Министерство цифрового развития

Федеральное государственное бюджетное образовательное учреждение высшего образования

«Сибирский государственный университет телекоммуникаций и информатики»

(СибГУТИ)

Кафедра прикладной математики и кибернетики

Отчёт

по лабораторной работе № 4 «Сравнительный анализ методов регрессии»

Выполнил:

студент группы ИП-312

Дорогин Н.С.

Работу проверил: старший преподаватель кафедры ПМиК Дементьева К.И.

Новосибирск 2025 г.

# **Введение (задание)**

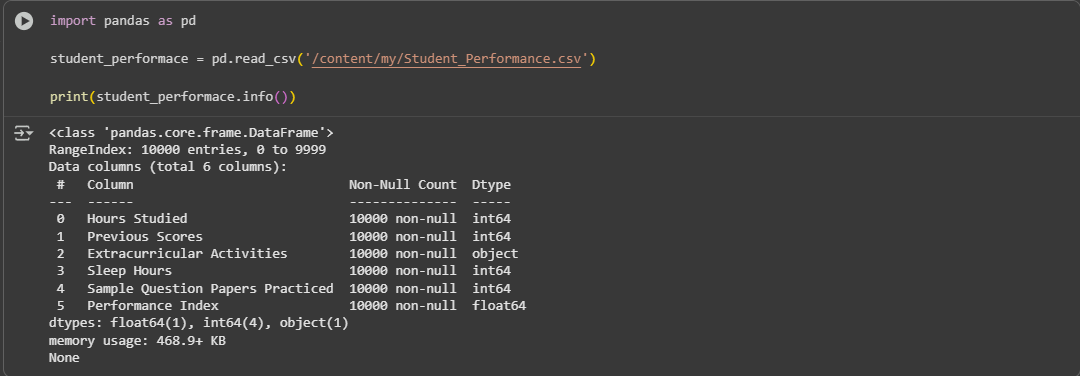
Цель: сформировать комплексное понимание различных методов регрессионного анализа и выработать навыки осознанного выбора моделей в зависимости от характеристик данных и решаемой задачи.

**Основная часть:**

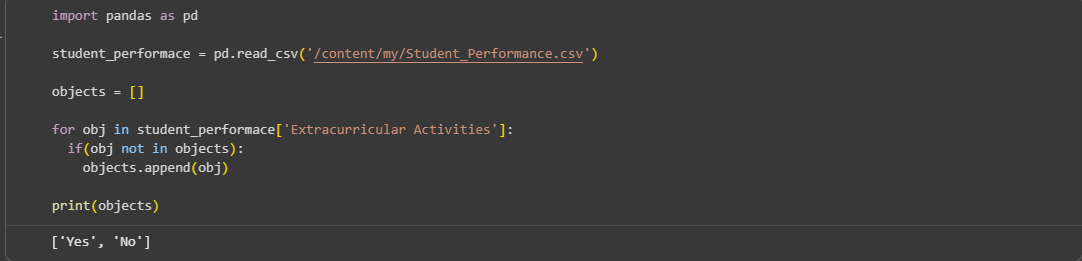
**1. Подготовка данных**

1.1 Выбираем подходящий для задачии датасет с Kaggle. Я выбрал для работы набор данных об академической успеваемости студентов и факторах, влияющих на неё. Целевой переменной будет индекс успеваемости (Performance Index), представляющий собой число с плавающей точкой.

