

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ  
РОССИЙСКОЙ ФЕДЕРАЦИИ  
Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«Московский государственный технический университет имени Н.Э.  
Баумана  
(национальный исследовательский университет)»

**ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА**  
**по курсу**  
**«Data Science»**

Слушатель

Орлянский Алексей Леонидович

Москва, 2022

# СОДЕРЖАНИЕ

Введение.....	3
1. Аналитическая часть.....	5
1.1. Постановка задачи.....	5
1.2. Подготовка датасета.....	7
1.3. Разведочный анализ данных. ....	7
2. Практическая часть.....	15
2.1. Предобработка данных. ....	15
2.2. Разработка и обучение моделей.....	18
2.3. Нейронная сеть.....	23
3. Выводы.....	24
Список использованной литературы .....	25

## **Введение**

Тема: Прогнозирование конечных свойств новых материалов (композиционных материалов).

Композиционный материал - искусственно созданный неоднородный сплошной материал, состоящий из двух или более компонентов с четкой границей раздела между ними. Засчет применения такой структуры достигаются характеристики материала, не уступающие или превышающие характеристики исходных однородных материалов.

Композиционный материал состоит, как правило, из армирующих элементов и матрицы, а его характеристики определяются как характеристиками исходных однородных материалов, так и прочностью связи между материалами, а также технологией совмещения этих компонентов и их структурой. Несмотря на наличие в композиционном материале нескольких однородных материалов, он является монолитным, и не может быть разделен на входящие в его состав материалы без разрушения общей конструкции композиционного

К примеру, использование композиционных материалов в броне позволяет значительно увеличить ее сопротивляемость удару засчет необходимости прохождения ударной волной через материалы с разными характеристиками и частичного отражения ударной волны в местах их соприкосновения, а их использование в лыжных палках позволяет их сделать не менее крепкими, чем металлические, но значительно снизить их вес.

В связи с высокими характеристиками композиционные материалы широко применяются не только при изготовлении узкоспециализированной продукции, но и в товарах и продукции, неразрывно присутствующих в нашей повседневной жизни.

Тем не менее, применение в композиционных материалах веществ, зачастую сильно отличающихся по своим свойствам, не позволяет заранее предсказать не только свойства итогового композиционного материала, но и наиболее оптимальные для повышения его характеристик соотношения структуры исходных компонентов, а также технологию их совмещения. Это приводит к необходимости проведения большого количества экспериментов и нахождения оптимального соотношения компонентов и технологии их совмещения эмпирическим путем.

Задача, исследуемая в данной работе, направлена на поиск способов прогнозирования свойств композиционного материала путем оценки свойств входящих в него компонентов, а также их структуры, соотношения и применяемых при их совмещении технологических процессов.

## **1. Аналитическая часть**

### **1.1. Постановка задачи**

На входе имеются данные о начальных свойствах компонентов композиционных материалов (количество связующего, наполнителя, температурный режим отверждения и т.д.). На выходе необходимо спрогнозировать ряд конечных свойств получаемых композиционных материалов.

Созданные прогнозные модели помогут сократить количество проводимых испытаний, а также пополнить базу данных материалов возможными новыми характеристиками материалов, и цифровыми двойниками новых композитов.

При постановке задачи на входе имеются два датасета с различными параметрами композиционного материала.

Всего в двух датасетах приведены данные о 13 характеристиках:

- соотношение матрица-наполнитель;
- плотность;
- модуль упругости;
- количество отвердителя;
- содержание эпоксидных групп;
- температура вспышки;
- поверхностная плотность;
- модуль упругости при растяжении;
- прочность при растяжении;
- потребление смолы;

- угол нашивки, град;
- шаг нашивки;
- плотность нашивки.

На основании исследования приведенных в датасете характеристик необходимо научиться прогнозировать следующие конечные свойства композиционного материала в зависимости от остальных его характеристик:

- модуль упругости при растяжении;
- модуль прочности при растяжении;
- соотношение матрица-наполнитель.

В первом датасете приведено 10 характеристик, всего 1023 строки, во втором датасете – 3 характеристики, всего 1040 строк.

По условиям задачи данные датасеты необходимо объединить в один по индексу с типом объединения INNER.

