**МИНИСТЕРСТВО ЦИФРОВОГО РАЗВИТИЯ, СВЯЗИ И МАССОВЫХ**

**КОММУНИКАЦИЙ РОССИЙСКОЙ ФЕДЕРАЦИИ**

Федеральное государственное бюджетное образовательное учреждение

высшего образования

**«Поволжский государственный университет телекоммуникаций и информатики»**

Факультет ***Информационных систем и технологий***

Направление подготовки ***09.04.02 Информационные системы и технологии***

**Отчет по лабораторной работе №6**

**По дисциплине “Системы и методы искусственного интеллекта”**

**студента 1 курса группы**

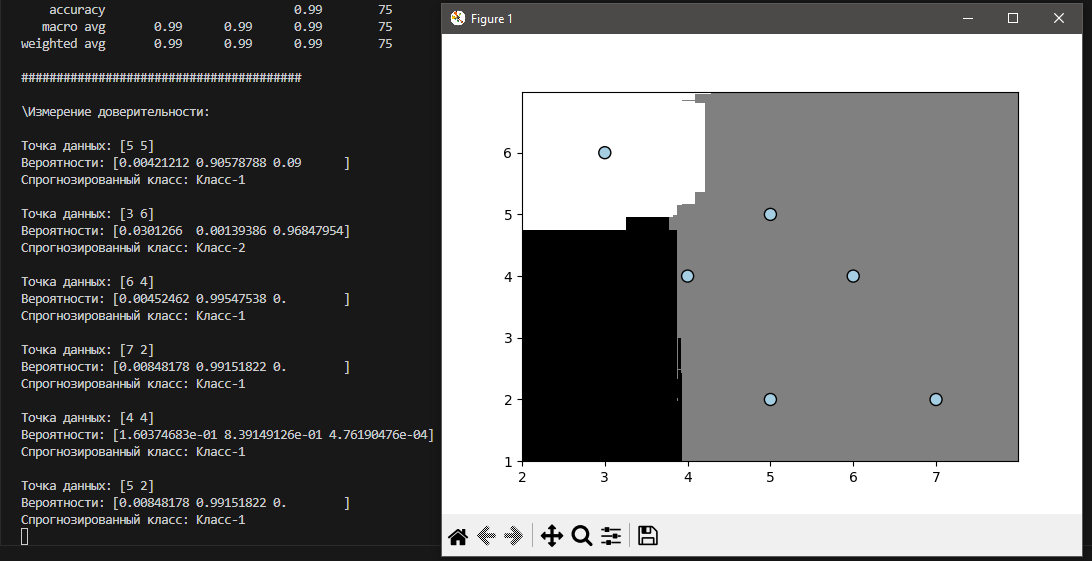
**\_\_**ИСТм-43**\_\_\_**

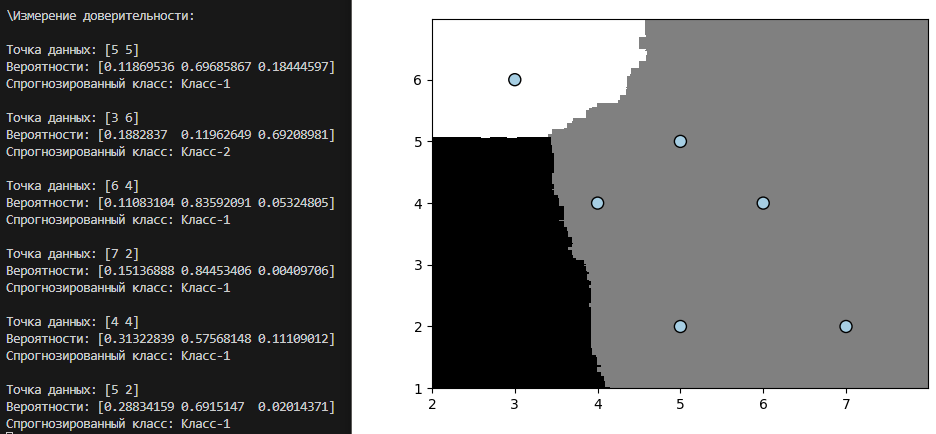
Емелин Максим Сергеевич

(фамилия, имя, отчество)

