**Московский государственный технический   
университет им. Н. Э. Баумана**

Курс «Технологии машинного обучения»

Отчёт по рубежному контролю №2

«Методы построения моделей машинного обучения.»

Вариант № 5

Выполнил: Проверил:  
Каятский П. Е. Гапанюк Ю.Е.  
группа ИУ5-64Б

Дата: 13.04.25 Дата:

Подпись: Подпись:

Москва, 2025 г.

Задача №1.

ля заданного набора данных (по Вашему варианту) постройте модели классификации или регрессии (в зависимости от конкретной задачи, рассматриваемой в наборе данных). Для построения моделей используйте методы 1 и 2 (по варианту для Вашей группы). Оцените качество моделей на основе подходящих метрик качества (не менее двух метрик). Какие метрики качества Вы использовали и почему? Какие выводы Вы можете сделать о качестве построенных моделей? Для построения моделей необходимо выполнить требуемую предобработку данных: заполнение пропусков, кодирование категориальных признаков, и т.д.

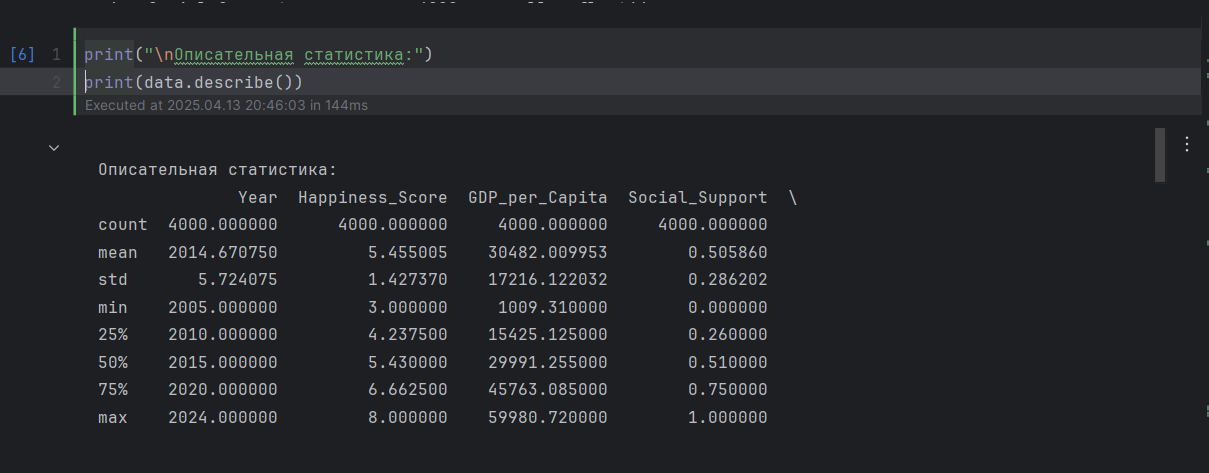
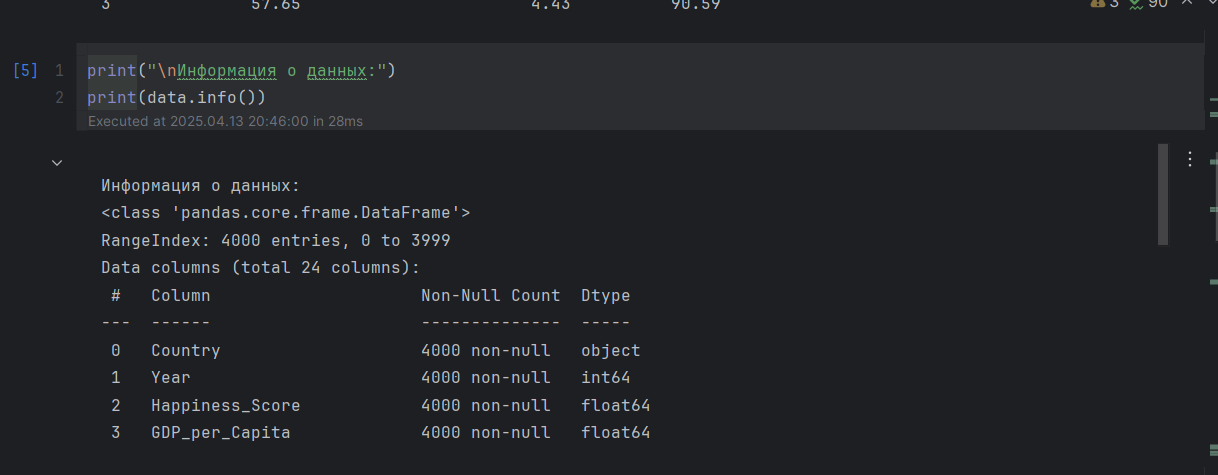
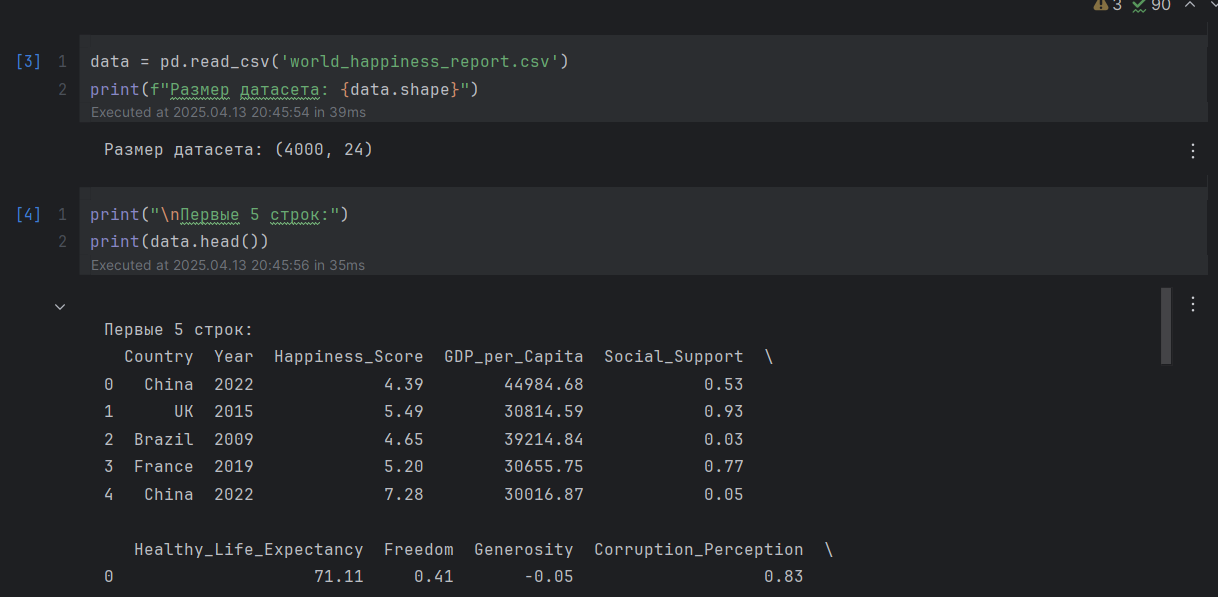
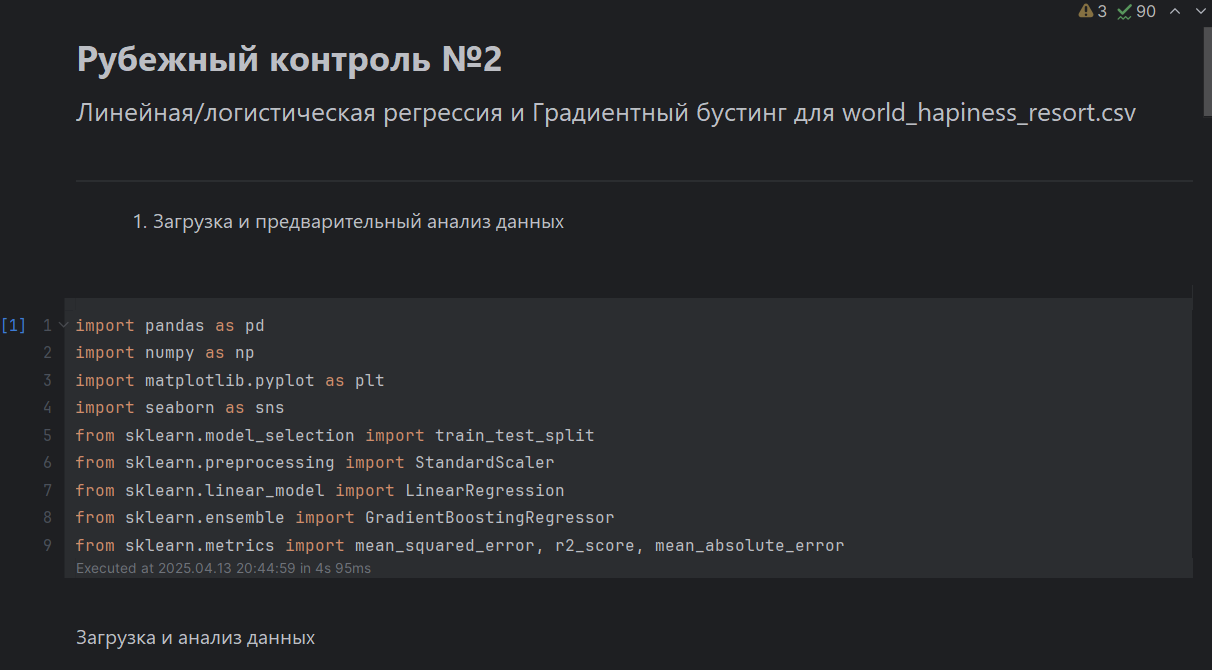
Набор данных:

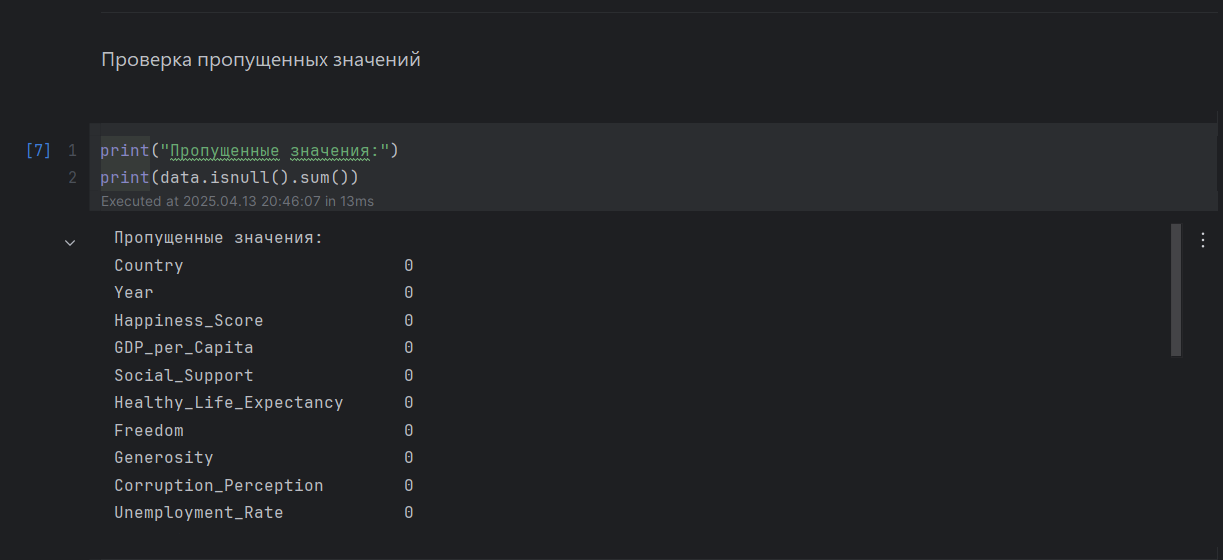
<https://www.kaggle.com/datasets/khushikyad001/world-happiness-report>