Для выполнения задачи необходимо провести разведочный анализ данных, их предобработку (удаление выбросов, пропусков, их нормализацию, и т.д.), обучить модели для прогноза модуля упругости при растяжении и прочности при растяжении, а также написать нейронную сеть, которая будет рекомендовать соотношение матрица-наполнитель.

Анализ данных, предобработка данных, построение моделей нейронной сети выполняются с использованием языка программирования Python.

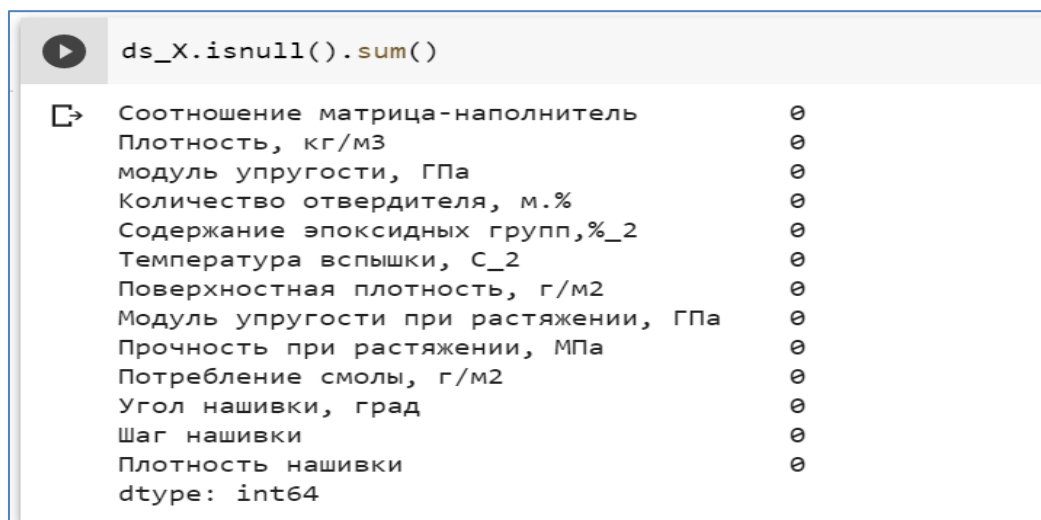
## 1.2. Исследование и подготовка датасета.

В целях подготовки датасета к его обработке производится объединение двух исходных датасетов  $X_{br}$  и  $X_{nir}$  в один.

Так как по условиям задачи объединение производится по индексам по типу INNER, итоговый датасет имеет 1023 строки. Строки датасета  $X_{nir}$  с 1024 по 1040 в итоговом датасете отсутствуют, тем не менее, в связи с небольшим количеством потерянных при объединении данных, данное уменьшение данных можно считать несущественным. В связи с тем, что отсутствует массив данных для строк 1024-1040 в датасете  $X_{br}$  (крайних строк датасета), и данные строки пропущены полностью, а не фрагментарно, аугментация данных в таком случае нецелесообразна и может привести к неверным результатам анализа общего массива данных и построения моделей.

## 1.3. Разведочный анализ данных.

В процессе разведочного анализа проверено наличие пропусков данных. Согласно полученным результатам, пропуски в итоговом датасете отсутствуют (Рисунок 1).



<code>ds_X.isnull().sum()</code>	
Соотношение матрица-наполнитель	0
Плотность, кг/м3	0
модуль упругости, ГПа	0
Количество отвердителя, м.%	0
Содержание эпоксидных групп,%_2	0
Температура вспышки, C_2	0
Поверхностная плотность, г/м2	0
Модуль упругости при растяжении, ГПа	0
Прочность при растяжении, МПа	0
Потребление смолы, г/м2	0
Угол нашивки, град	0
Шаг нашивки	0
Плотность нашивки	0
dtype: int64	

Рисунок 1. Поиск пропусков данных.

