

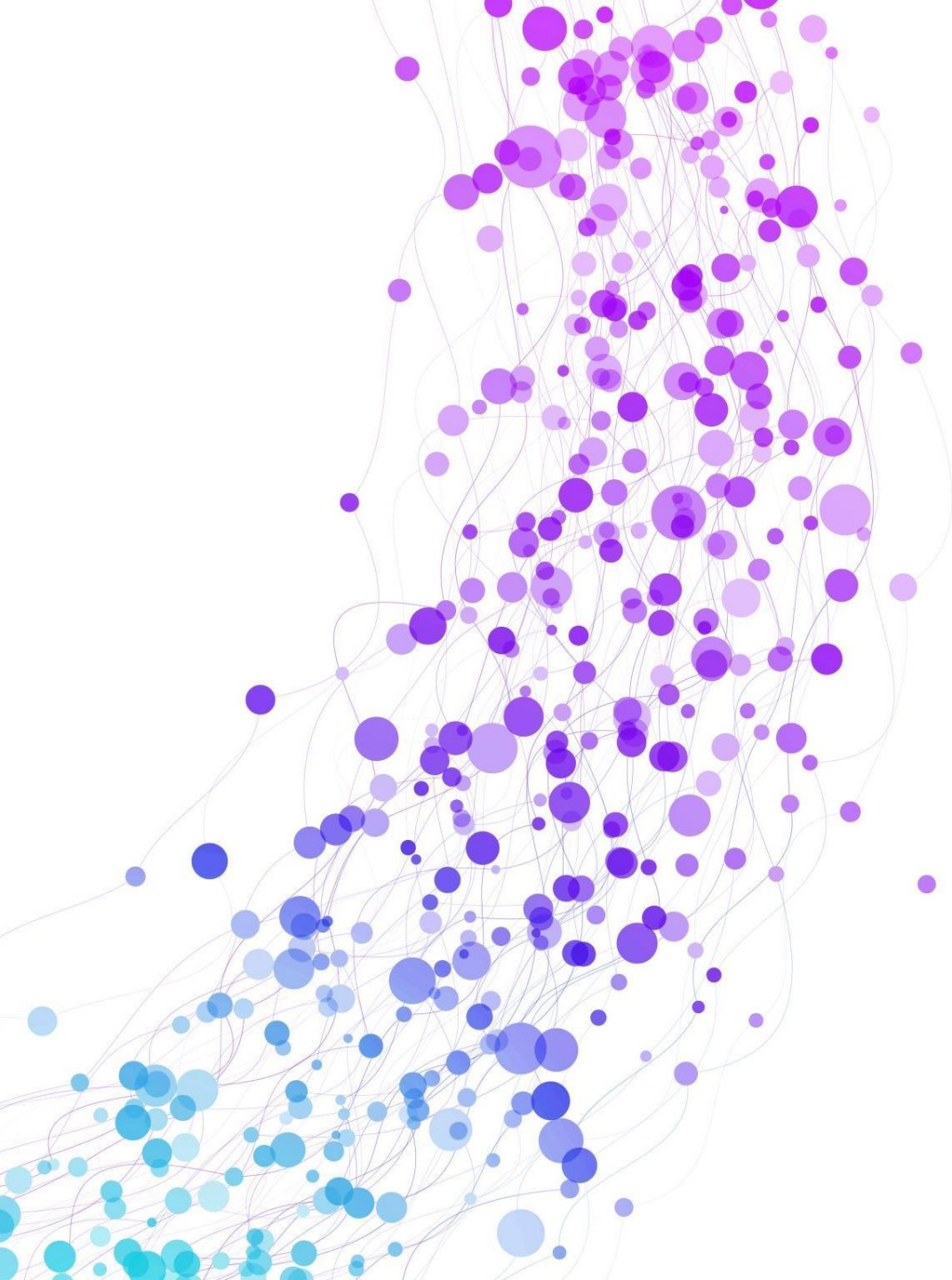
Классификация типов воды по качеству



Постановка задачи

Одна из важных задач мониторинга экосистемы - это состояние водных ресурсов, и в частности подготовка [очистка] питьевой воды.