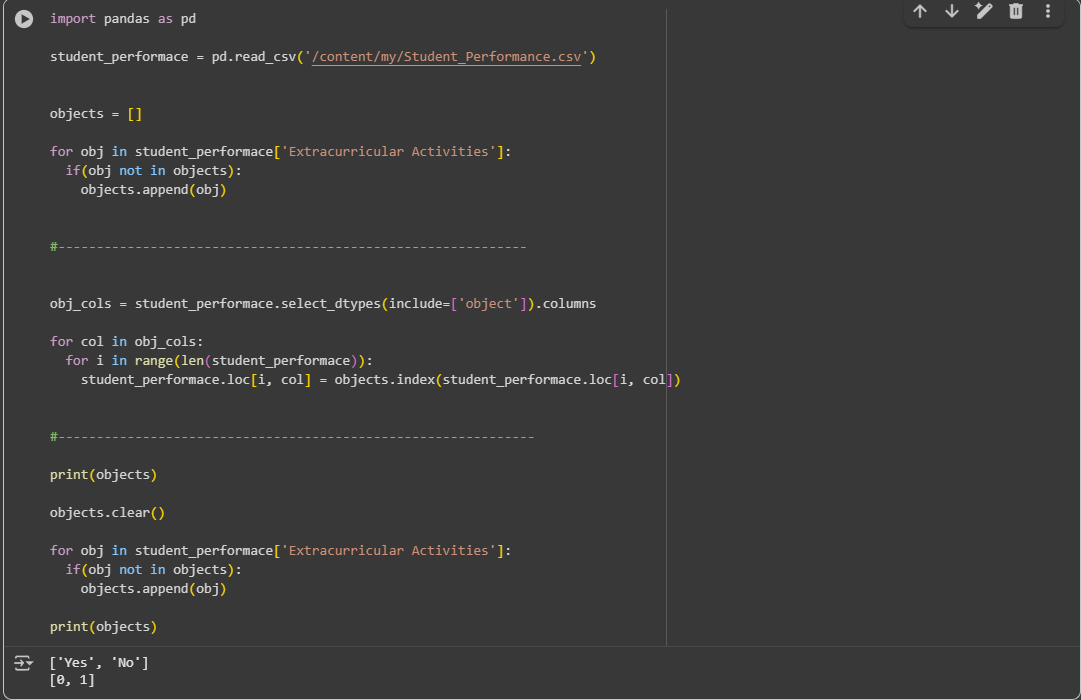
1.2. Загружаем и выполняем предобработку.



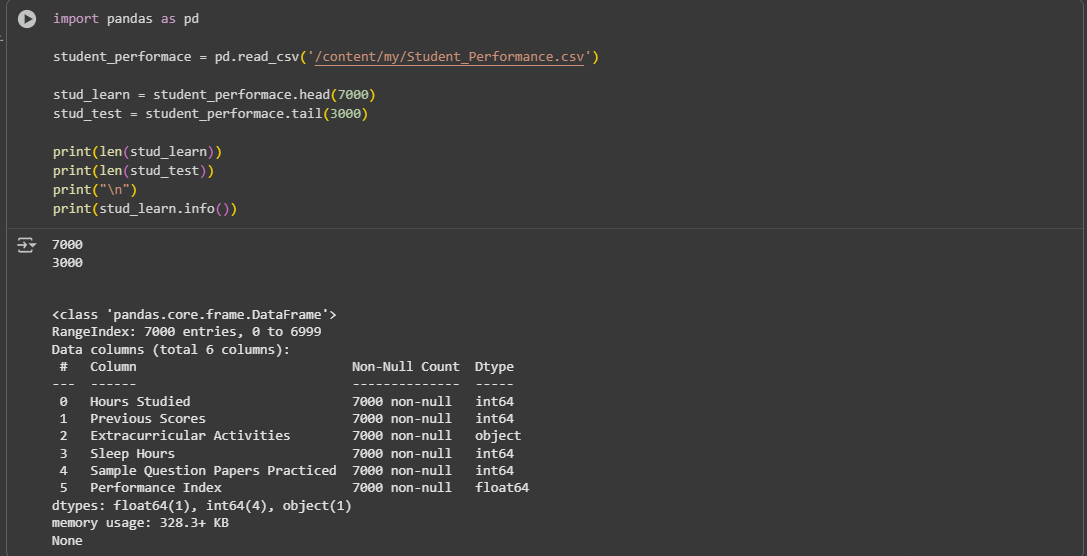
Пустых значений нет, но есть объект в качестве признака, посмотрим какие значения в нём есть.



Видим, что это простейший категориальный признак с двумя вариантами ответа, который легко закодировать.