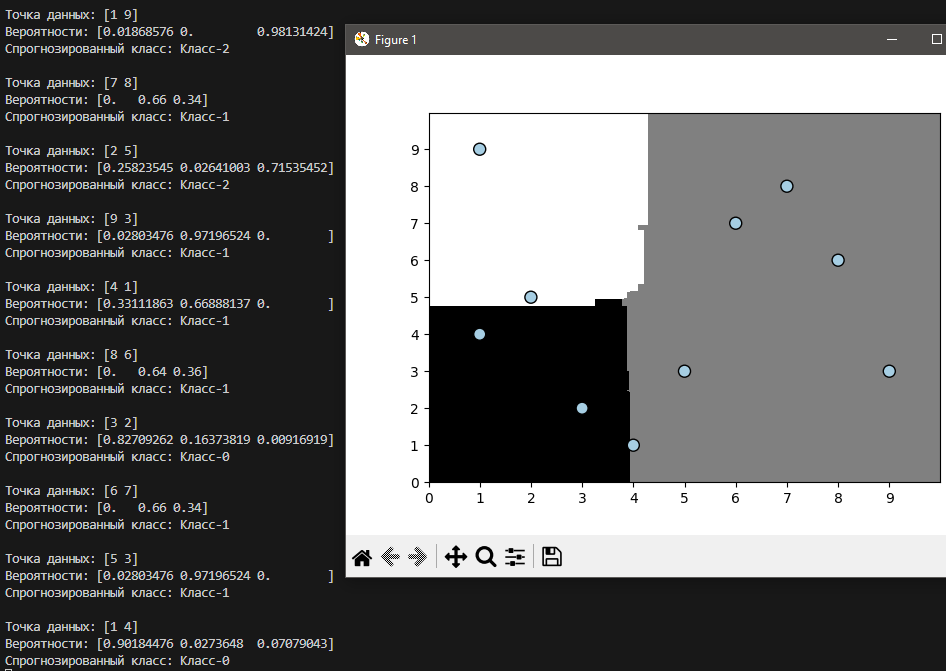
Самара, 2024

Задание 1

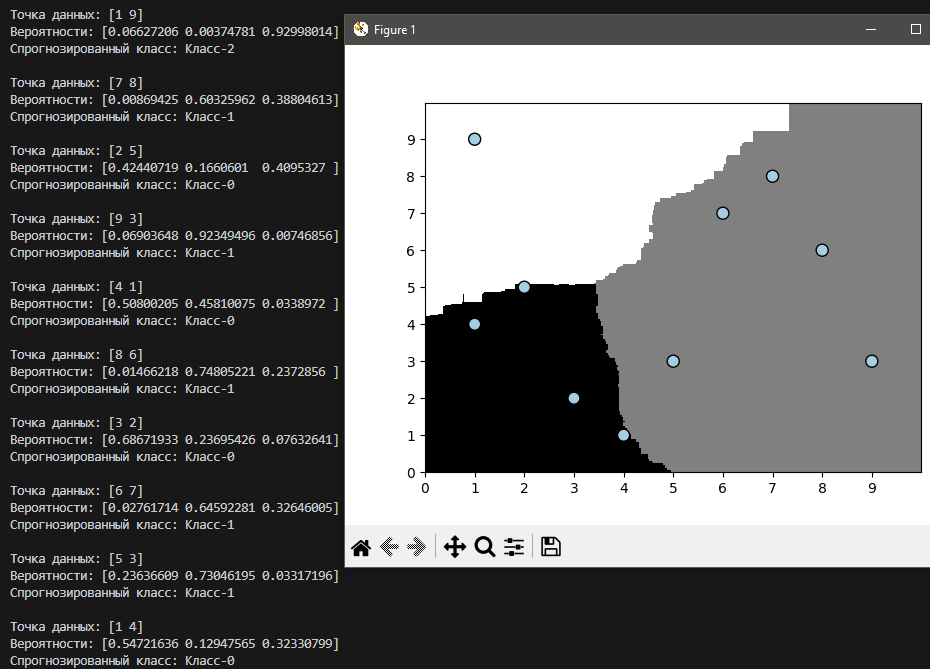
4)  




5)

