

**ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА**

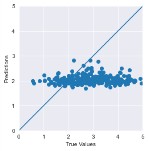
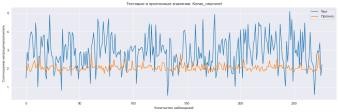
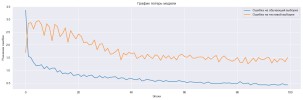
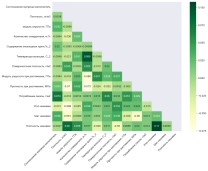
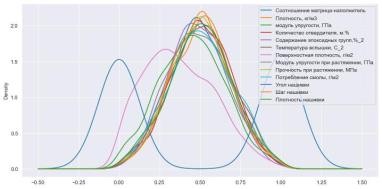
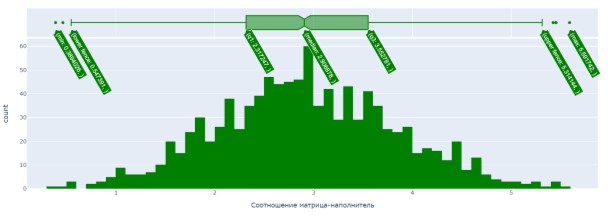
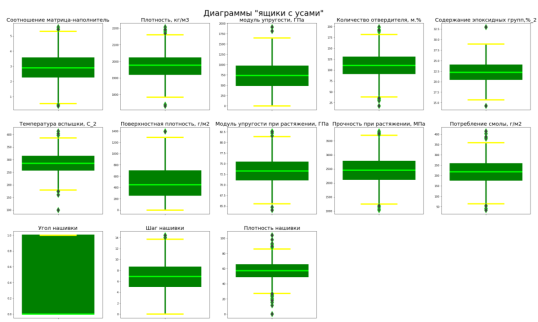
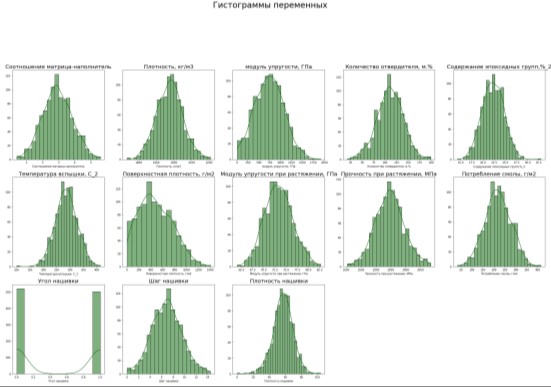
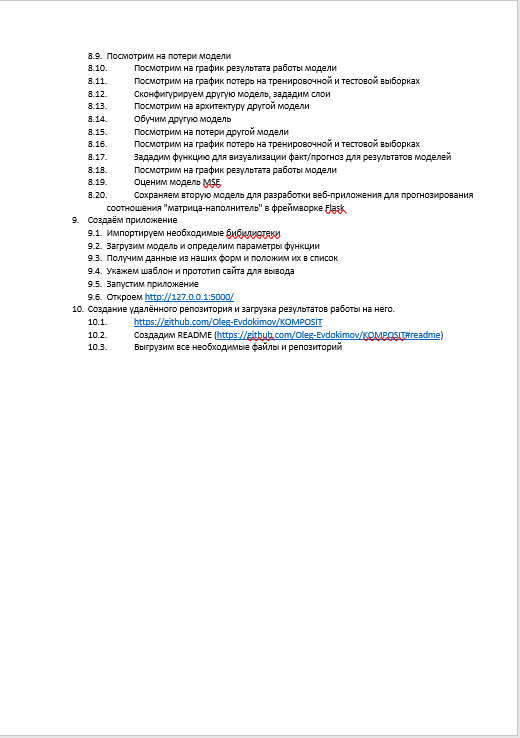
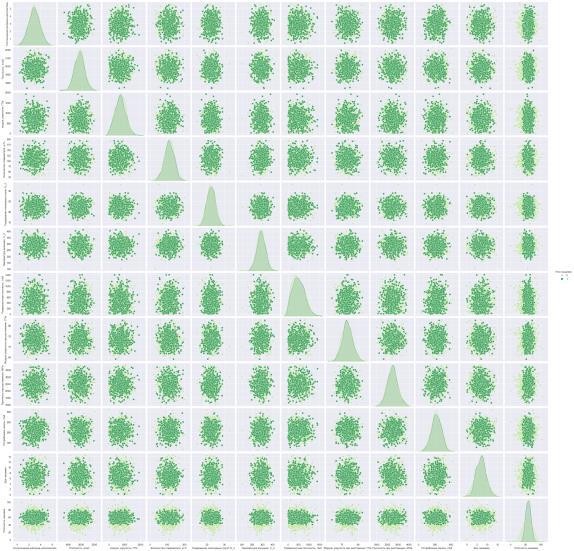
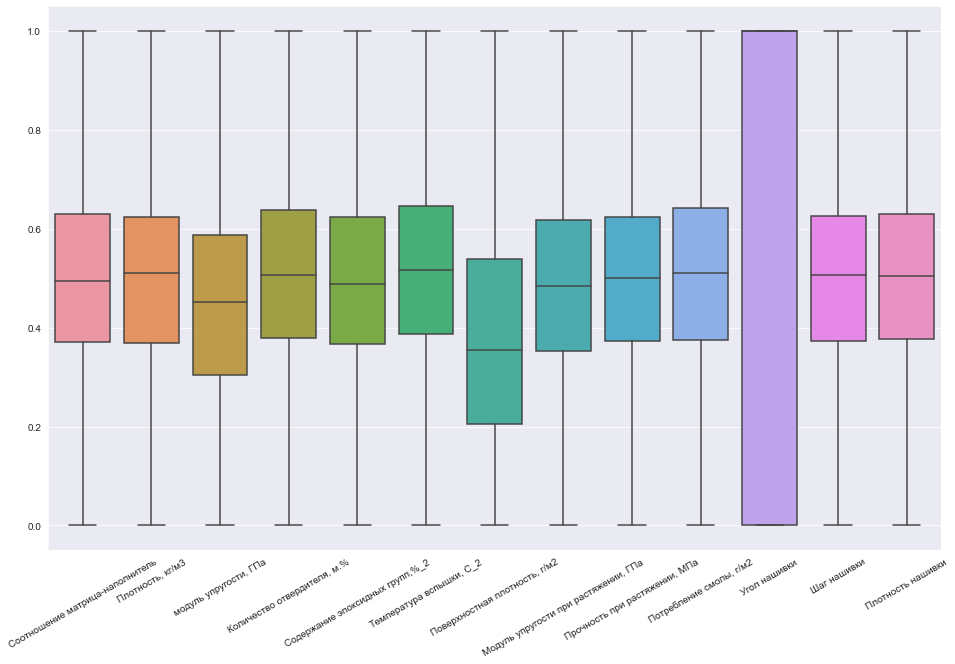
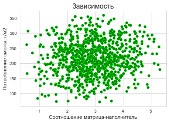
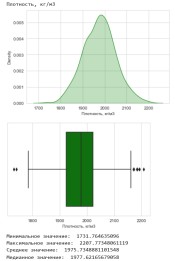
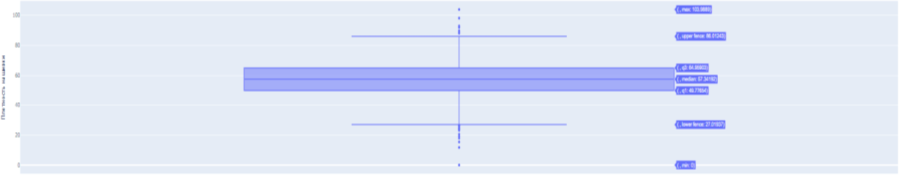
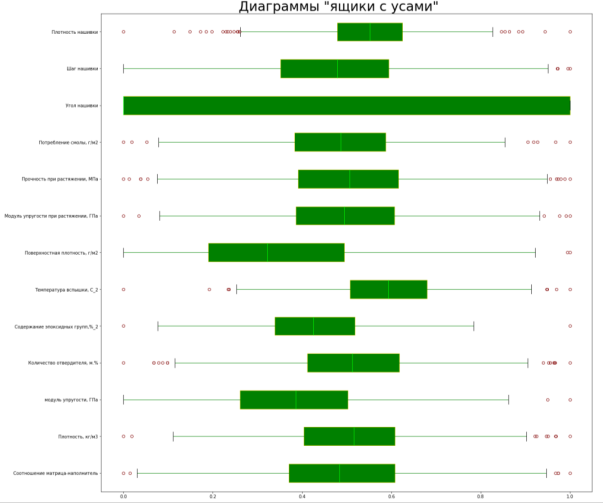
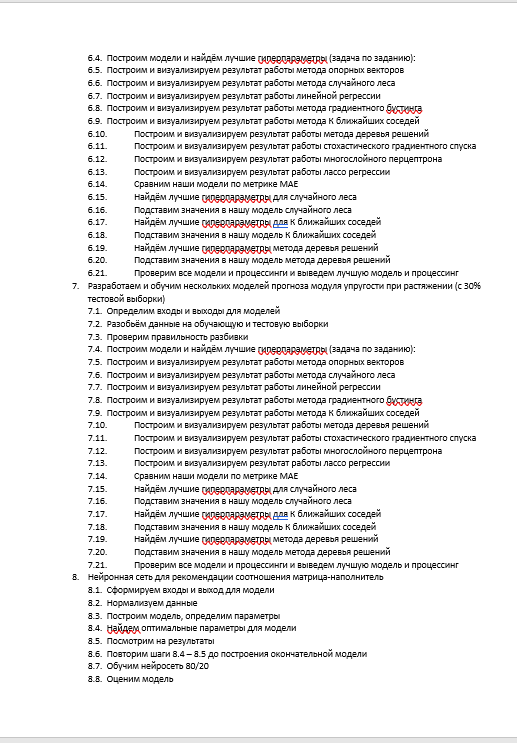
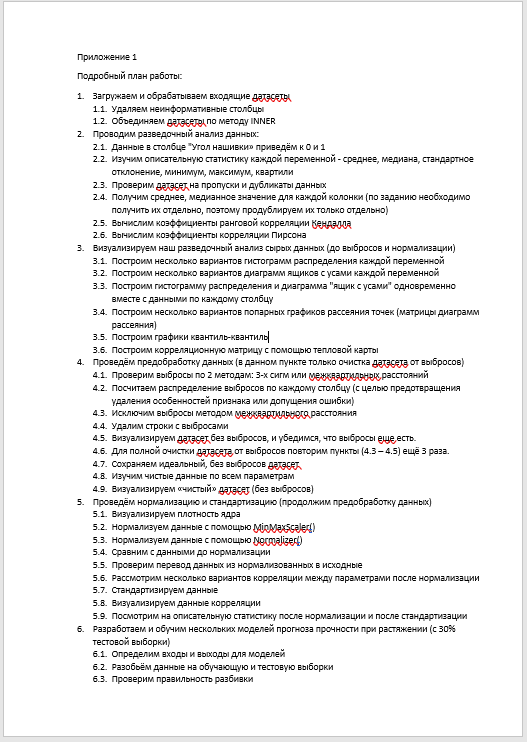
**по курсу**