Далее для каждой из колонок датасета получено среднее и медианное значение (Рисунок 2).

```
[36] #Для каждой колонки получаем среднее значение#
ds_X.mean(axis=0)
```

Соотношение матрица-наполнитель	2.930366
Плотность, кг/м3	1975.734888
модуль упругости, ГПа	739.923233
Количество отвердителя, м.%	110.570769
Содержание эпоксидных групп,%_2	22.244390
Температура вспышки, C_2	285.882151
Поверхностная плотность, г/м2	482.731833
Модуль упругости при растяжении, ГПа	73.328571
Прочность при растяжении, МПа	2466.922843
Потребление смолы, г/м2	218.423144
Угол нашивки, град	44.252199
Шаг нашивки	6.899222
Плотность нашивки	57.153929
dtype: float64	

```
[37] #Для каждой колонки получаем медианное значение#
ds_X.median(axis=0)
```

Соотношение матрица-наполнитель	2.906878
Плотность, кг/м3	1977.621657
модуль упругости, ГПа	739.664328
Количество отвердителя, м.%	110.564840
Содержание эпоксидных групп,%_2	22.230744
Температура вспышки, C_2	285.896812
Поверхностная плотность, г/м2	451.864365
Модуль упругости при растяжении, ГПа	73.268805
Прочность при растяжении, МПа	2459.524526
Потребление смолы, г/м2	219.198882
Угол нашивки, град	0.000000
Шаг нашивки	6.916144
Плотность нашивки	57.341920
dtype: float64	

Рисунок 2. Среднее и медианное значение для каждой характеристики

Получена информация об отсутствии пропусков данных в датасете, а также о большом разбросе значений характеристик в датасете.

Далее в процессе исследования датасета получены гистограммы распределения для каждой из характеристик датасета (Рисунок 3).



В целях визуализации наличия выбросов также получены диаграммы «ящик с усами» для каждой из характеристик (Рисунок 4).