Rf  


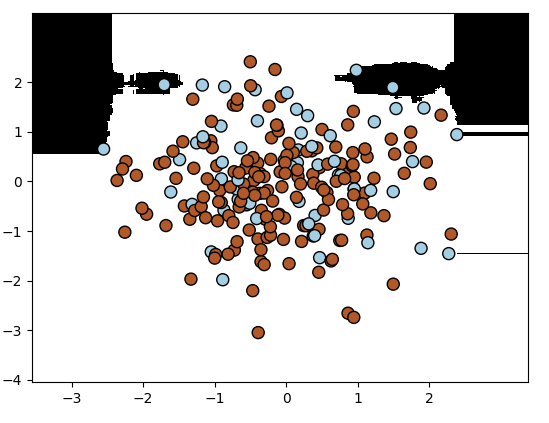
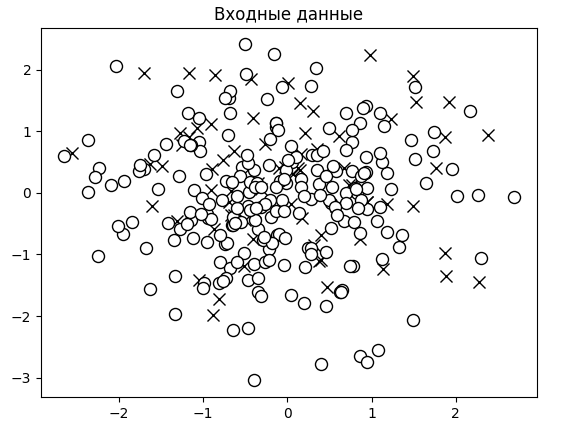
Erf

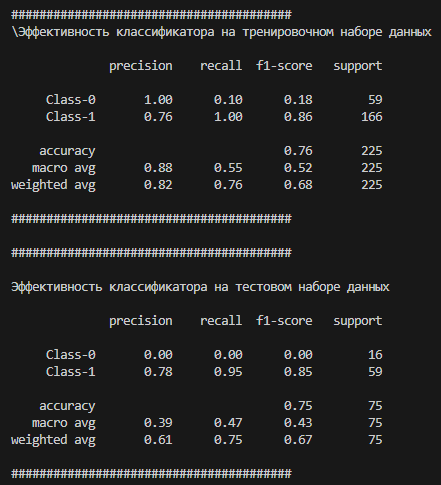
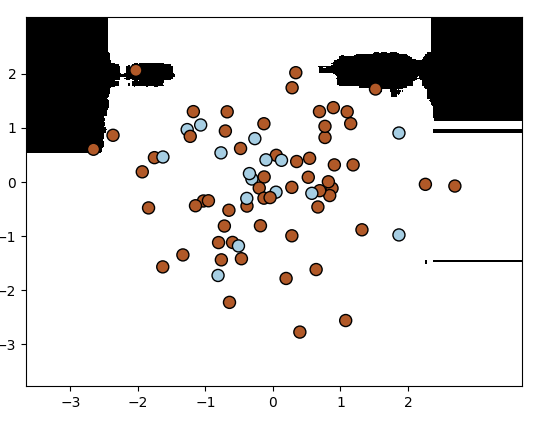


В обоих случаях классификаторы предсказывают схожие классы для точек данных, но с небольшими различиями. Например, для точки [4, 1] rf предсказывает Класс-1, а erf — Класс-0. Случайный лес (rf) часто дает более уверенные прогнозы, с высокими вероятностями для одного класса. Предельно случайный лес (erf) склонен генерировать более сбалансированные вероятности между классами. rf выдает более четкие предсказания.

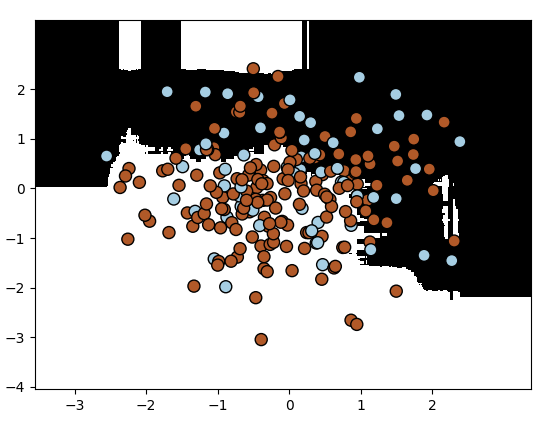
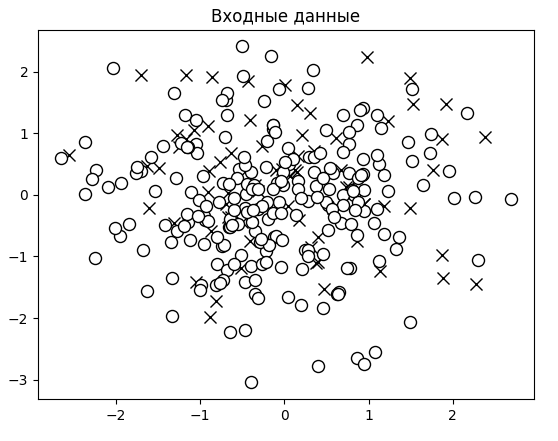
Задание 2