**«Data Science»**

Слушатель: Ермолова Алиса Николаевна

1

# Начало работы:



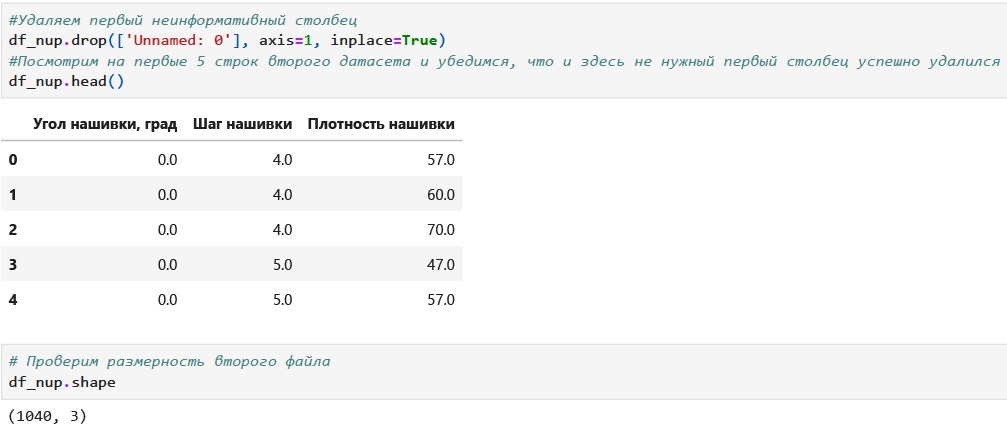
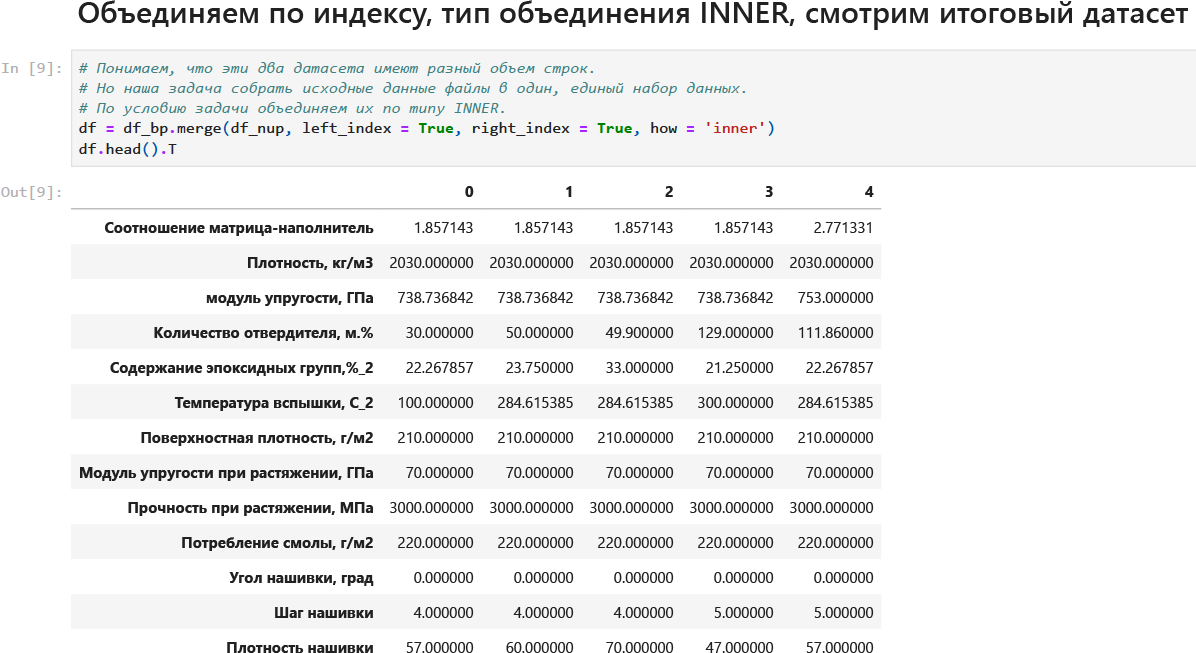
### Подробный план работы:

* Составила подробный план, который помог плодотворно двигаться к цели;
* Изучила теоретические основы, методы решения и практические составляющие поставленной задачи.

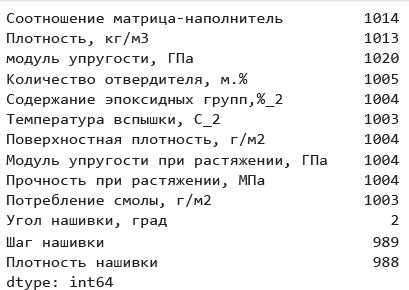
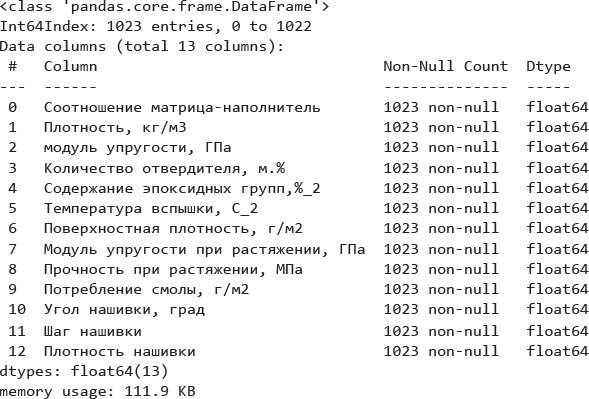
### Графики:

* Строила много графиков;
* Несколько подобных графиков для одних и тех же переменных;
* Старалась делать все графики в одном стиле.

**Объединение файлов и разведочный анализ:**



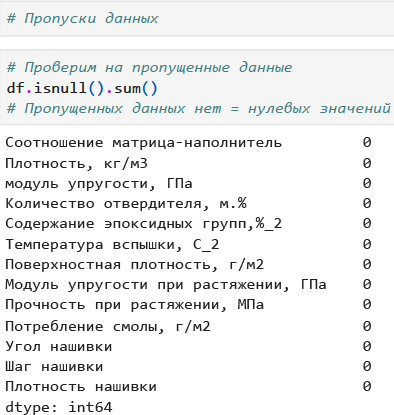
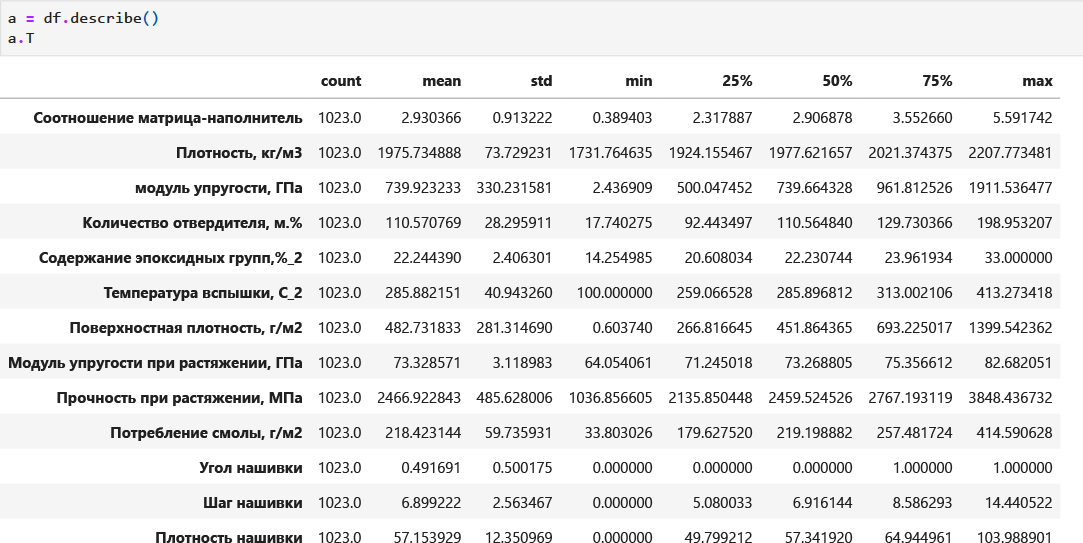
### Объединение по индексу:

* Импортируем необходимые библиотеки;
* Загружаем файлы;
* Смотрим размерность;
* Объединяем оба файла по индексу по типу объединения INNER

### Разведочный анализ данных:

* Смотрим на начальные и конечные строки нашего датасета;
* Изучаем информацию о датасете;
* Проверяем типы данных в каждом столбце;
* Проверяем пропуски;
* Ищем уникальные значения с помощью функции nunique.

