|  |
| --- |
| МИНОБРНАУКИ РОССИИ |
| Федеральное государственное бюджетное образовательное учреждение высшего образования **«МИРЭА − Российский технологический университет»**  **РТУ МИРЭА** |

**Институт информационных технологий (ИИТ)**

**Кафедра прикладной математики (ПМ)**

**ОТЧЕТ ПО ПРАКТИЧЕСКОЙ РАБОТЕ**

по дисциплине «Технологии и инструментарий анализа больших данных»

**Практическое занятие №7**

|  |  |  |  |
| --- | --- | --- | --- |
| Студент группы ИНБО-01-17 | *ИНБО-07-21, Веригин Никита Алексеевич* | (подпись) | |
| Преподаватель | *Царёв Роман Юрьевич* | (подпись) | |
| Отчет представлен | «\_\_\_»\_\_\_\_\_\_\_\_2024 г. | |  | |

Москва 2024 г.

Введение

Цели работы:

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

2) Реализовать баггинг.

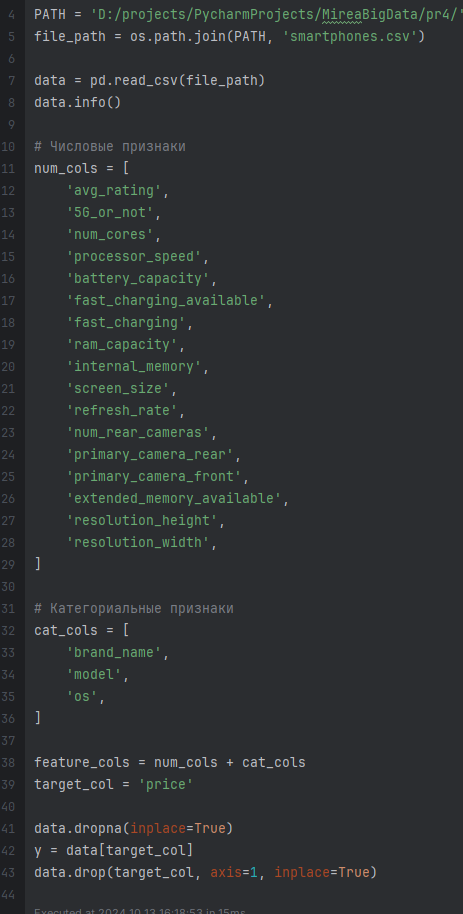
3) Реализовать бустинг на тех же данных, что использовались для баггинга.

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

5) Оформить отчет о проделанной работе.

ХОД РАБОТЫ

Код программ и результаты их выполнения представлены на Рисунках 1-5.



**Рисунок 1 *—* Задача 1**

**Изображение выглядит как текст, снимок экрана, программное обеспечение

Автоматически созданное описание**

**Рисунок 2 *—* Задача 1**

**Изображение выглядит как текст, снимок экрана, программное обеспечение

Автоматически созданное описание**

**Рисунок 3 *—* Задача 2**

**Изображение выглядит как текст, снимок экрана, Шрифт, программное обеспечение

Автоматически созданное описание**

**Рисунок 4 — Задача 2**

**Изображение выглядит как текст, снимок экрана, программное обеспечение

Автоматически созданное описание**

**Рисунок 5 *—* Задача 3**

**Изображение выглядит как текст, снимок экрана, программное обеспечение

Автоматически созданное описание**

**Рисунок 6 *—* Задача 3**

Выводы

Время обучения - Random Forest обучается значительно быстрее, чем Gradient Boosting. Это ожидаемо, так как бустинг исполняет последовательные обучающие этапы, тогда как баггинг сделан для обучения деревьев параллельно.

Качество - Random Forest демонстрирует лучшее качество предсказаний на тестовом наборе данных по обеим метрикам (MAE и RMSE) в сравнении с Gradient Boosting. Разница в качестве может быть связана с параметрами, количеством деревьев, а также шагом learning rate, которые существенно влияют на результаты.

ЗАКЛЮЧЕНИЕ

В ходе выполнения практической работы были изучены различные библиотеки ансамблевого обучения особенности работы с ними в среде программирования Python.