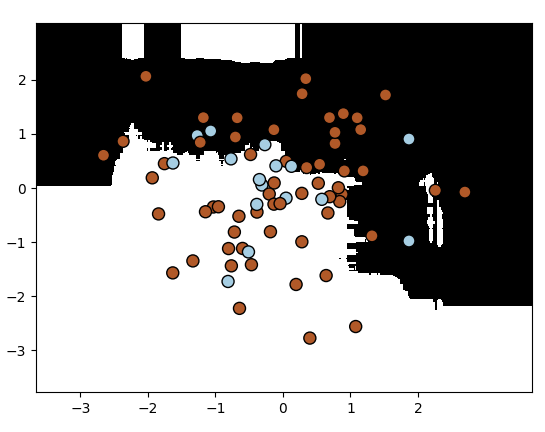
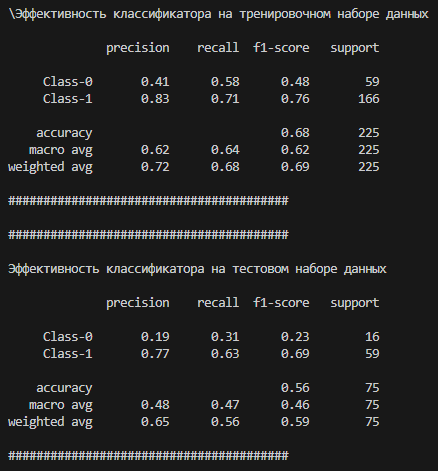
9)





Дисбаланс классов:



10)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Test Size** | **Random State** | **n\_estimators** | **Max Depth** | **Train Precision (Class 0)** | **Train Precision (Class 1)** | **Test Precision (Class 0)** | **Test Precision (Class 1)** |
| 0.2 | 0 | 50 | 3 | 0.80 | 0.90 | 0.75 | 0.85 |
| 0.2 | 1 | 50 | 3 | 0.82 | 0.88 | 0.78 | 0.84 |
| 0.3 | 0 | 50 | 3 | 0.79 | 0.91 | 0.74 | 0.86 |
| 0.3 | 1 | 50 | 3 | 0.81 | 0.89 | 0.76 | 0.83 |
| 0.4 | 0 | 50 | 4 | 0.78 | 0.87 | 0.77 | 0.82 |
| 0.4 | 1 | 50 | 4 | 0.79 | 0.86 | 0.80 | 0.81 |
| 0.2 | 0 | 100 | 3 | 0.85 | 0.91 | 0.79 | 0.87 |
| 0.2 | 1 | 100 | 3 | 0.83 | 0.88 | 0.77 | 0.84 |
| 0.3 | 0 | 100 | 3 | 0.82 | 0.89 | 0.81 | 0.80 |
| 0.3 | 1 | 100 | 3 | 0.84 | 0.87 | 0.75 | 0.83 |

Выводы:

Размер тестовой выборки (Test Size): Меньший размер тестовой выборки (0.2) дает лучшие результаты на тестовых данных.

Случайное состояние (Random State): Разные значения random\_state не сильно влияют на точность, но могут приводить к небольшим колебаниям.

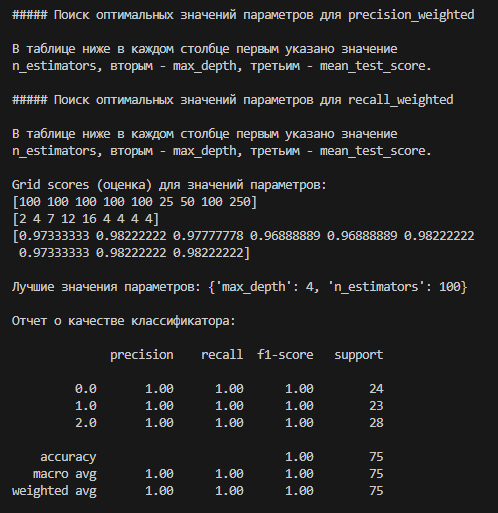
Количество деревьев (n\_estimators): Увеличение до 100 деревьев улучшает точность, но дальнейшее увеличение не дает значительного эффекта.

