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

|  |  |
| --- | --- |
| **Отчет по практической работе №7.** | |
| **по дисциплине** | |
| **«Технологии и инструментарий анализа больших данных»** | |
|  | |
| Выполнил студент группы ИКБО- 22-20 | Гологузов К.А. |
| Принял | Парамонов А.А. |

Москва 2023

Решение

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

Для решения задачи классификации был выбран датасет из библиотеки sklearn “load\_wine”, содержащий следующие признаки:

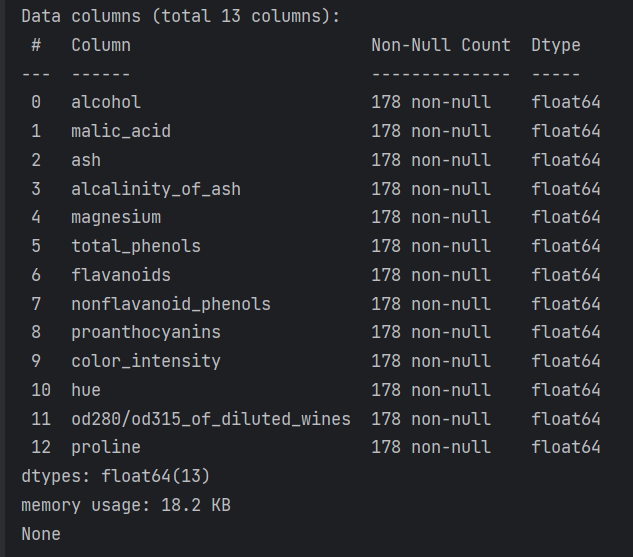


Рис.1 – Данные

2) Реализовать баггинг

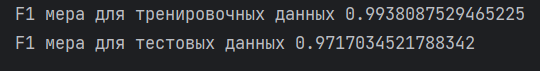


Рис.2 – Результат работы кода

Баггинг был реализован алгоритмом случайного леса.

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

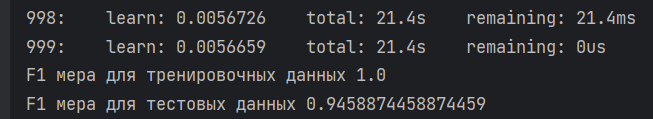


Рис.3 – Результат работы кода

Для реализации был использован бустинг CatBoost.

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

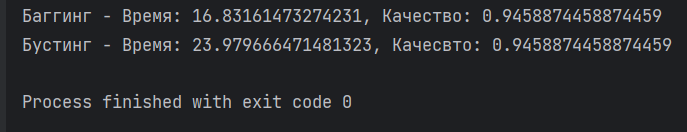


Рис.4 – Сравнение алгоритмов

Алгоритм случайного леса работает быстрее, чем градиентный бустинг. Бустинг обычно точнее, однако для выбранных данных разницы нет.