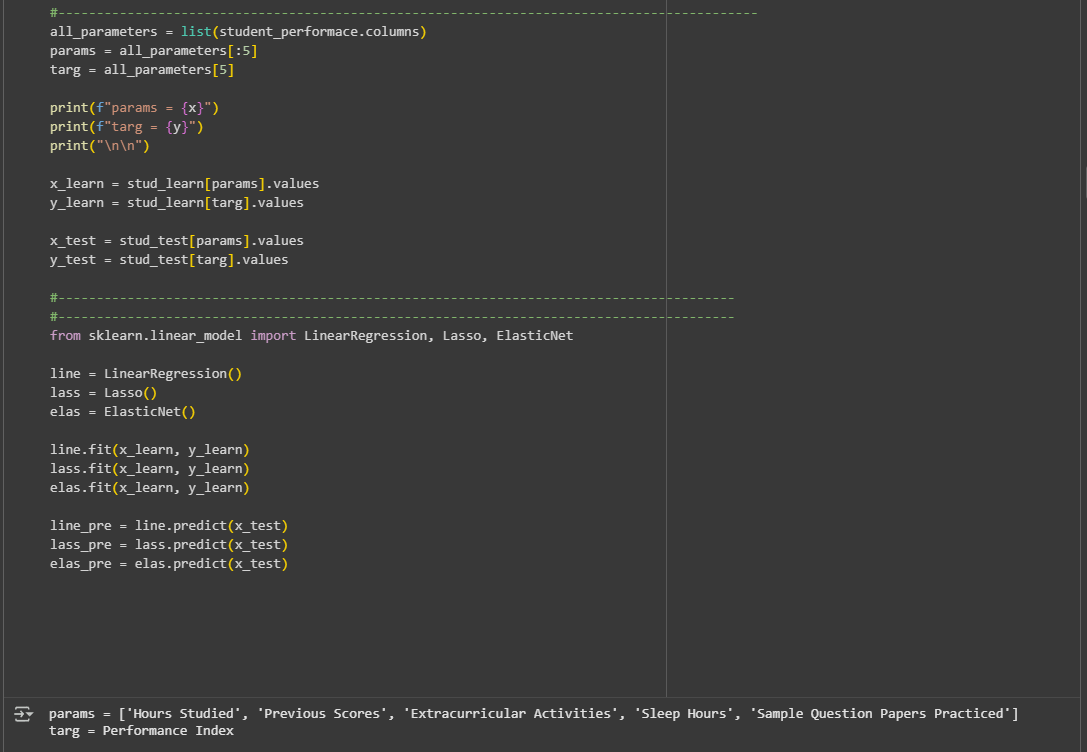


Разделим на обучающую и тестовую выборки в соотношении 70/30, благо у нас для этого идеальные 10000 студентов.