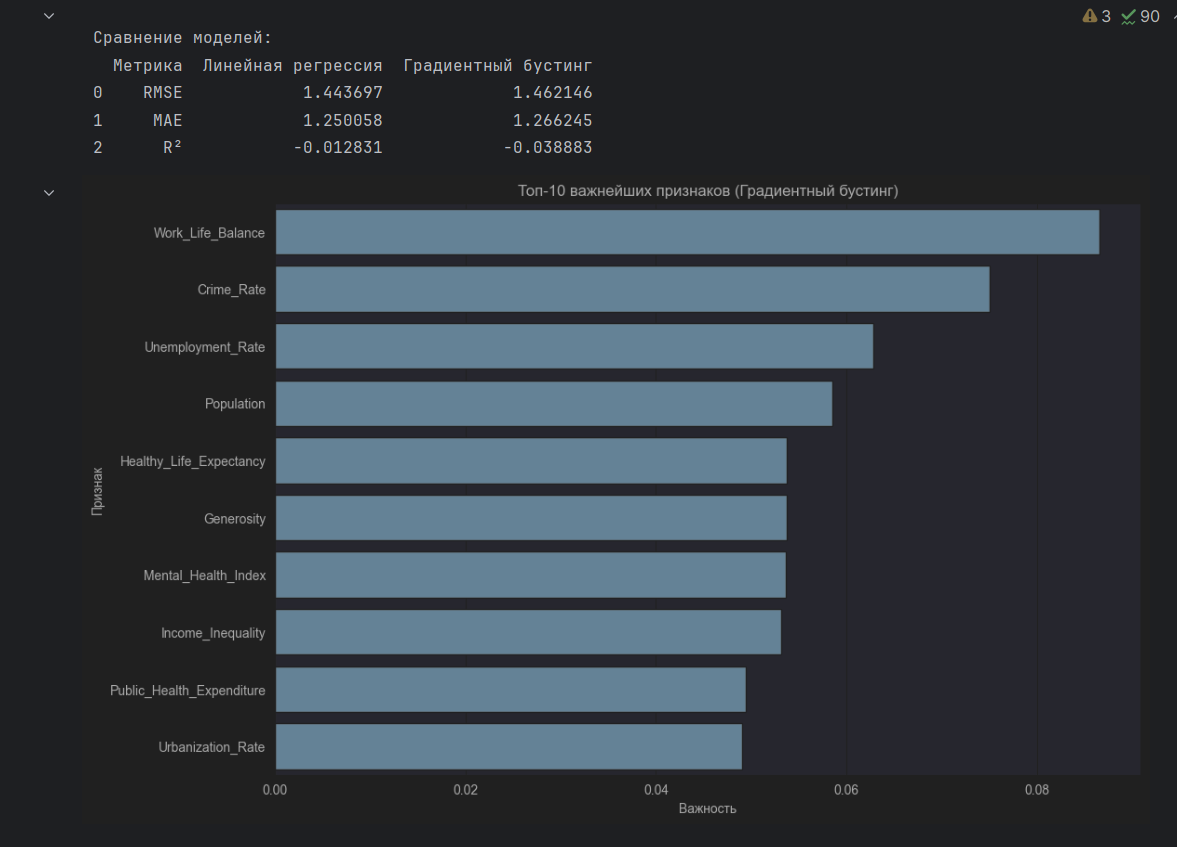
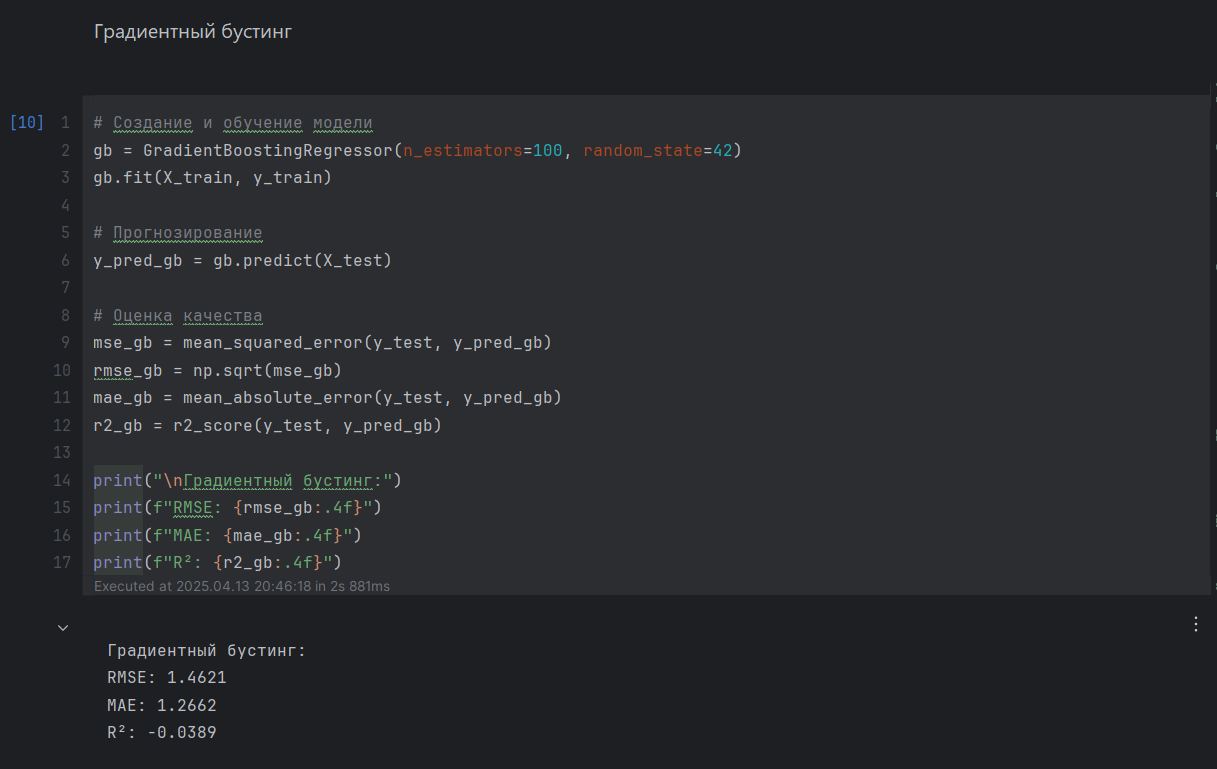