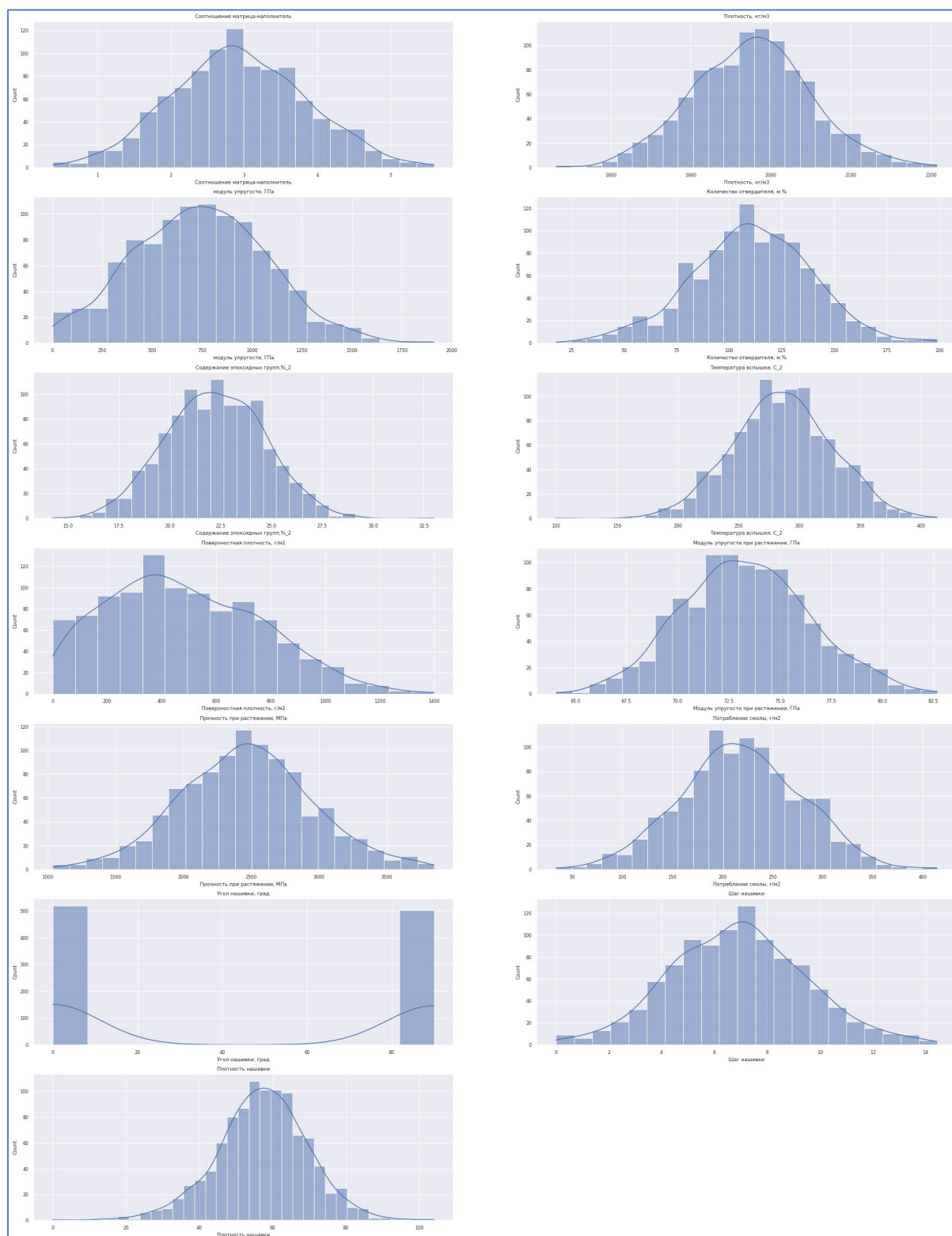
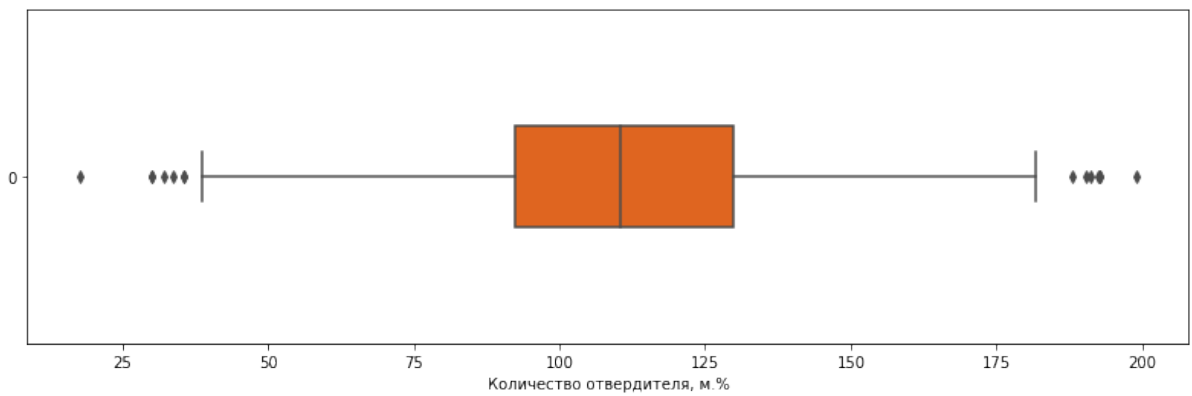
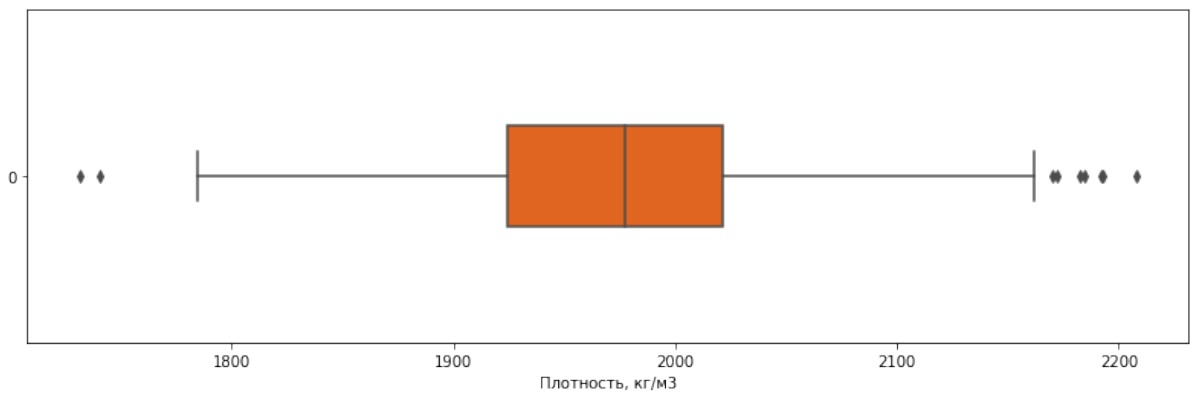
