|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | |  |  |  | | 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 | Фомичев Р.А. |
| 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). Код представлен на рисунке 1.

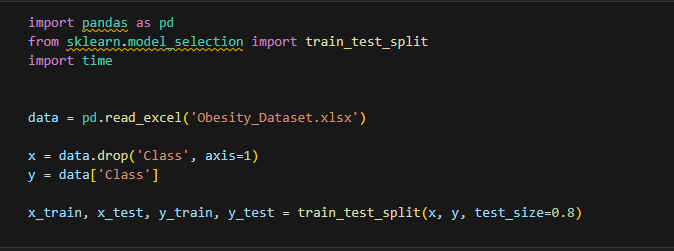


Рисунок 1 – Код программы

Задание 2

Реализовать баггинг. Код программы с выводом представлены на рисунках 2, 3.

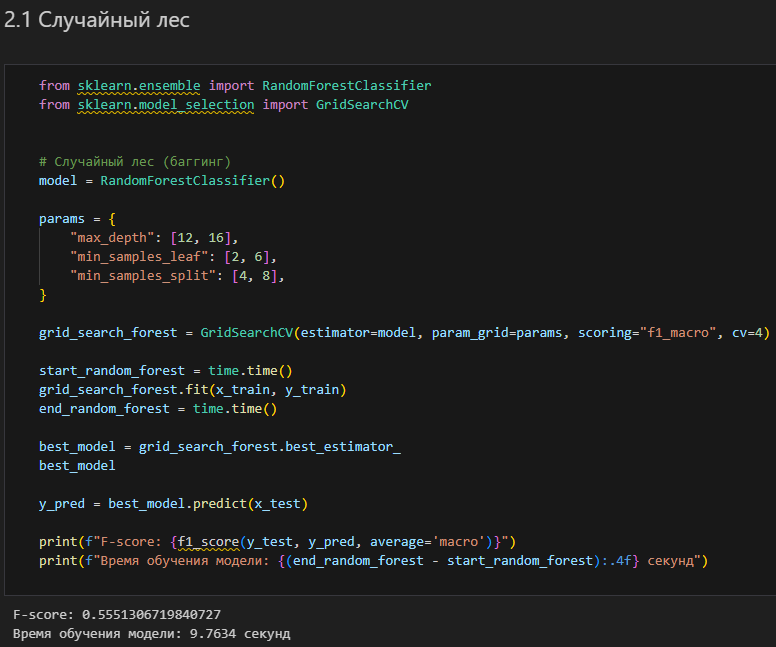


Рисунок 2 – Код программы с выводом



Рисунок 3 – Код программы с выводом

**Задание 3**

Реализовать бустинг на тех же данных, что использовались для баггинга. Код программы с выводом представлены на рисунках 4, 5.

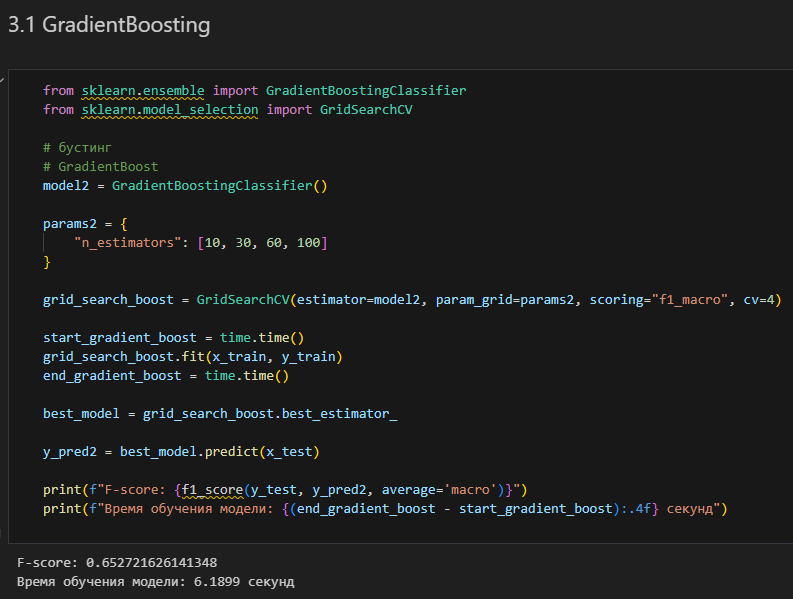


Рисунок 4 – Код программы с выводом

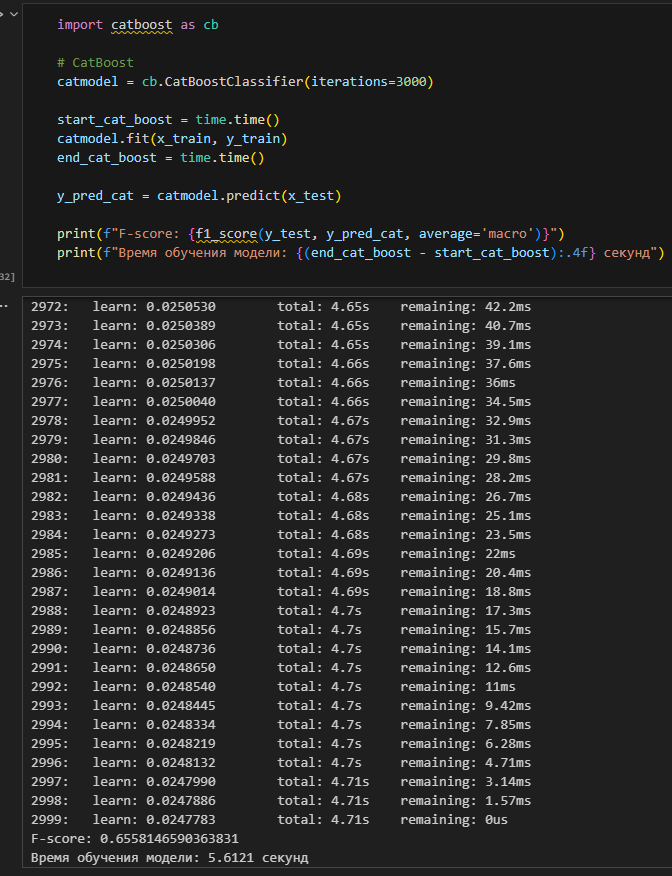


Рисунок 5 – Код программы с выводом

**Задание 4**

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

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

Случайный лес (RandomForestClassifier) оказался наименее точным согласно метрике F1. Также время его работы оказалось наибольшим.

Бустинг находится посередине между представленными выше методами по времени и по метрике F1.