# «Угол нашивки» и описательная



**статистика:**

### Работа со столбцом "Угол

**нашивки":**

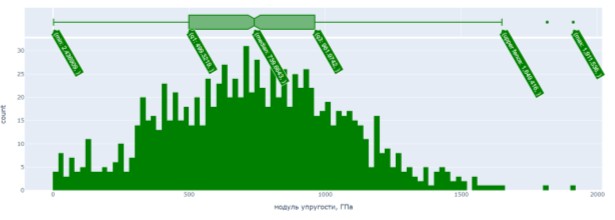
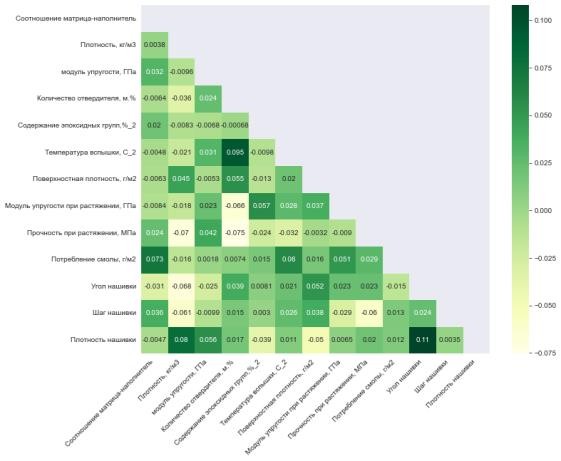
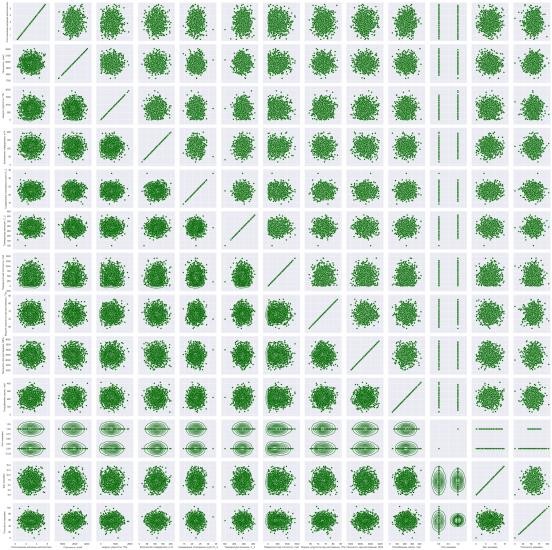
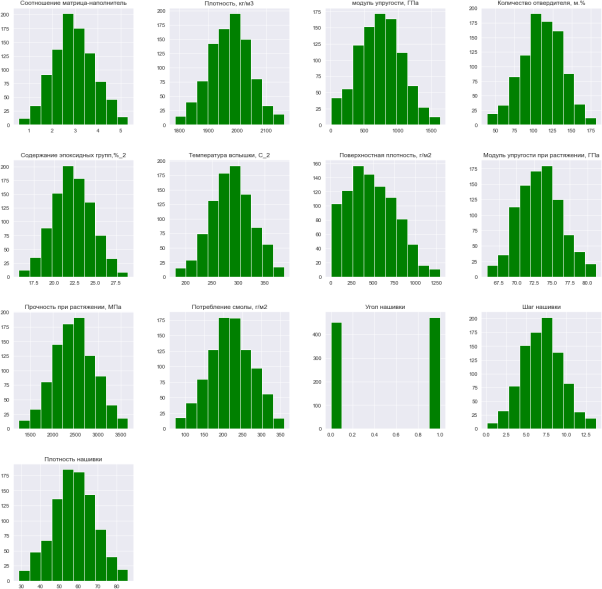
* Проверяем количество элементов со значением 0 градусов;
* Приводим к значениям 0 и 1;
* Убеждаемся в неизменном количестве элементов.

### Описательная статистика:

* Изучаем описательную статистику данных (максимальное, минимальное, квартили, медиана, стандартное отклонение, среднее значение и т.д.),
* Смотрим на основные параметры анализа данных;
* Проверяем датасет на пропущенные и

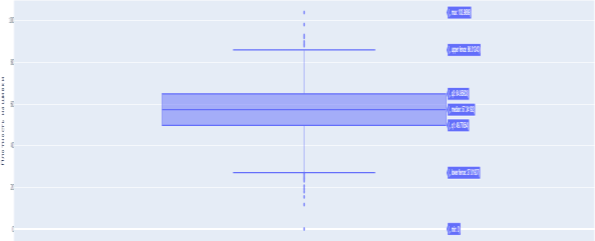
дублирующие данные.