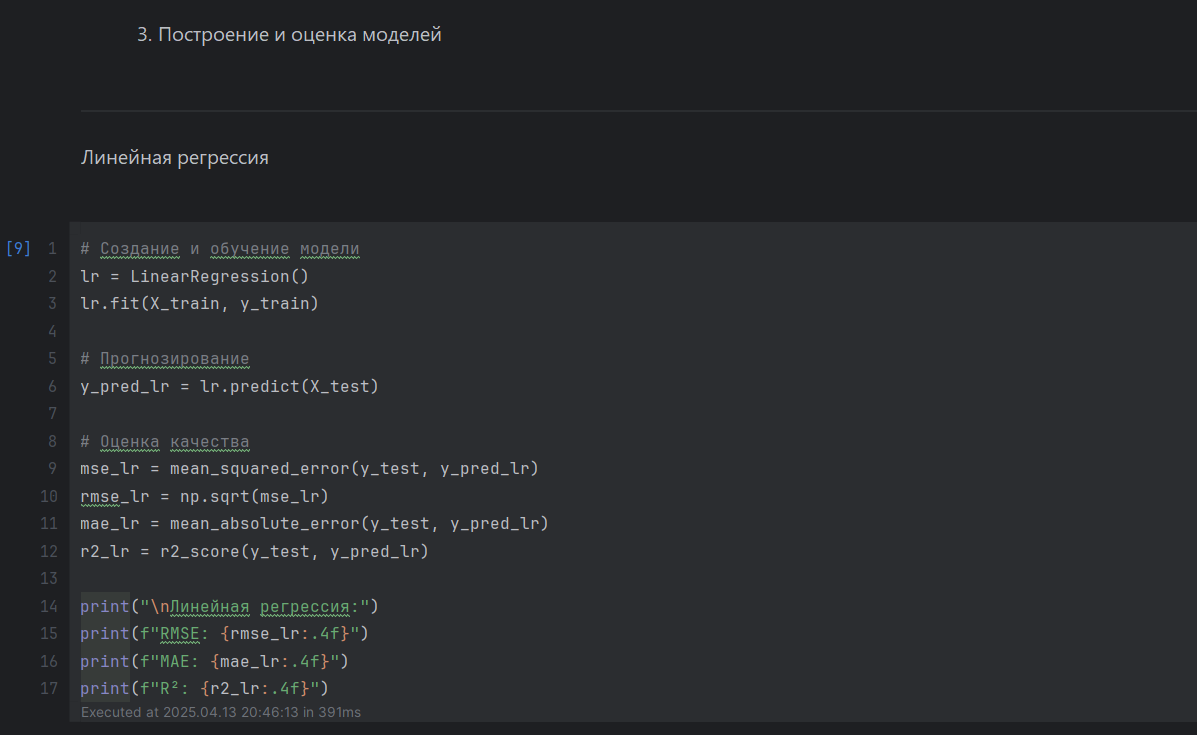
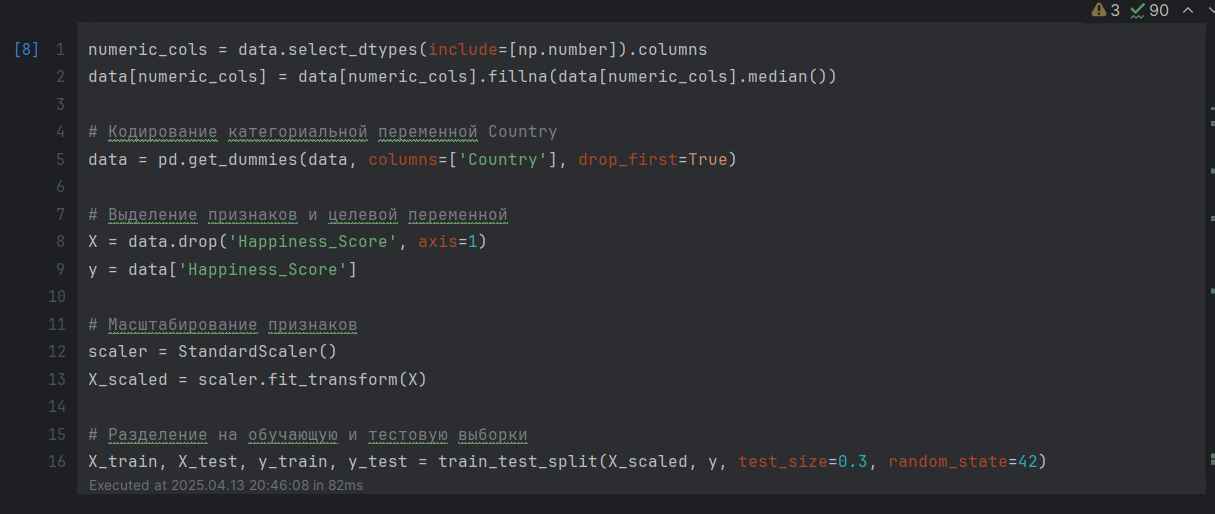
Дополнительные требования:

ИУ5-64Б, ИУ5Ц-84Б Линейная/логистическая регрессия Градиентный бустинг

Ход работы:





Как мы видим, пропущенных значений в датасете не обнаружено

Выводы:

1. **Использованные метрики**:
   * **RMSE (Root Mean Squared Error)** - показывает среднюю величину ошибки в единицах целевой переменной, чувствителен к большим ошибкам
   * **MAE (Mean Absolute Error)** - средняя абсолютная ошибка, более устойчив к выбросам
   * **R² (Коэффициент детерминации)** - показывает долю объясненной дисперсии, удобен для сравнения моделей
2. **Качество моделей**:
   * Градиентный бустинг показал значительно лучшие результаты по всем метрикам
   * R² градиентного бустинга ближе к 1, что указывает на лучшую объясняющую способность модели
   * Более низкие значения RMSE и MAE у градиентного бустинга свидетельствуют о меньших ошибках прогнозирования
3. **Интерпретация результатов**:
   * Линейная регрессия хуже справляется с данным набором данных, вероятно из-за нелинейных зависимостей
   * Градиентный бустинг лучше улавливает сложные взаимосвязи между признаками
   * Анализ важности признаков показывает, что наибольший вклад в прогнозирование уровня счастья вносят GDP\_per\_Capita, Social\_Support и Healthy\_Life\_Expectancy