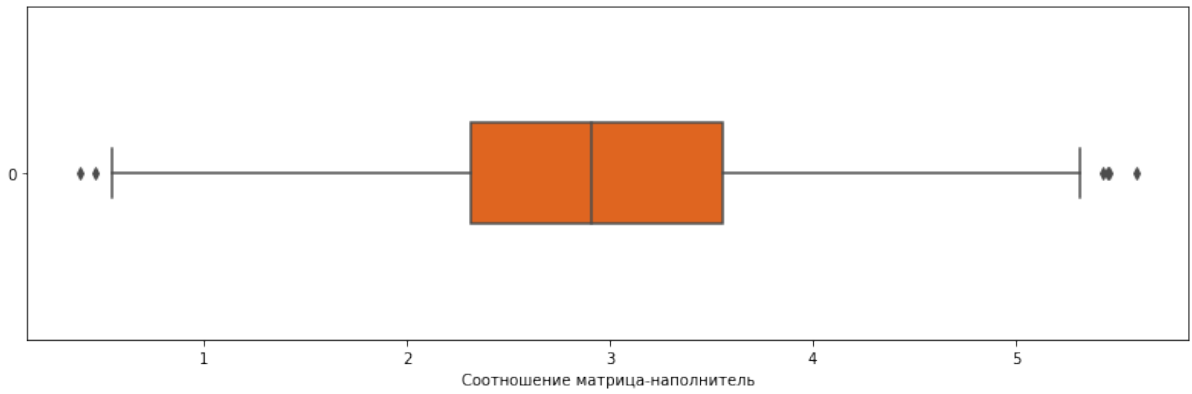
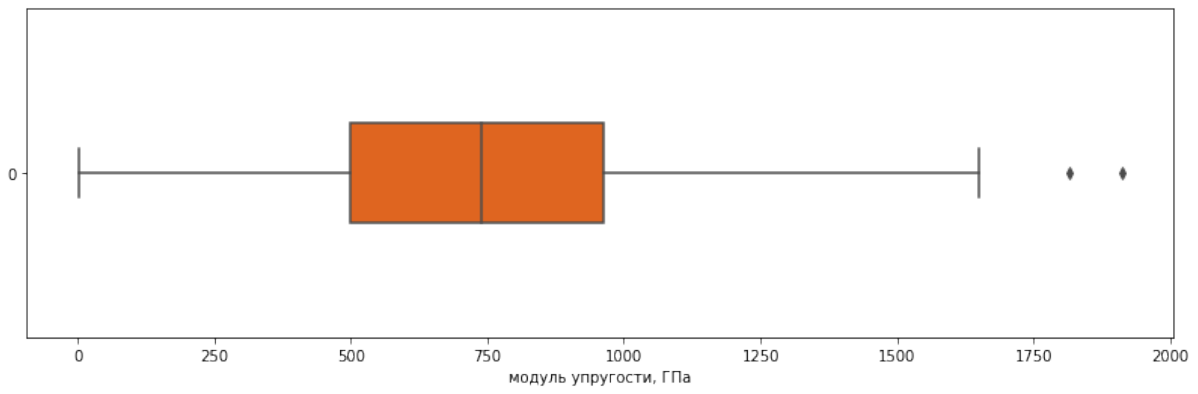
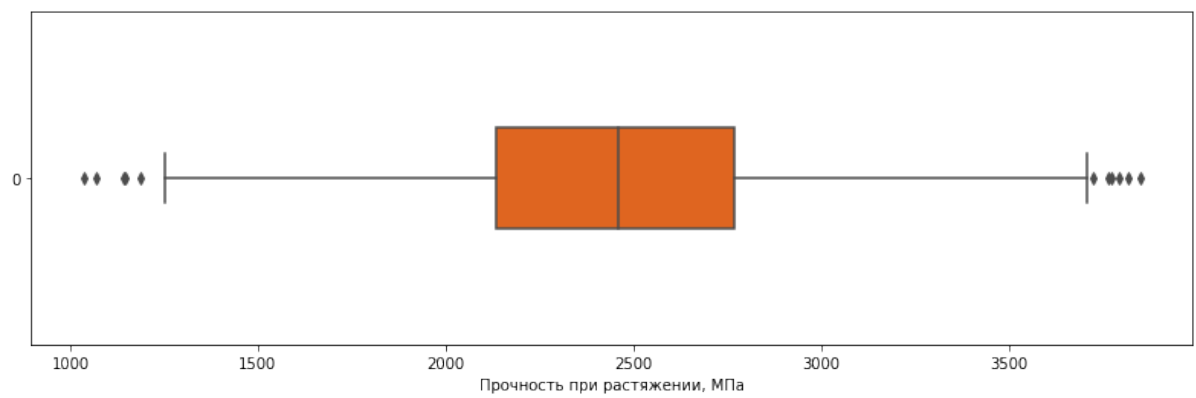