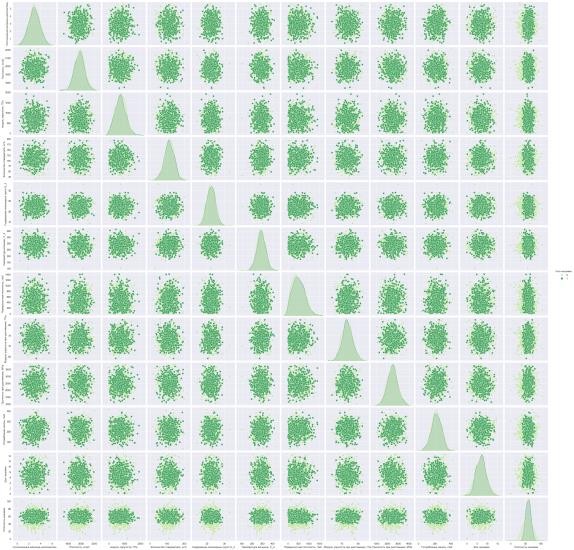
# Визуализация

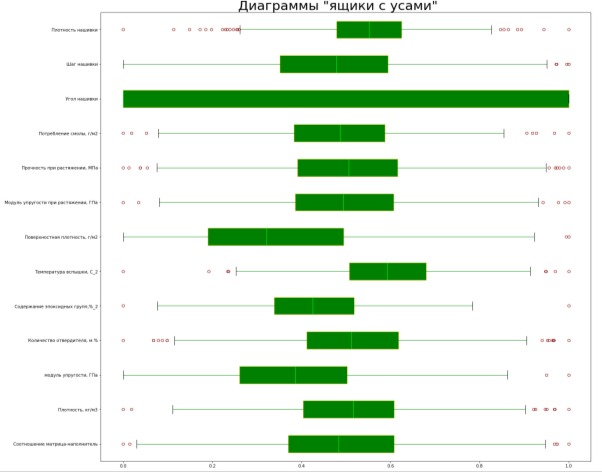


**«сырых» данных:**

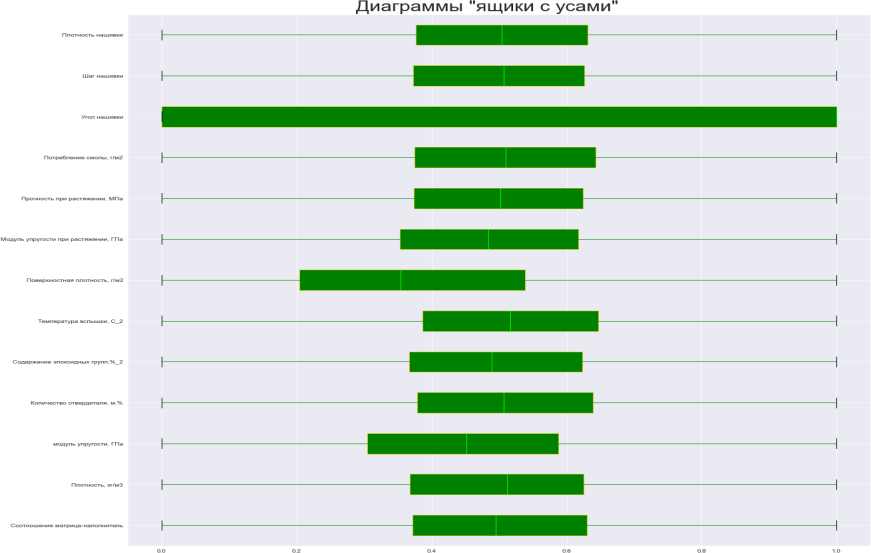
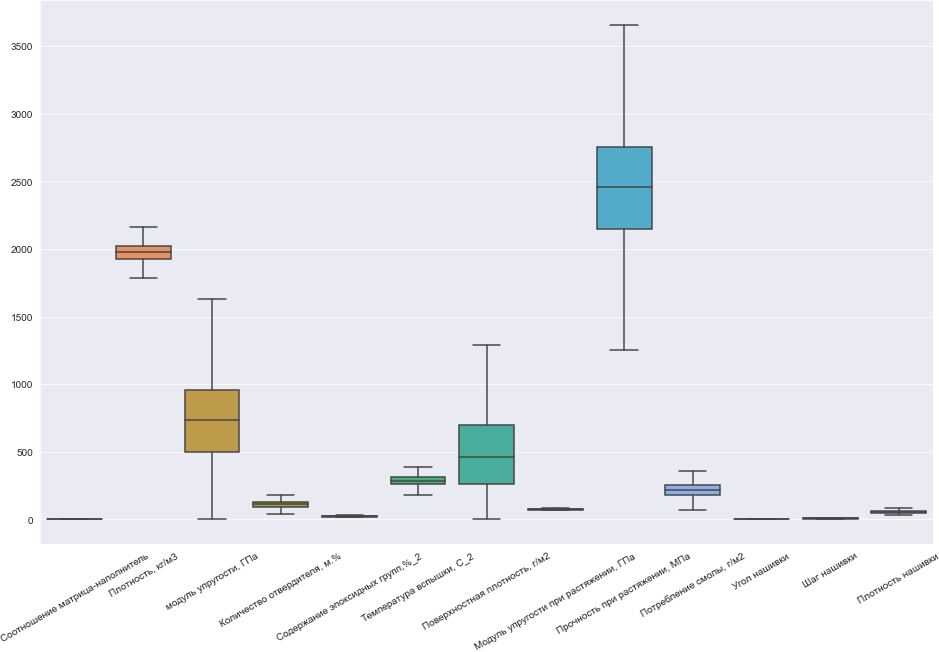
### Графики без нормализации и исключения шумов :

* Строим гистограммы распределения каждой из переменных (несколько вариантов);
* Строим диаграммы "ящиков с усами" (несколько вариантов);
* Строим попарные графики рассеяния точек

(несколько вариантов);

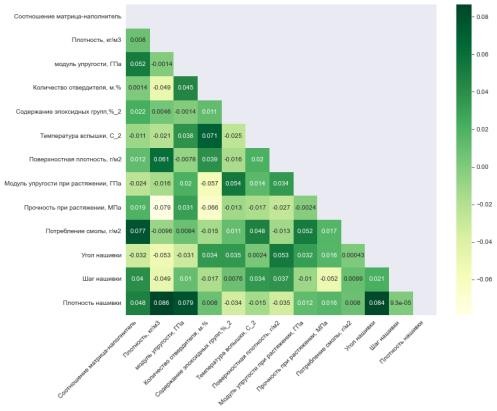
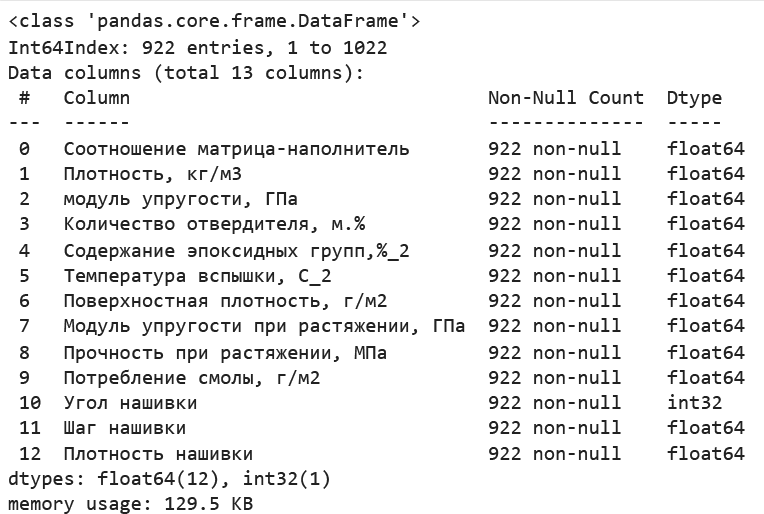
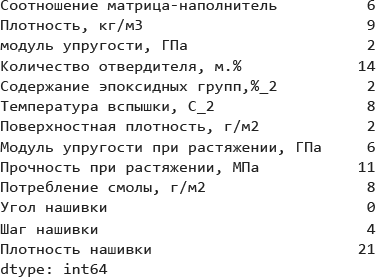


# Предобработка данных:



### Исключение выбросов:

* Исключаем выбросы методом межквартильного расстояния ;
* Проверяем результат;
* Строим графики;
* Убеждаемся, что выбросы ещё остались;
* Повторяем удаление выбросов ещё 4 раза до полного удаления;



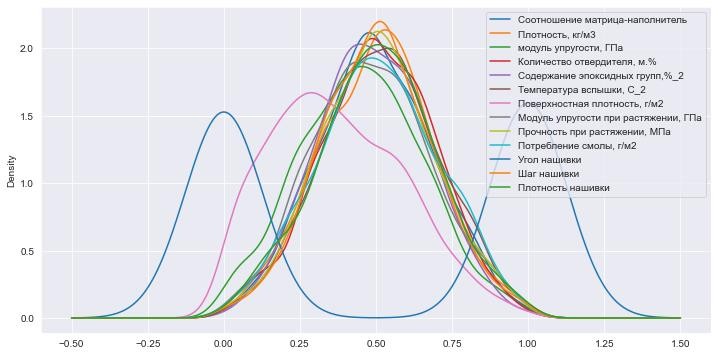
* Проверяем чистоту датасета от выбросов;
* Строим дополнительные возможные графики