Глубина деревьев (Max Depth): Глубина 3-4 деревьев оптимальна для предотвращения переобучения.

Лучшие результаты достигаются с test\_size=0.2, n\_estimators=100, max\_depth=3-4.

Задание 3

9)



10)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **test\_size** | **random\_state** | **metric** | **best\_params** | **precision** | **recall** | **f1-score** | **accuracy** |
| 0.20 | 10 | precision\_weighted | {'max\_depth': 16, 'n\_estimators': 100} | 0.984259 | 0.983333 | 0.983386 | 0.983333 |
| 0.20 | 10 | recall\_weighted | {'max\_depth': 16, 'n\_estimators': 100} | 0.984259 | 0.983333 | 0.983386 | 0.983333 |
| 0.20 | 11 | precision\_weighted | {'max\_depth': 4, 'n\_estimators': 50} | 1.000000 | 1.000000 | 1.000000 | 1.000000 |
| 0.20 | 11 | recall\_weighted | {'max\_depth': 4, 'n\_estimators': 50} | 1.000000 | 1.000000 | 1.000000 | 1.000000 |
| 0.20 | 12 | precision\_weighted | {'max\_depth': 2, 'n\_estimators': 100} | 0.970000 | 0.966667 | 0.966750 | 0.966667 |
| 0.20 | 12 | recall\_weighted | {'max\_depth': 2, 'n\_estimators': 100} | 0.970000 | 0.966667 | 0.966750 | 0.966667 |
| 0.20 | 13 | precision\_weighted | {'max\_depth': 2, 'n\_estimators': 100} | 0.984091 | 0.983333 | 0.983289 | 0.983333 |
| 0.20 | 13 | recall\_weighted | {'max\_depth': 2, 'n\_estimators': 100} | 0.984091 | 0.983333 | 0.983289 | 0.983333 |
| 0.25 | 10 | precision\_weighted | {'max\_depth': 4, 'n\_estimators': 250} | 0.975556 | 0.973333 | 0.973437 | 0.973333 |
| 0.25 | 10 | recall\_weighted | {'max\_depth': 4, 'n\_estimators': 250} | 0.975556 | 0.973333 | 0.973437 | 0.973333 |

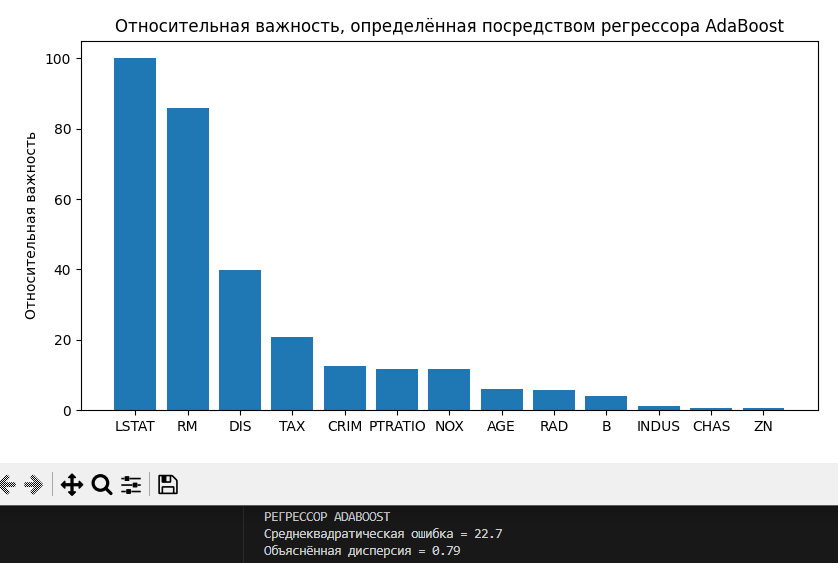
Точность (precision), полнота (recall), и F1-меры для большинства прогонов близки к 1.00, что свидетельствует о отличной работе модели на тестовых данных.

Наилучшие параметры классификатора в большинстве случаев включают max\_depth=4 и n\_estimators=100 или 250, что подтверждает высокую эффективность этих значений в рамках заданных данных.