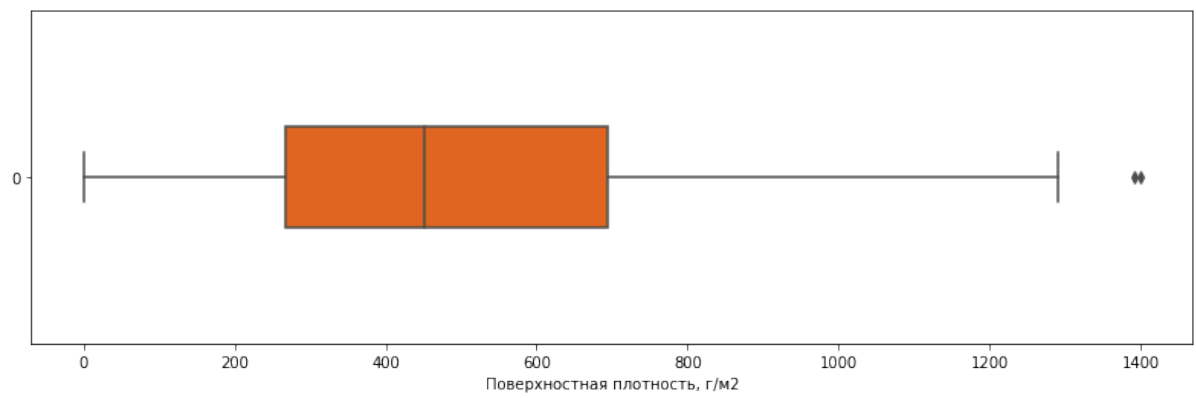
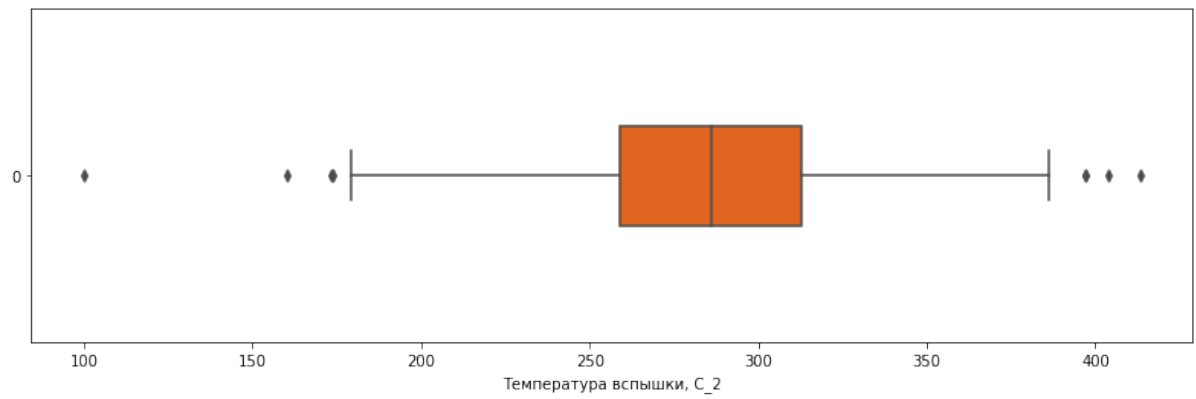