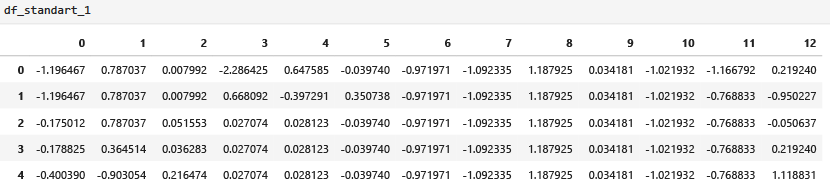
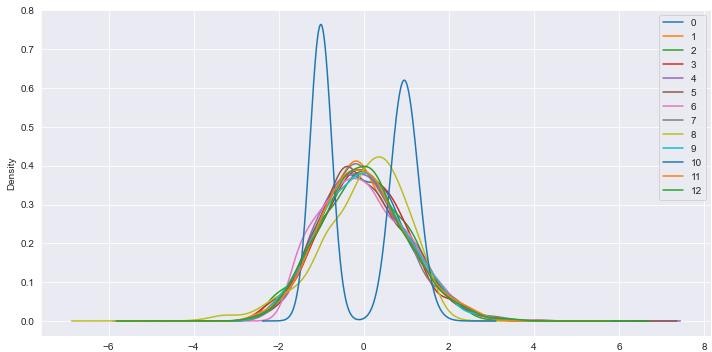
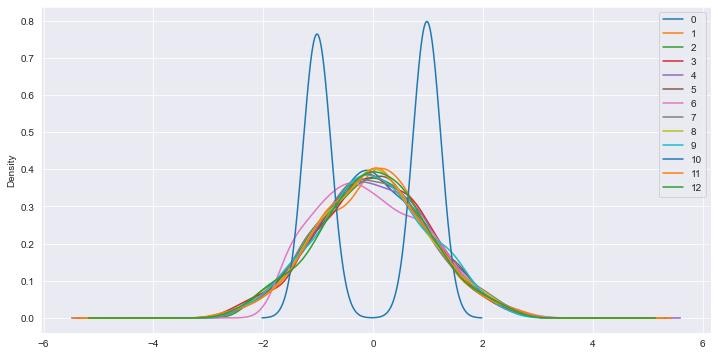
«чистого» датасета.

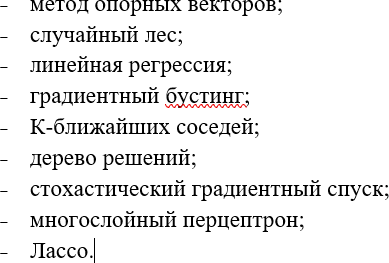
# Предобработка данных:

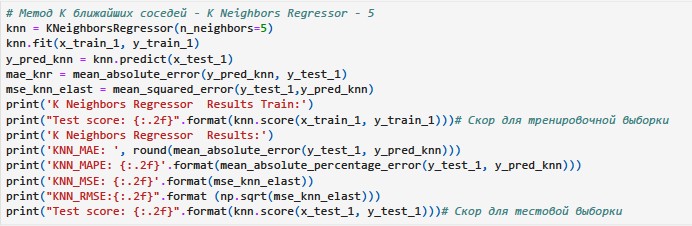
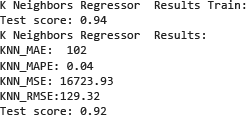
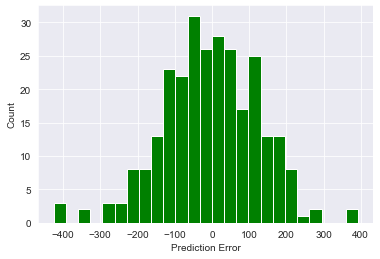
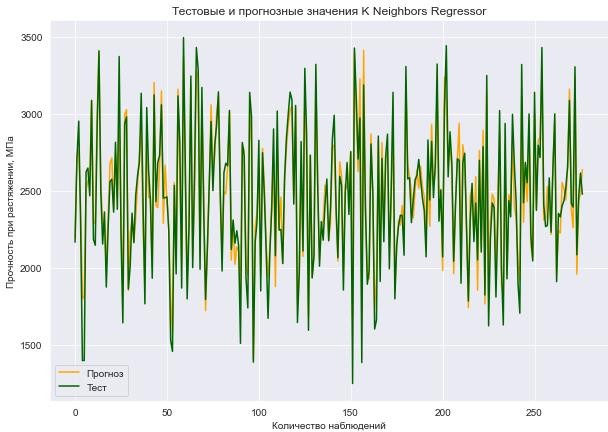


### Нормализация данных:

* Нормализуем данные MinMaxScaler();
* Строим график плотности ядра;
* Проверяем результат MinMaxScaler();
* Строим графики MinMaxScaler();
* Нормализуем данные с помощью Normalizer();
* Проверяем результат Normalizer();
* Строим графики Normalizer();