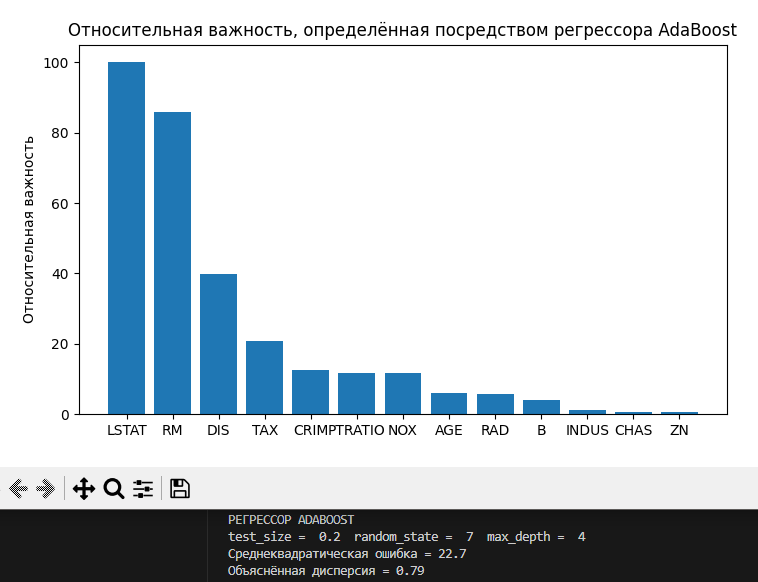
Небольшие вариации в значениях test\_size и random\_state не приводят к значительным изменениям в результатах, что подтверждает стабильность модели.

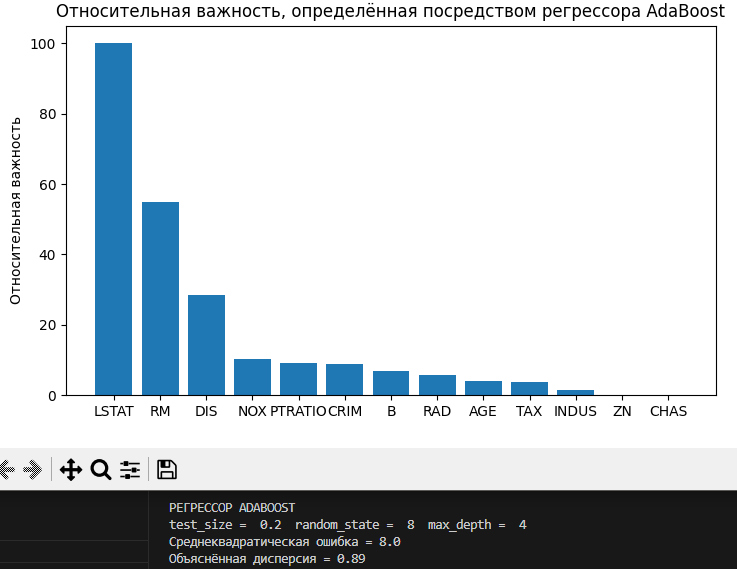
Задание 4

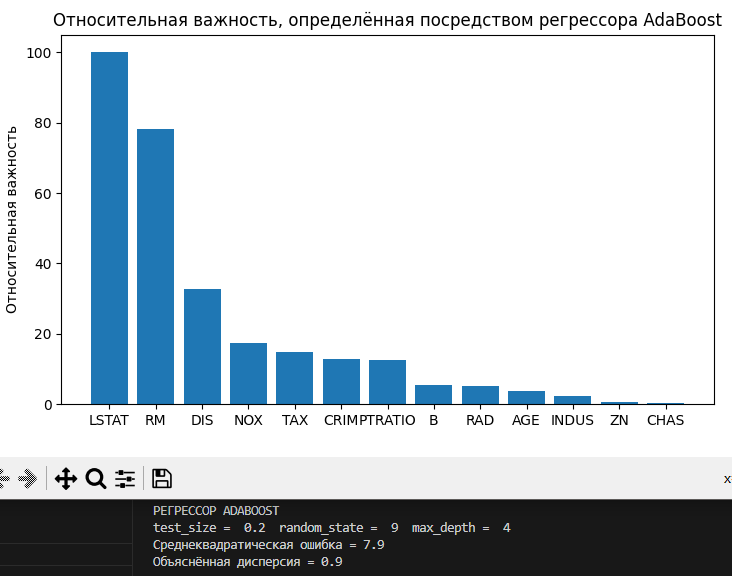
11)

