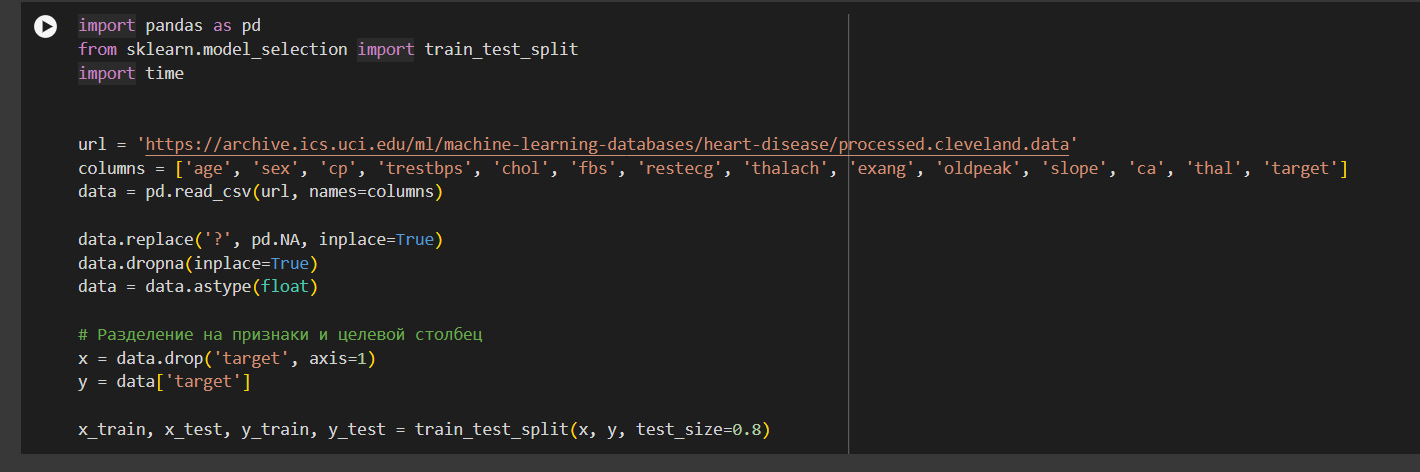


Рисунок 1 – выполнение задания 1

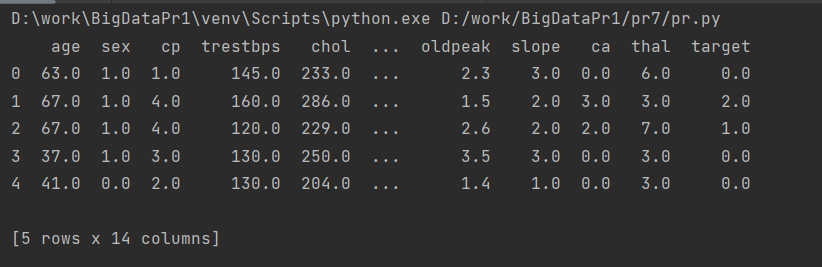


Рисунок 2 – Пример вывода

**Задание 2**

Реализовать баггинг.