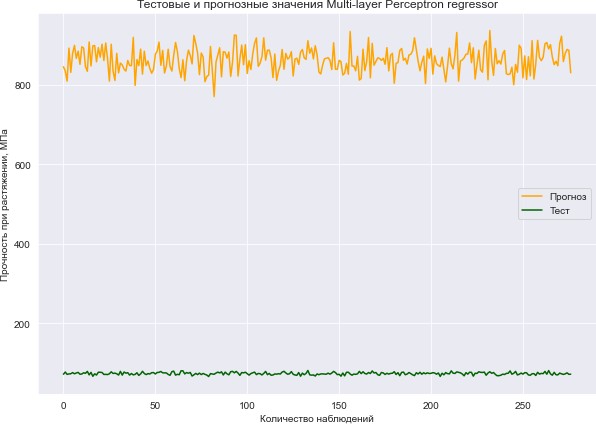
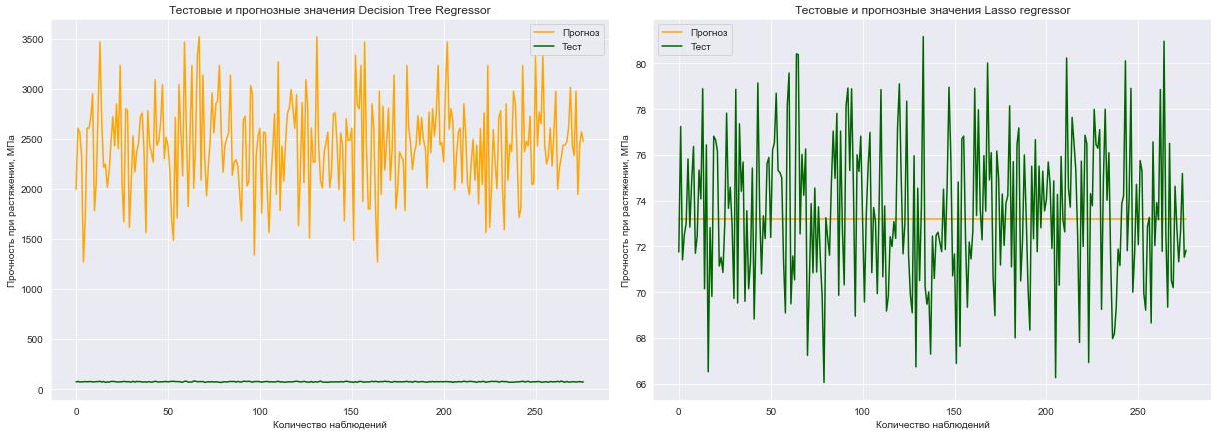
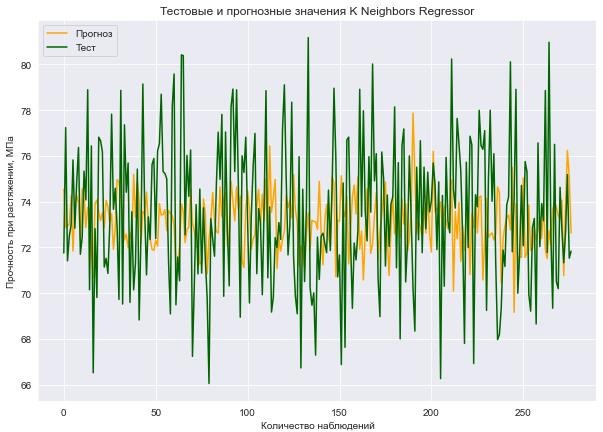
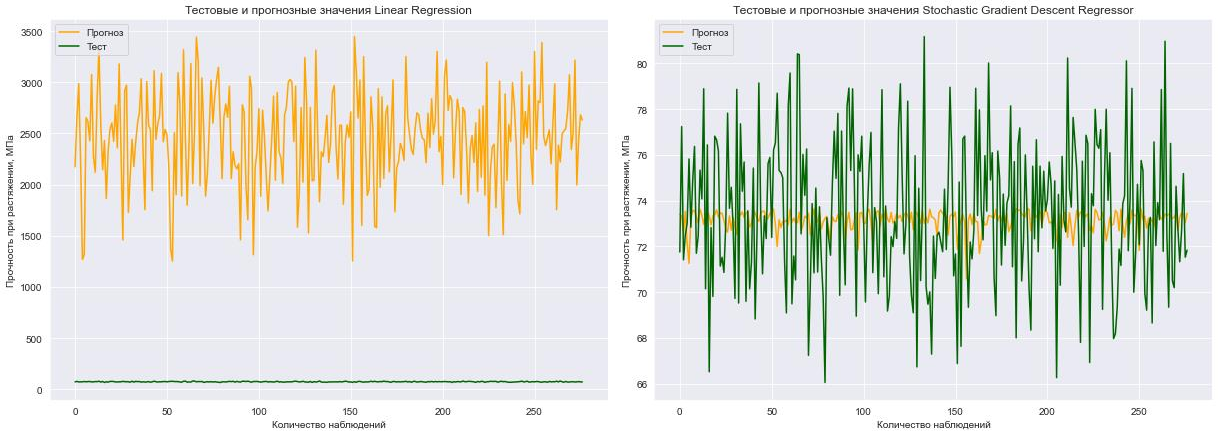
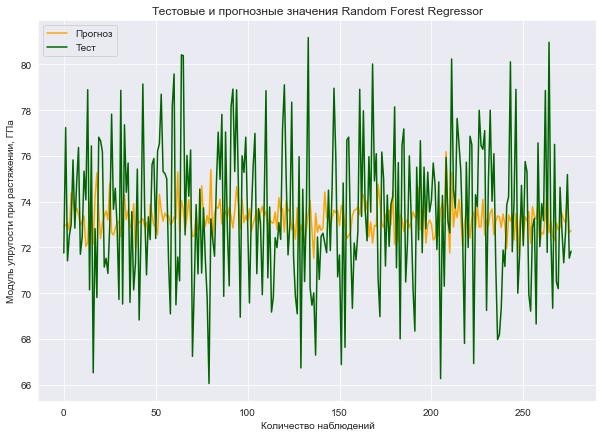
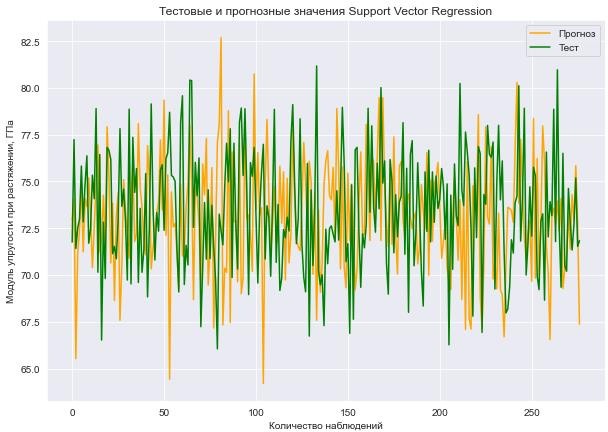
**Разработка и обучение моделей для прогноза прочности при растяжении:**



### Метод К ближайших соседей:

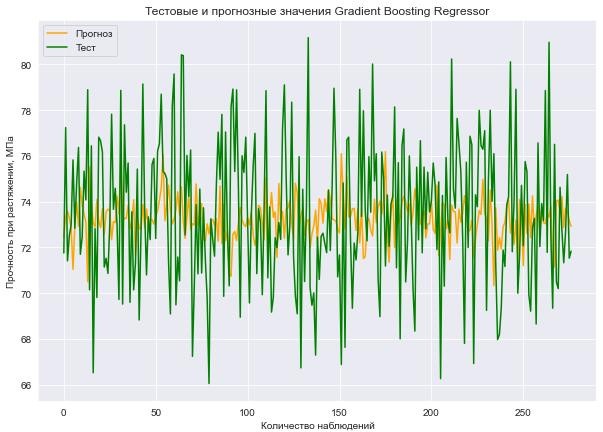
* Разбиваем данные на тестовую и тренировочную выборки;
* Обучаем модель;
* Вычисляем коэффициент детерминации;
* Считаем MAE, MAPE, MSE, RMSE, test score train и test score test;
* Сравниваем с результатами модели, выдающей среднее значение;
* Строим графики для тестовых и прогнозных значений;
* Строим гистограмму распределения ошибки