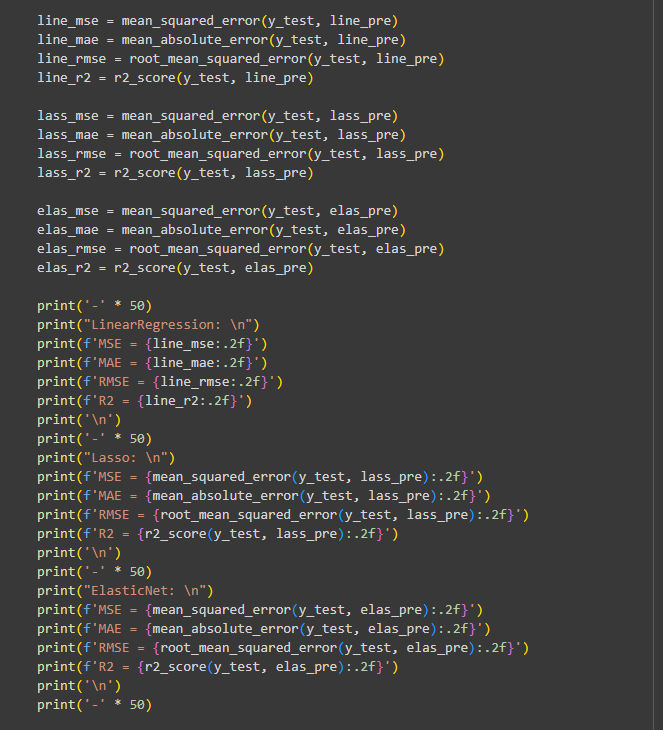


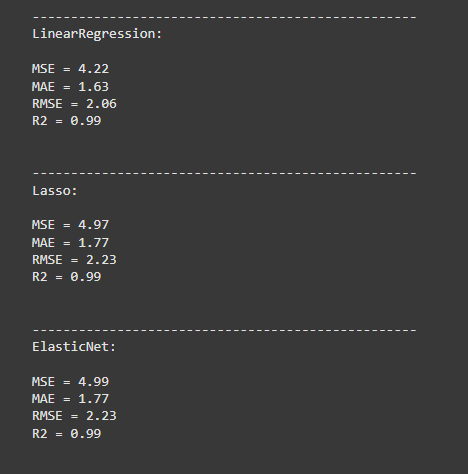
**2. Базовые модели**

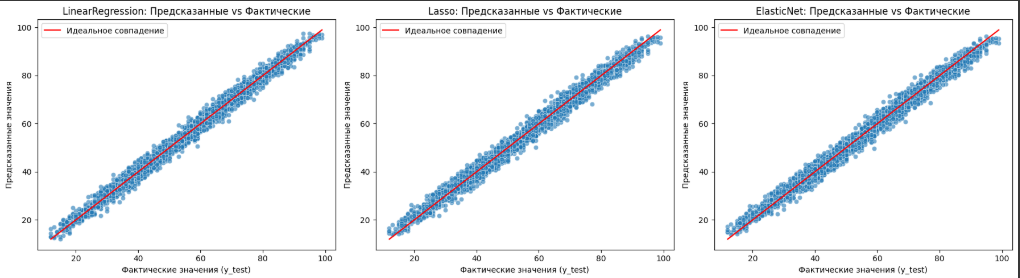
Обучим три базовые модели LinearRegression, Lasso и ElasticNet.