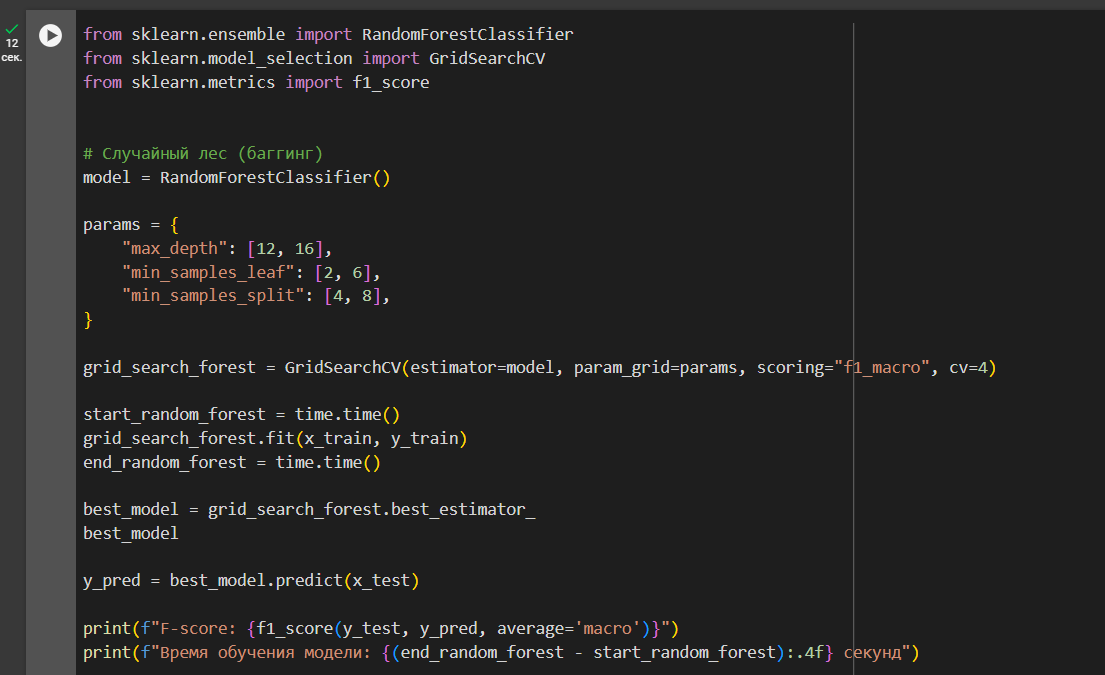


Рисунок 3 – выполнение задания 1

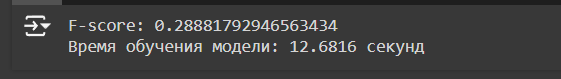
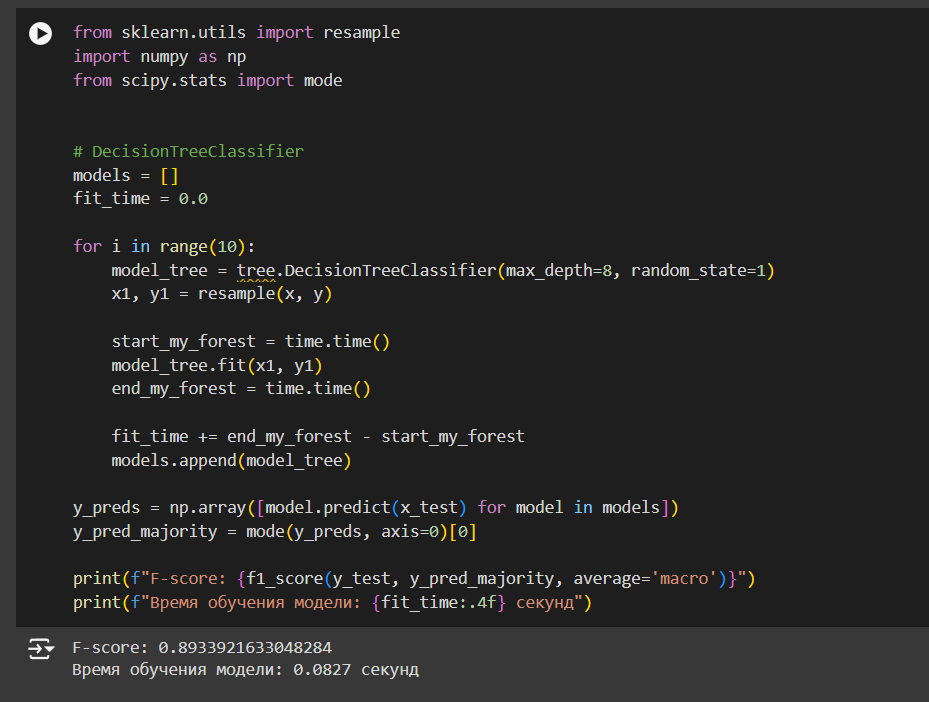


Рисунок 4 – Пример работы

Рисунок 5 – DecisionTreeClassifier



**Задание 3**

Реализовать бустинг на тех же данных, что использовались для баггинга.

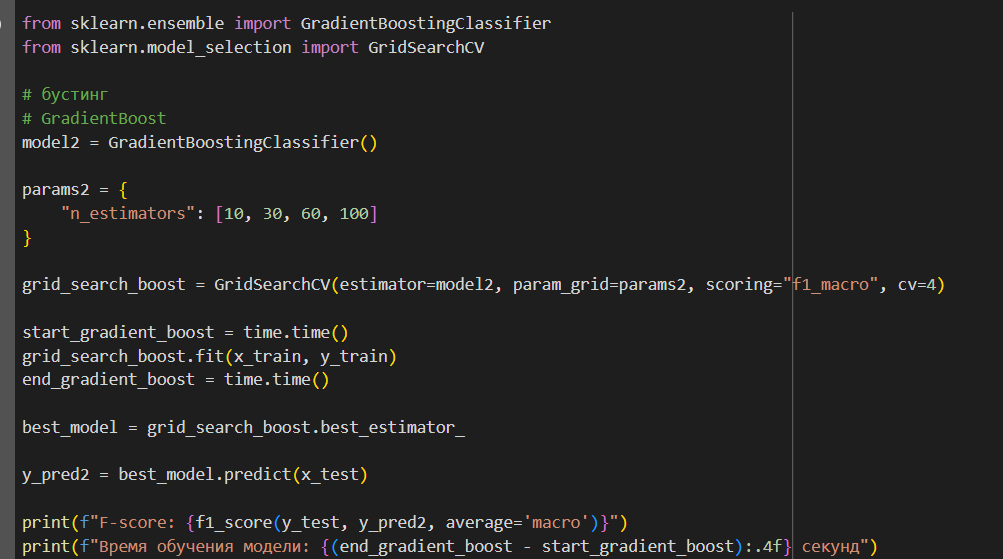


Рисунок 6 – Код к заданию 3

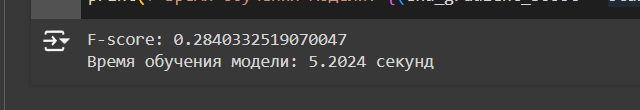


Рисунок 7 – Пример работы

**Задание 4**

Сравнить результаты работы алгоритмов (время работы и качество моделей). Сделать выводы.

Самым точным методом обучения на данных по классификации сердечных заболеваний оказался баггинг, реализованный на основе DecisionTreeClassifier.