**Разработка и обучение моделей для прогноза модуль упругости при растяжении:**

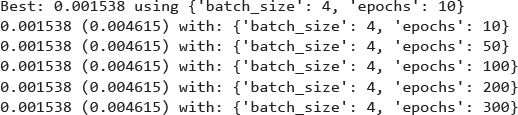
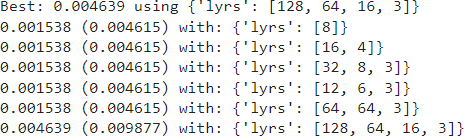
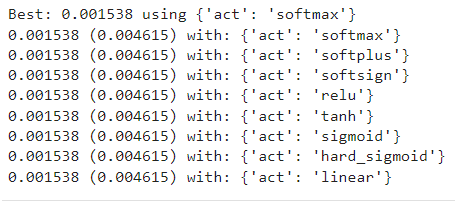
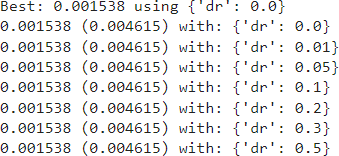
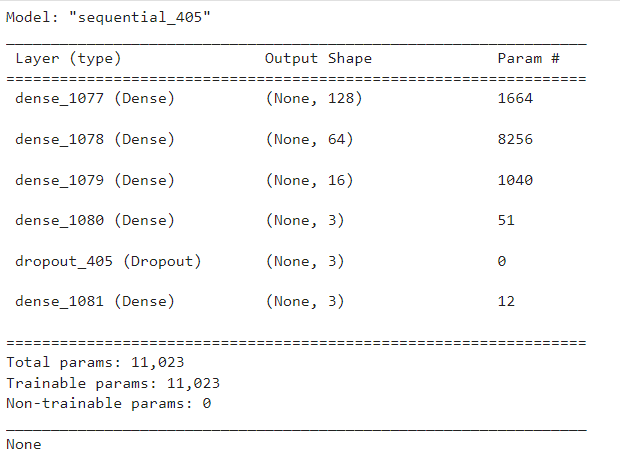
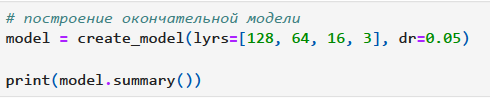
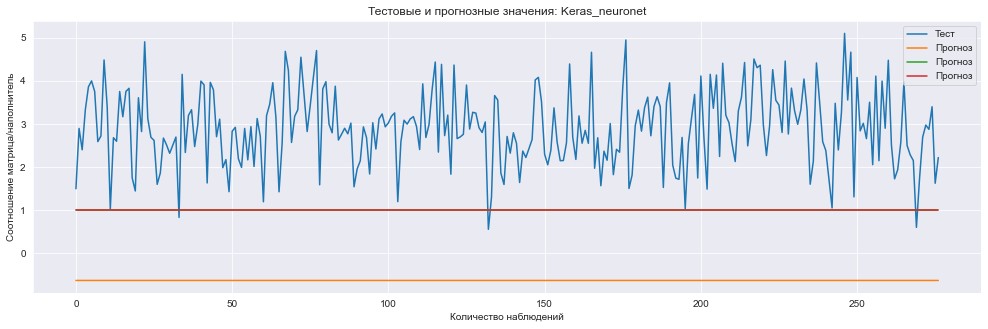
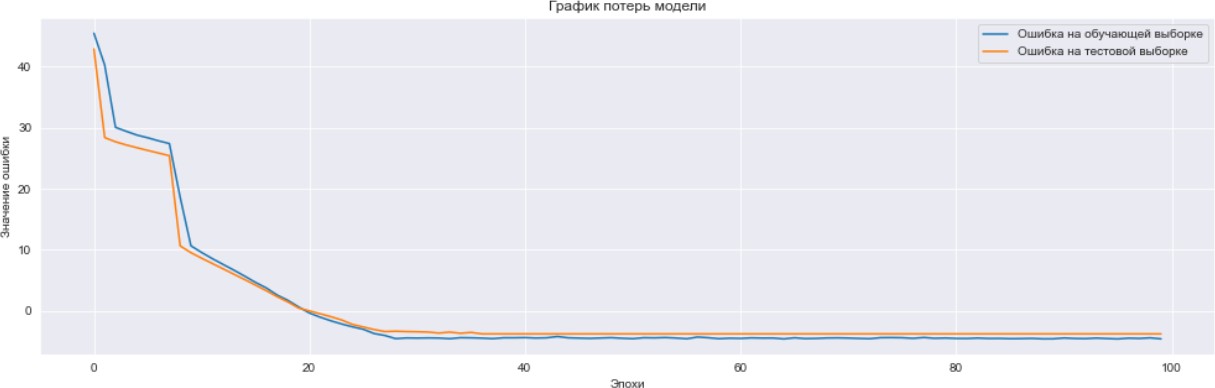


### Графики тестовых и прогнозных значений для разных методов

**(слева - направо и сверху – вниз):**

* Метод опорных векторов;
* Линейная регрессия;
* Стохастический градиентный спуск;
* Многослойный перцептрон;
* К-ближайших соседей;
* Градиентный бустинг;
* «Случайный лес»;
* Дерево принятия решений;
* Лассо.

**Нейронная сеть для соотношения**



**«матрица- наполнитель»:**

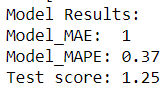
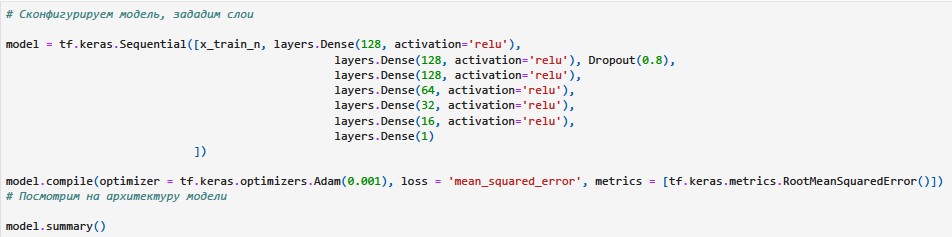
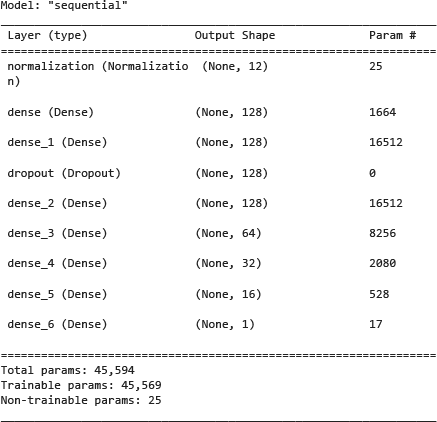
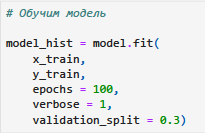
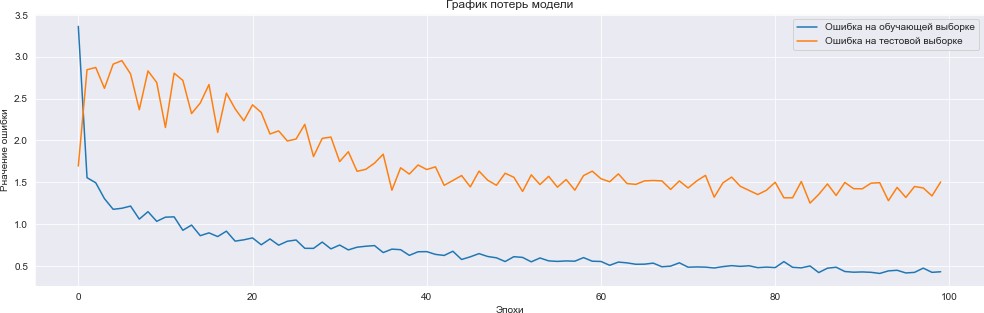
### Первая модель:

* Формируем входы и выход для модели;
* Разбиваем выборки на обучающую и

Тестовую;

* Нормализуем данные;
* Прописываем функцию для поиска наилучших параметров и слоёв;
* Обучаем нейросеть;
* Смотрим на потери модели;
* Строим график потерь на тренировочной и тестовой выборках.
* Строим график результата работы модели.

**Нейронная сеть для соотношения**



**«матрица- наполнитель»:**

### Вторая модель:

* Формируем входы и выход для модели;
* Разбиваем выборки на обучающую и тестовую;
* Нормализуем данные;
* Конфигурируем модель, задаем слои, смотрим на архитектуру модели;
* Обучаем модель;
* Строим график потерь на тренировочной и тестовой выборках;
* Строим график результата работы модели.

**Спасибо за**

**Спасибо за внимание!**