Целью данного исследования является создание математической модели, которая позволит сократить затраты водоподготовительных организаций для контроля качества воды



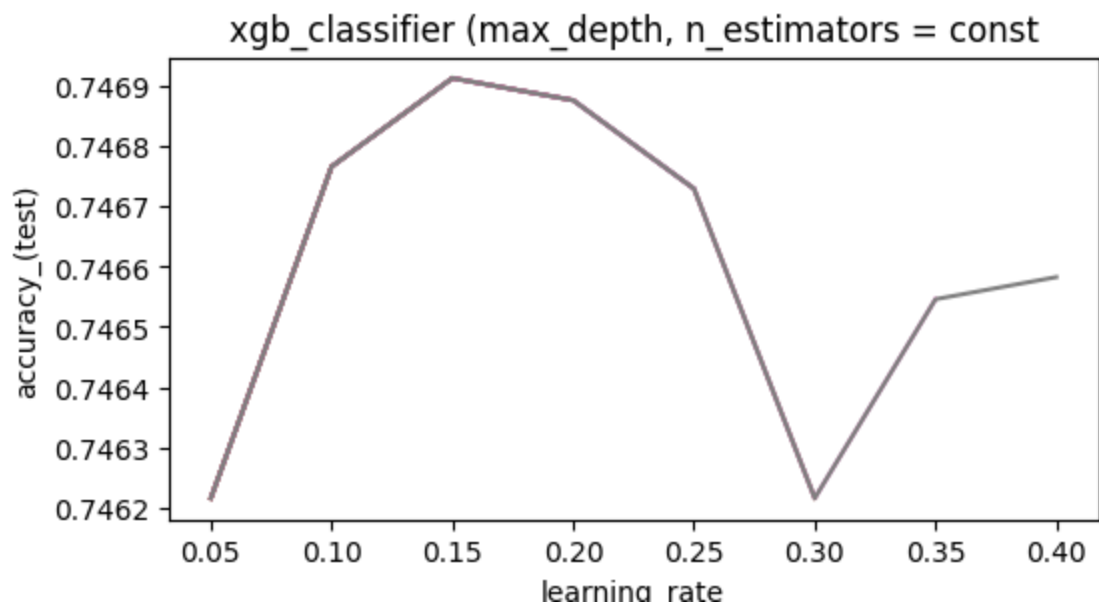
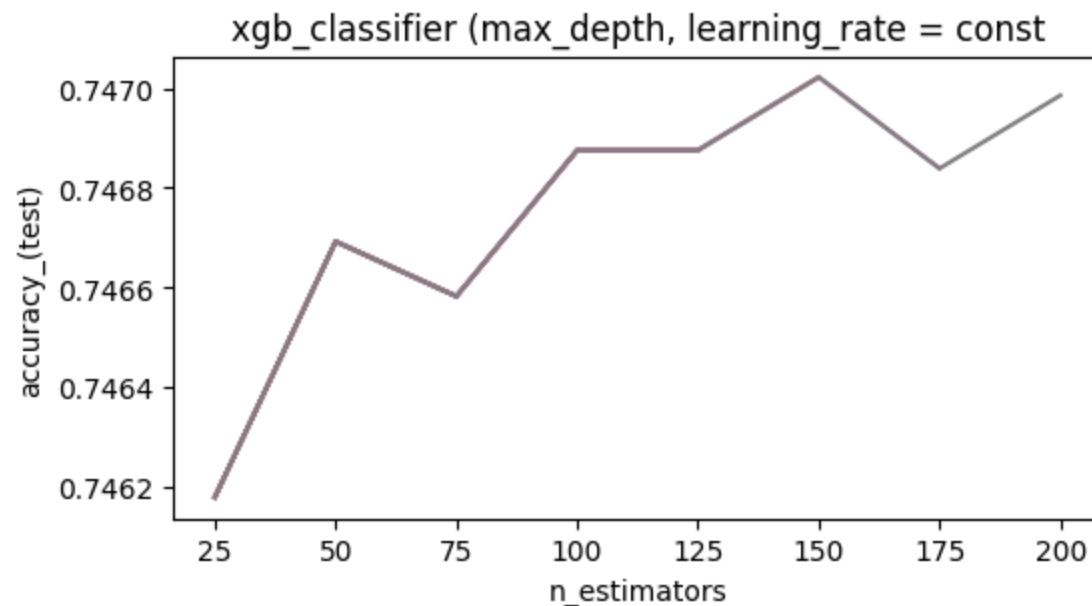
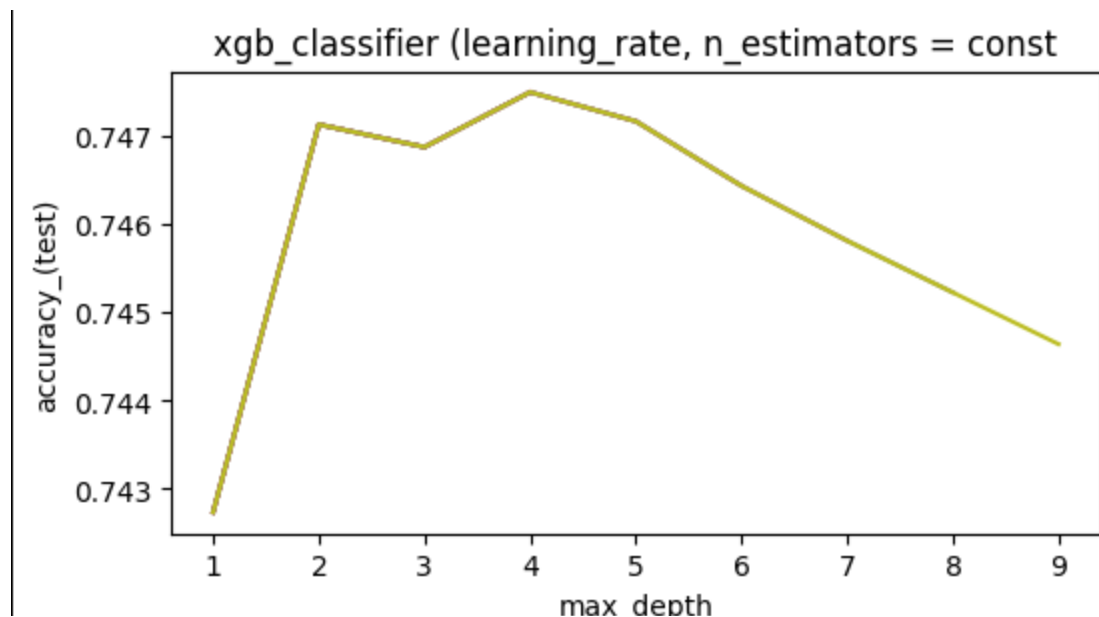
Использованные данные