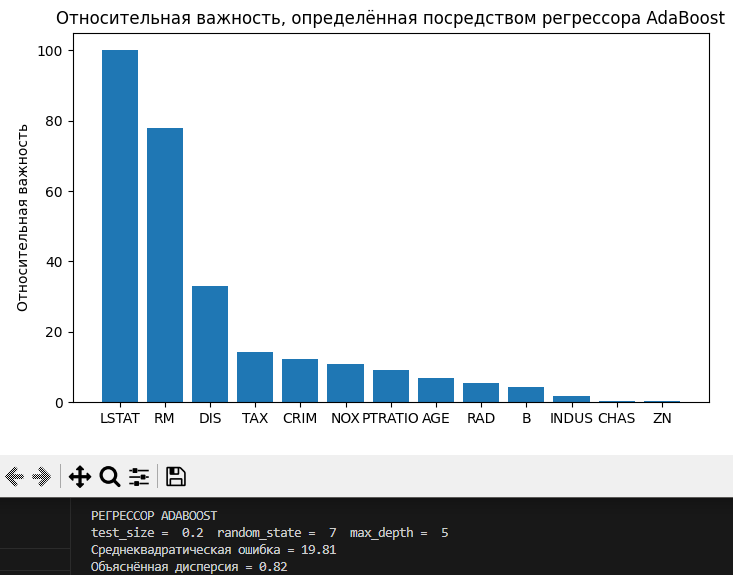


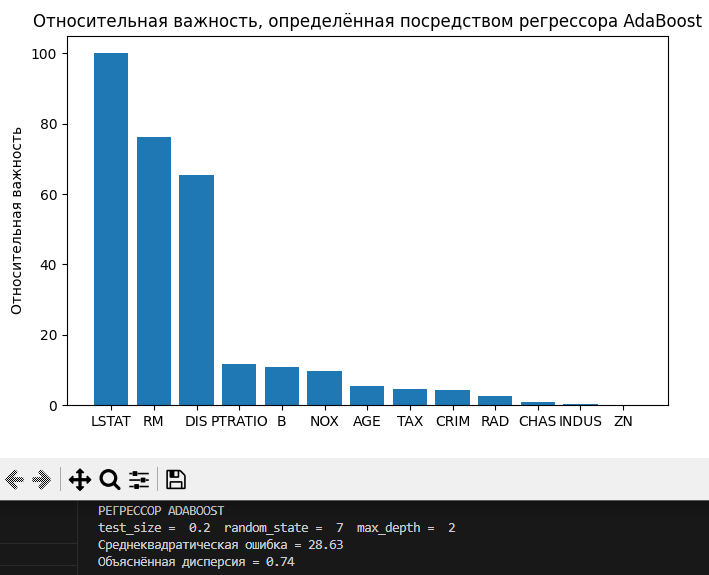
12)

