

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 прикладная математика (ПМ)

**ОТЧЕТ ПО ПРАКТИЧЕСКОЙ РАБОТЕ №7**

**по дисциплине**

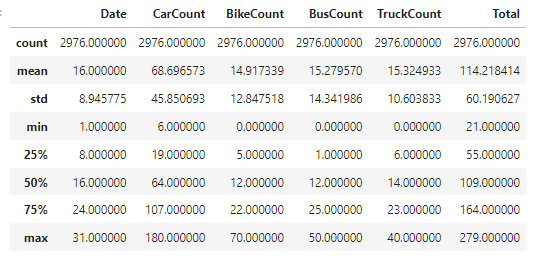
**«Технологии и инструментарий анализа больших данных»**

|  |  |  |  |
| --- | --- | --- | --- |
| Выполнил студент группы ИКБО-02-20 | |  | Злобин А.О. |
| Принял преподаватель |  |  | Горячев А.А. |
|  |  |  | |

Москва 2023

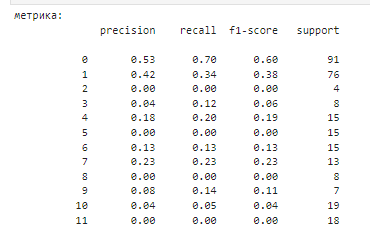
1. Найти данные для классификации. Данные в группе повторяться не должны. Предобработать данные, если это необходимо.

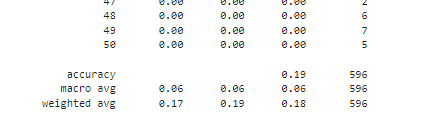


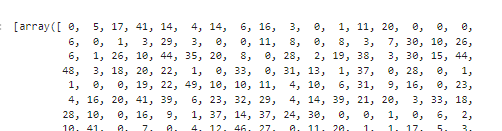


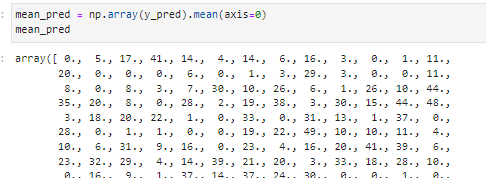
1. Реализовать баггинг.

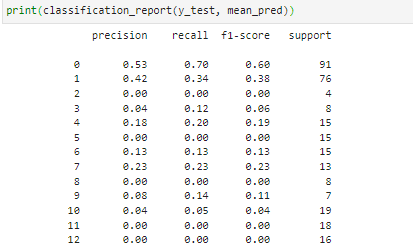


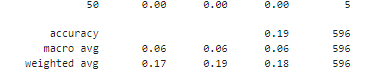




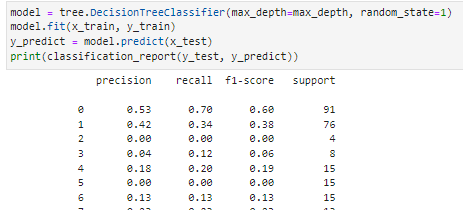


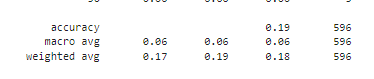




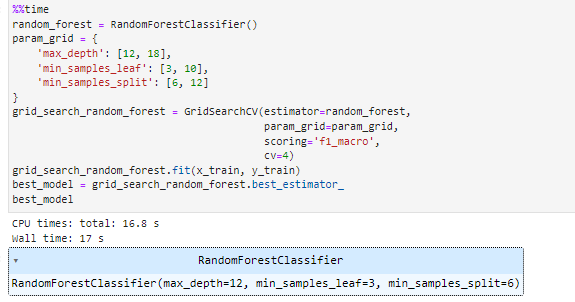


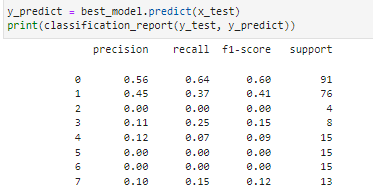
Полная выборка:

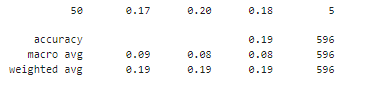




Вывод: Не выявлено никакой разницы  
Создадим случайный лес:

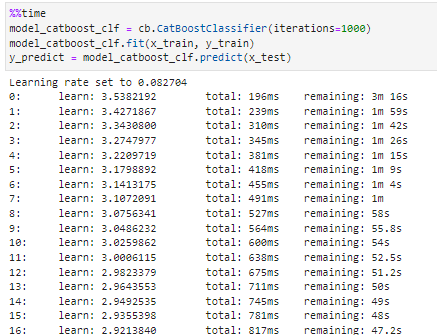


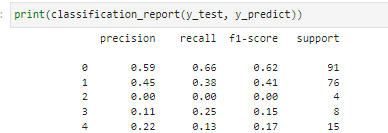


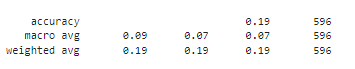


Вывод: Точность никак не изменилась

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







Результаты как у обычного дерева

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

Даже с учетом GridSearchCV, лес работает гораздо медленнее чем CatBoosting. Результаты при этом остались одинаковые.