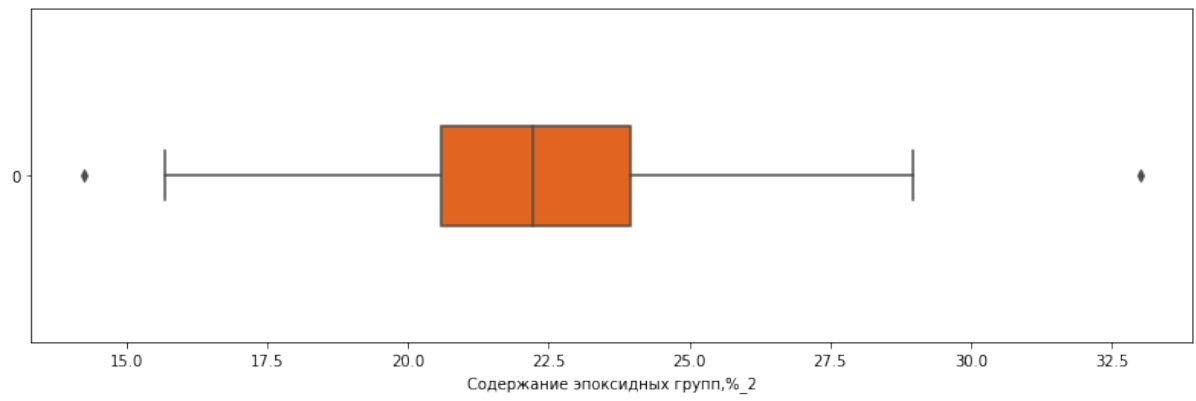
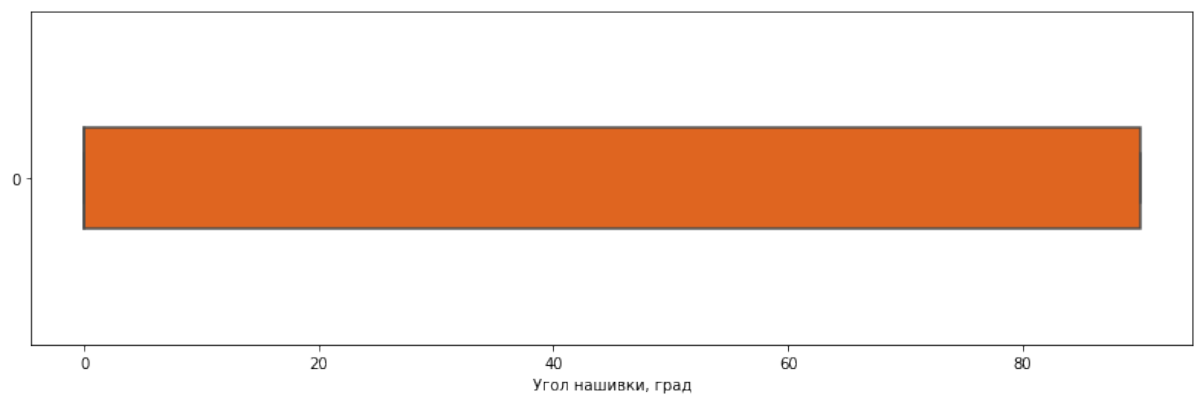
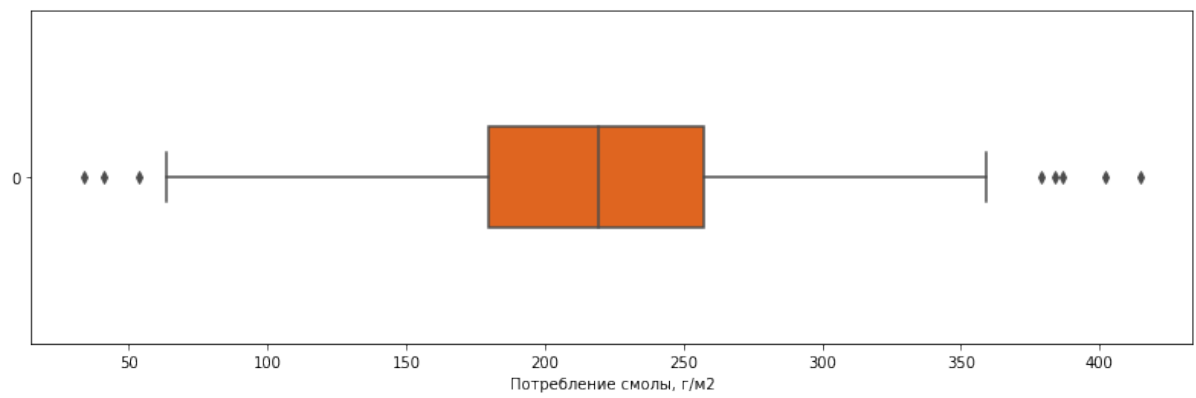