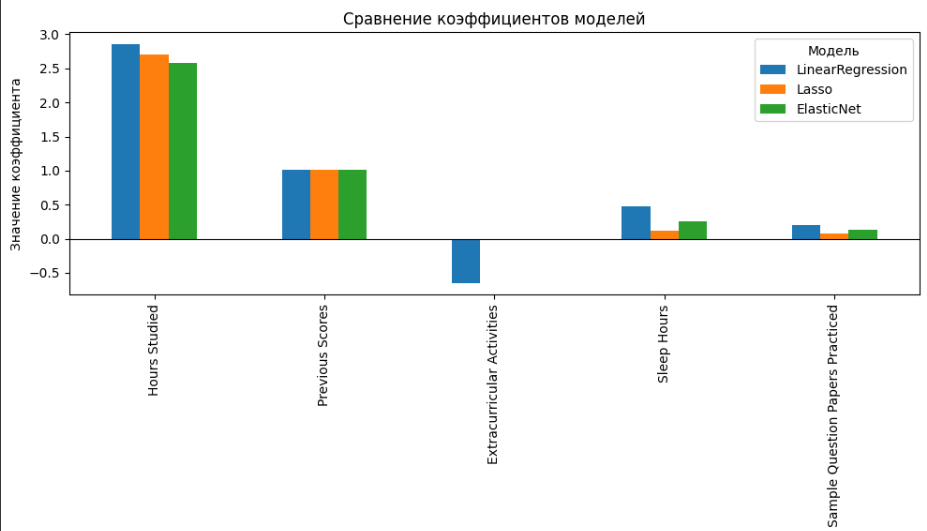


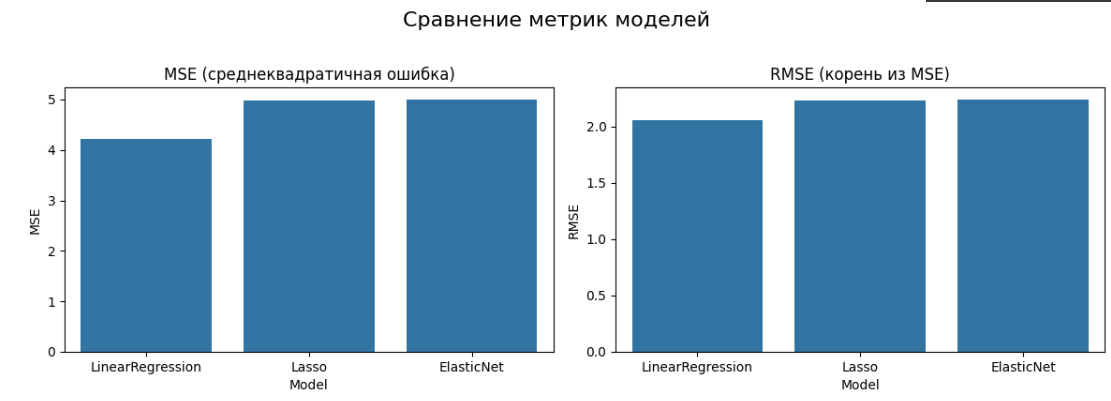
**3. Оценка качества моделей на тестовой выборке**