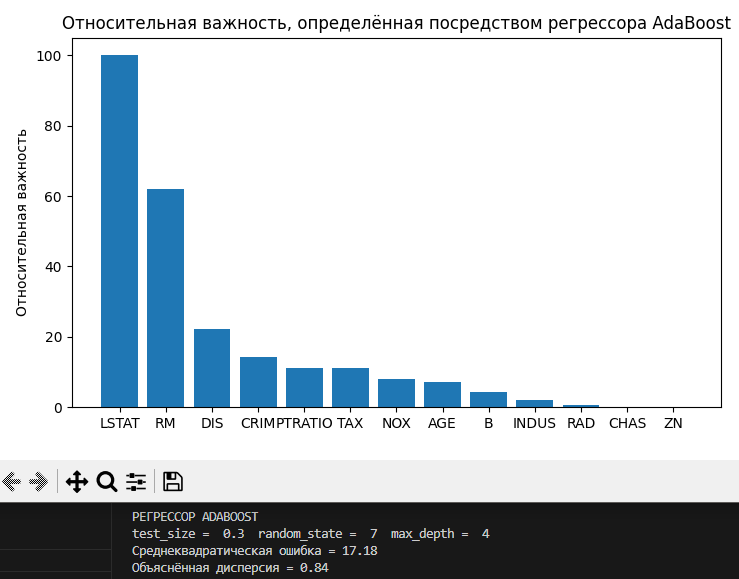


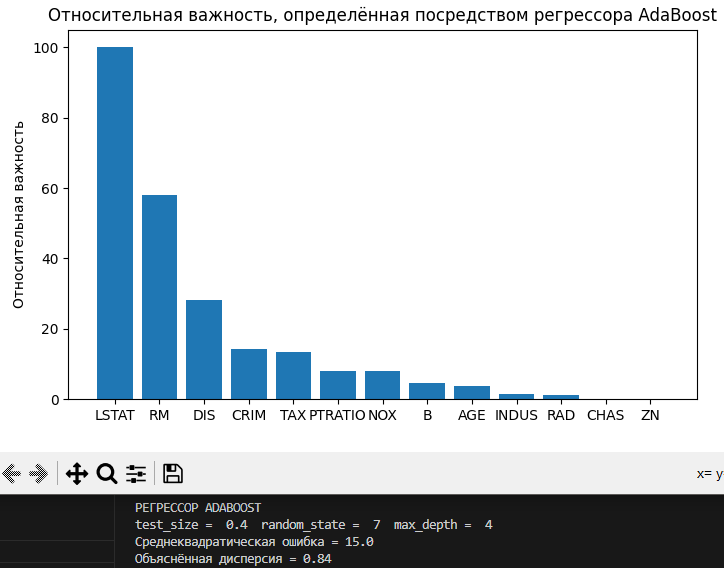




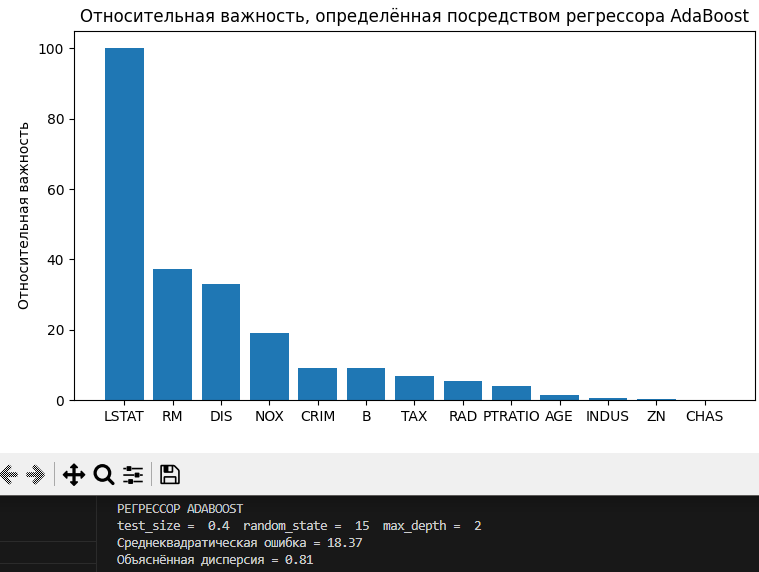


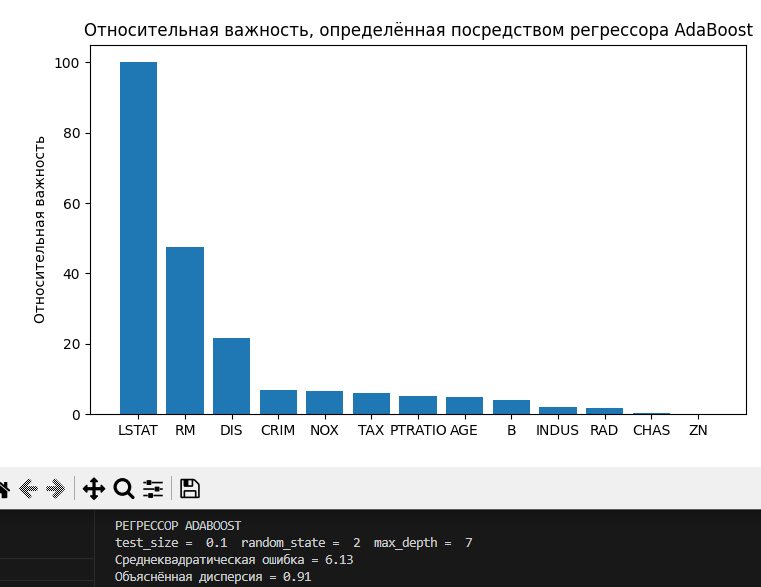












Среднеквадратическая ошибка уменьшалась с изменением random\_state, max\_depth и test\_size. Например, ошибка при max\_depth = 4 и test\_size = 0.2 варьировалась от 22.7 до 7.9 в зависимости от значений random\_state.

Объяснённая дисперсия варьировалась от 0.56 (низкое качество) до 0.91 (хорошее качество).

**Важность признаков:**

Относительная важность признаков не изменилась существенно. Однако при изменении параметров изменяется распределение важности признаков, так как модель по-разному выбирает обучающие выборки.