Исследование проведено на данных из открытого источника:

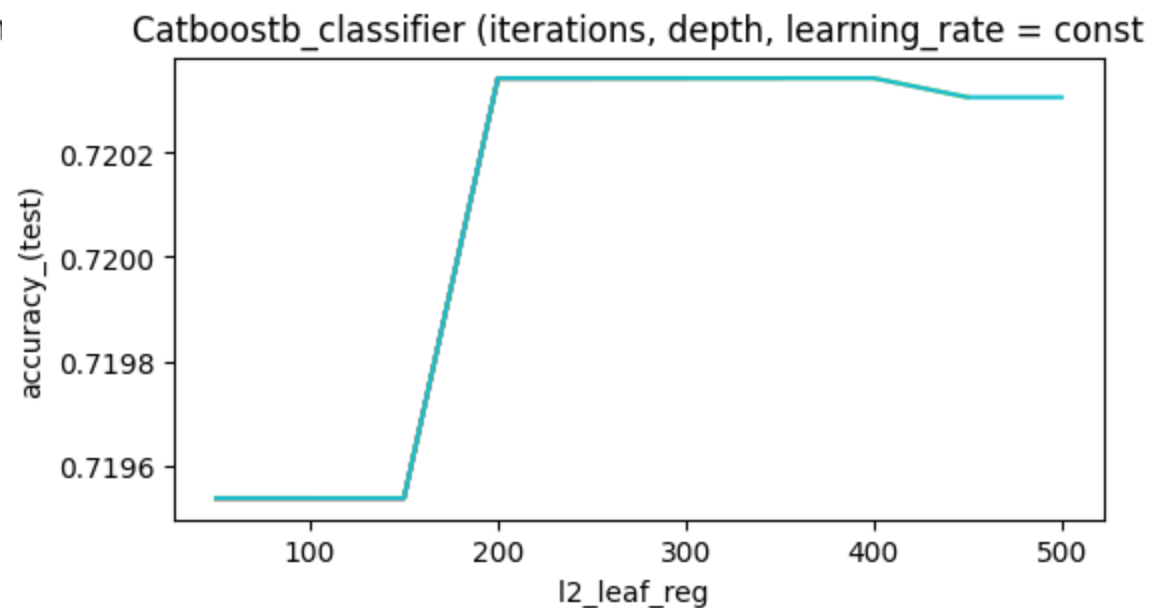
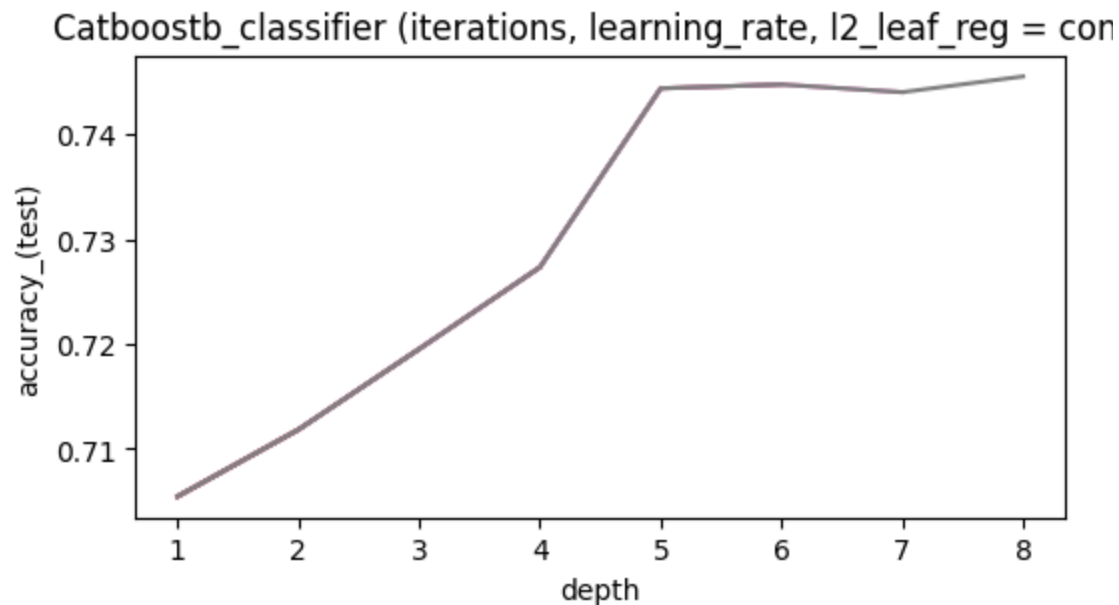
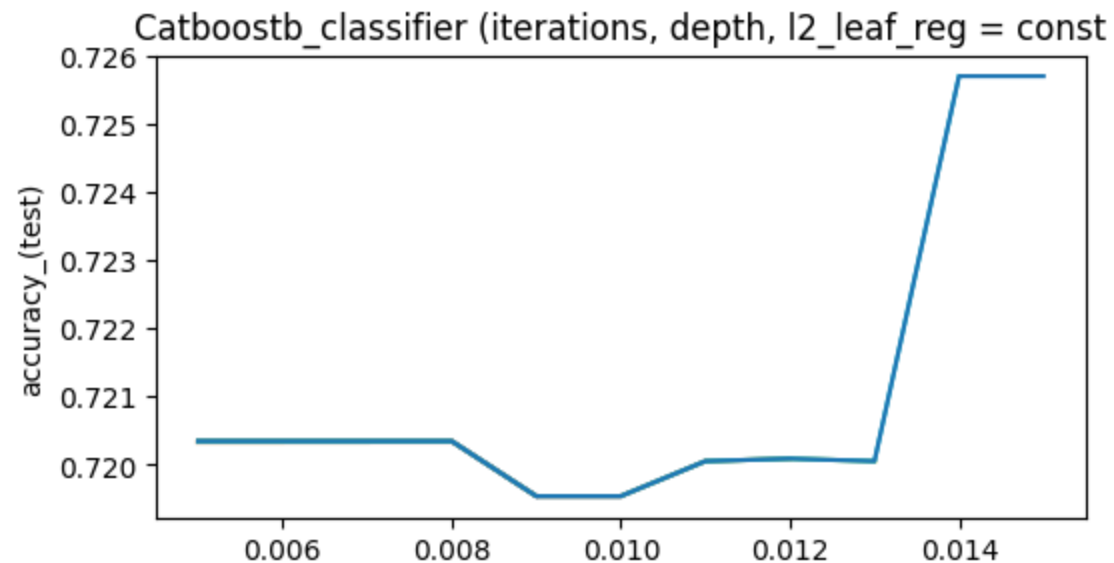
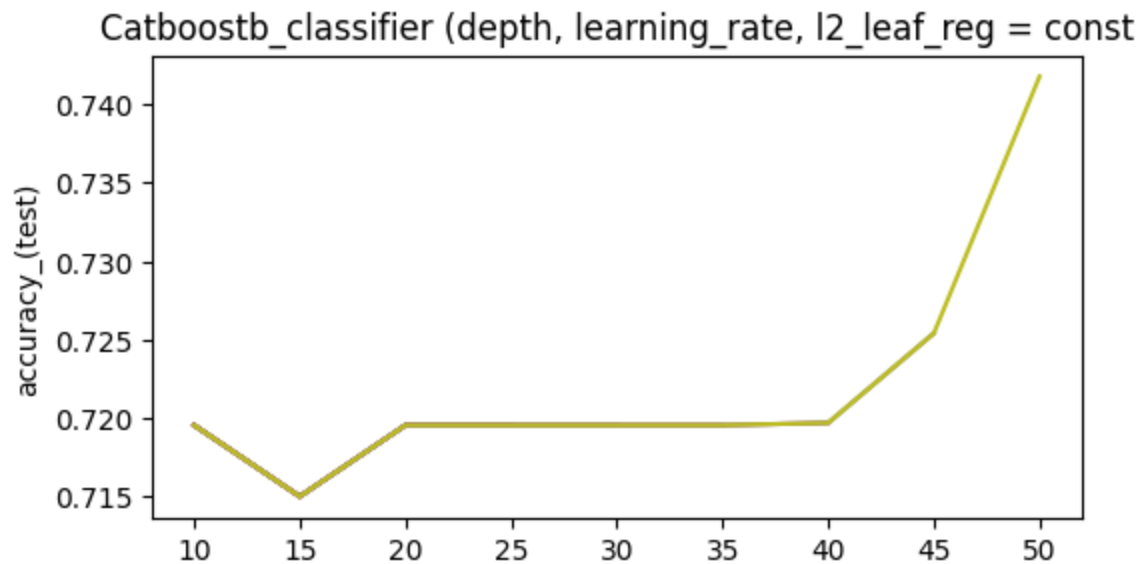
**Drinking Water Quality Distribution
Monitoring Data | NYC Open Data
(cityofnewyork.us)**

- Датасет содержит **136** тысяч строк (наблюдений) качества воды в Нью-Йорке с **2015** по **2023** год .

Подбор гиперпараметров модели **XGBclassifier**



Подбор гиперпараметров модели **CatBoost**





Выбор модели

В исследовании были обучены 4 модели, выбор был сделан в пользу **XGBoost**

Подобранные гиперпараметры:

`learning_rate = 0.2;`

`max_depth = 3;`

`n_estimators = 100;`

Точность на обучающем наборе:
0.745

Точность на тестовом наборе:
0.747



Заключение и перспективы

В настоящее время контроль качества воды ведется лабораторно по **20** показателям.

Предложенная модель позволяет осуществлять мониторинг качества воды по **5** показателям с точностью до **75%**. Для повышения точности в перспективе могут быть проведены аналогичные исследования при условии раскрытия данных водоподготовительных организаций.