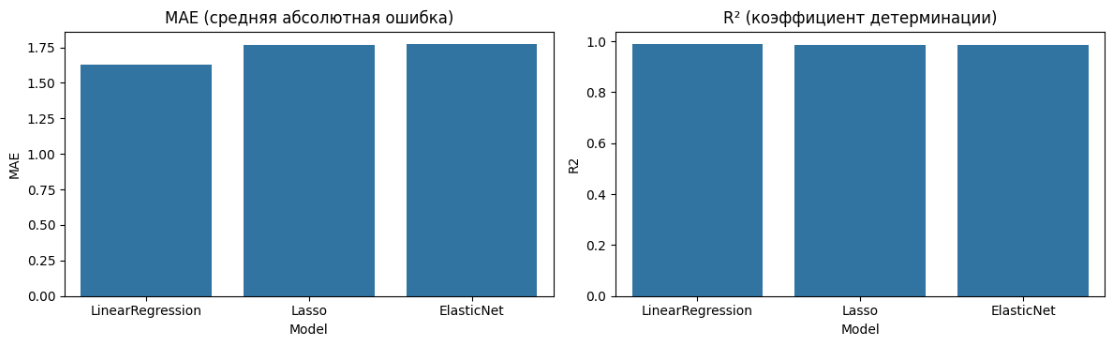
****

****

****

****

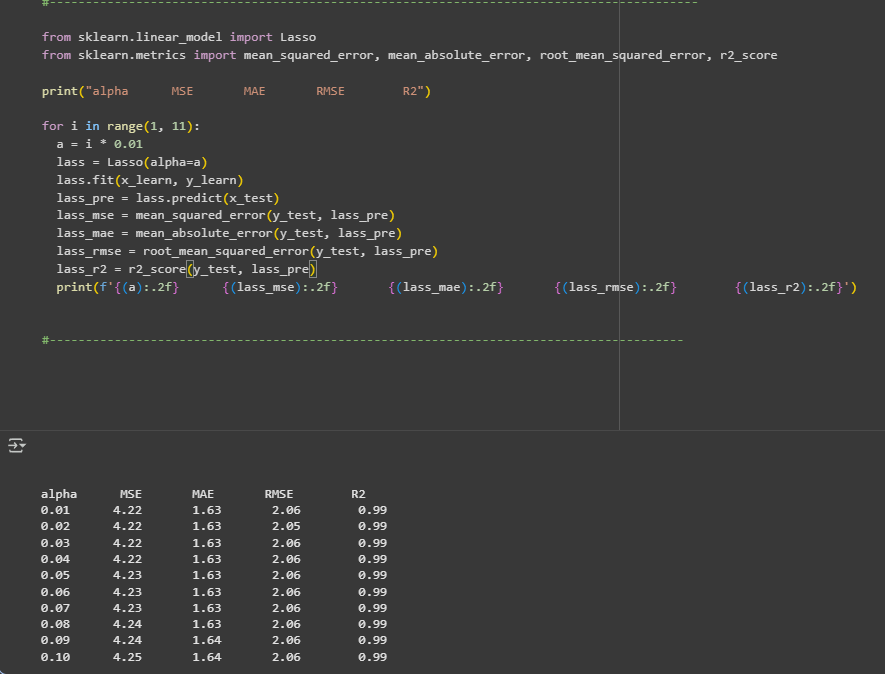
****

****

Из представленных результатов мы можем сказать, что LinearRegression с небольшим отрывом оказалась самой эффективной моделью на нашей тестовой выборке, обладая наименьшими ошибками из всех представленных. Но все модели при этом демонстрируют высокую эффективность

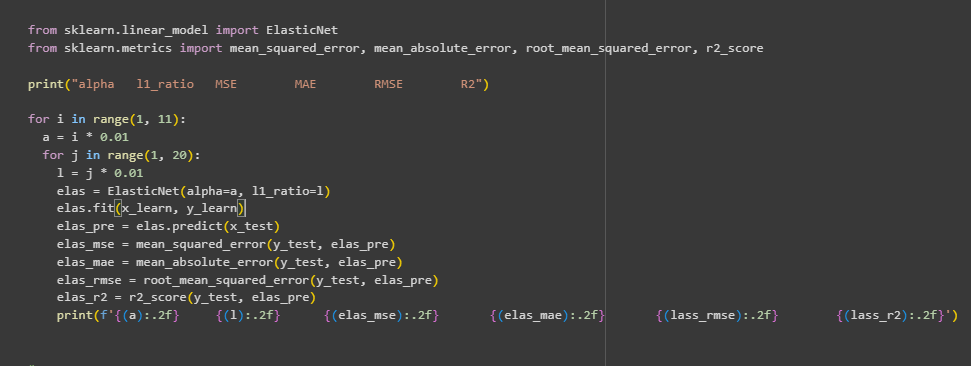
**4. Подбор гиперпараметров и кросс-валидация**

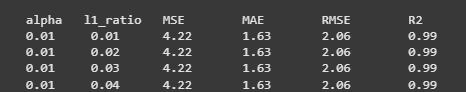
Подберём оптимальный параметр *alpha* для Lasso:

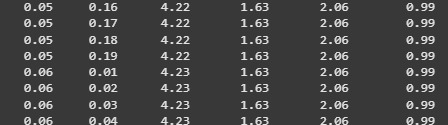


Итого: наблюдаем, что на alpha <= 0.07 Lasso полностью сравнивается по эффективности с LinearRegression.

Подберем оптимальные параметры *alpha* и *l1\_ratio* для ElasticNet:



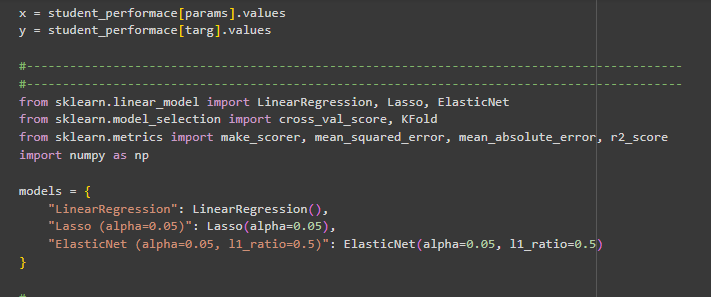


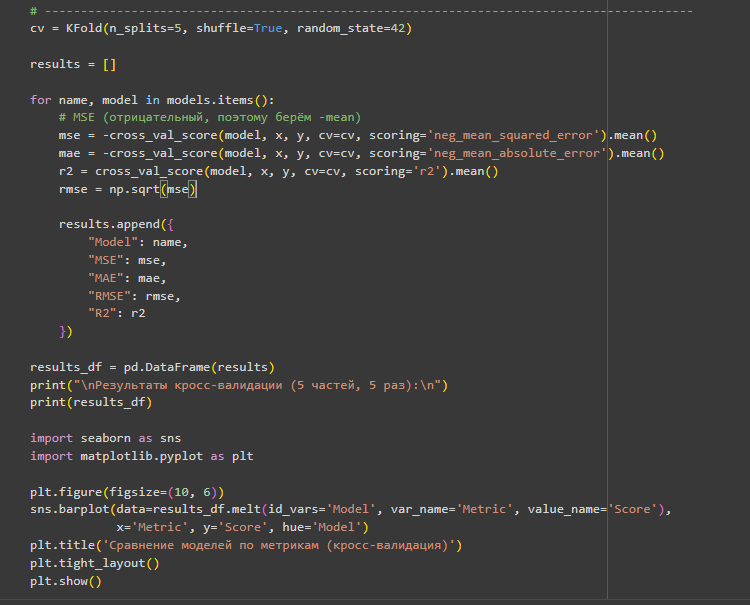


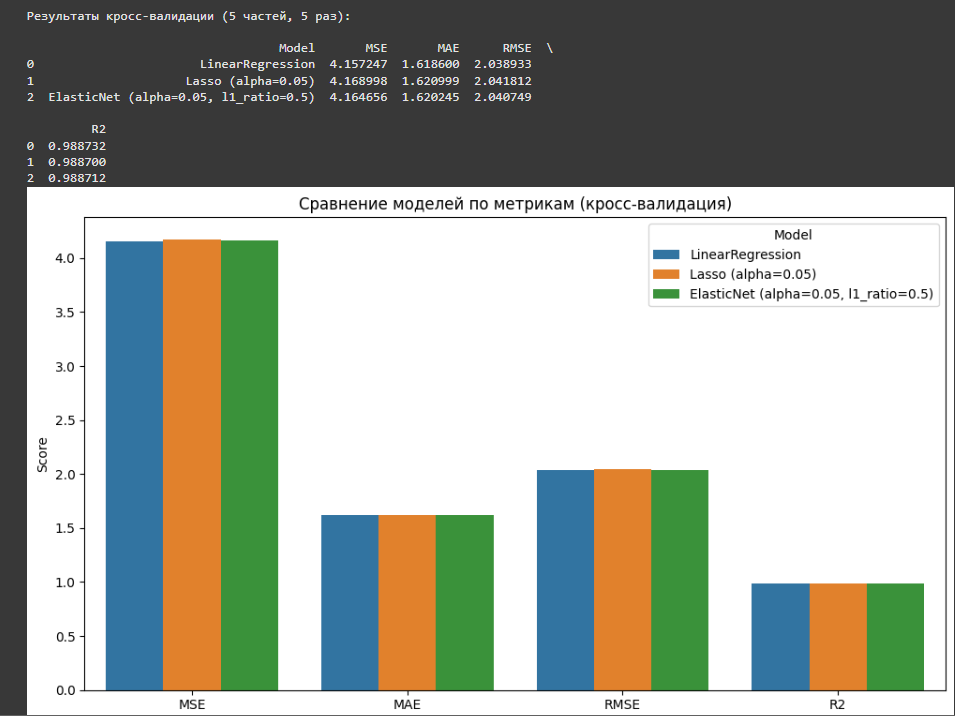
Итого: наблюдаем, что на alpha <= 0.05 ElasticNet полностью сравнивается по эффективности с LinearRegression. l1\_ratio же как-будто и вовсе не оказывает влияния.

То есть, подбор гиперпараметров хоть и немного, но улучшил качество регрессий.

Теперь выполним кросс-валидацию:







Получаем, что при 5-кратном обучения на 5 разных выборках все модели выдают примерно схожие результаты.

**5. Отчёт**

Ссылка на GoogleCollab: <https://colab.research.google.com/drive/1InYZSFYOxjGShodyJXLZvcIH-hd156SS#scrollTo=VkIoL3oM5FKQ>