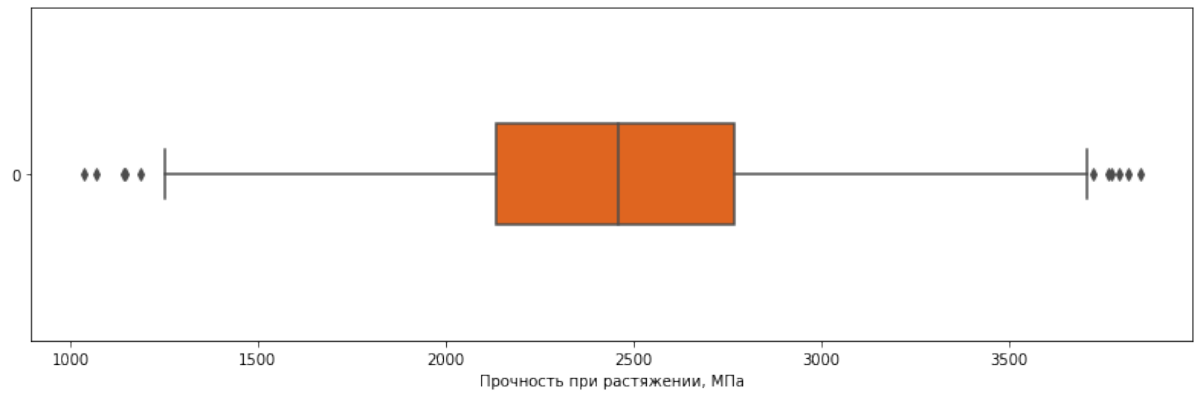
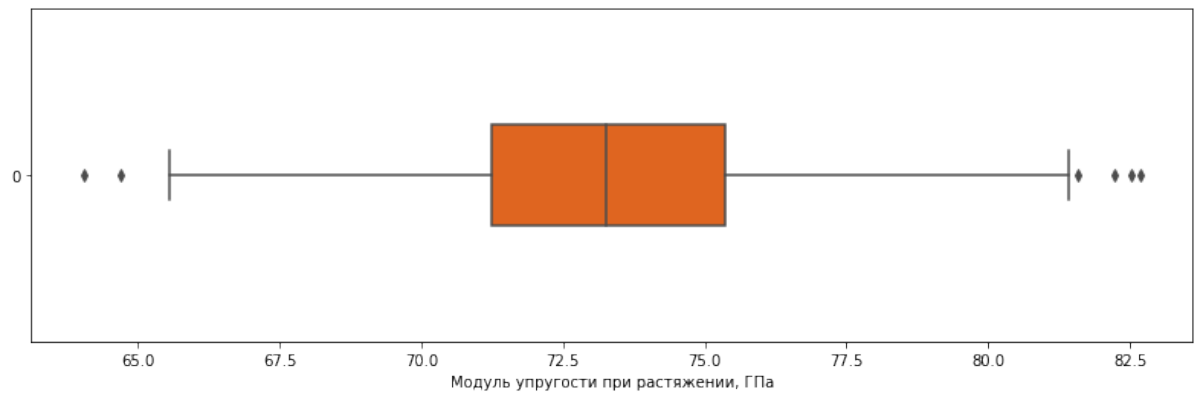


Рисунок 3. Гистограммы распределения.







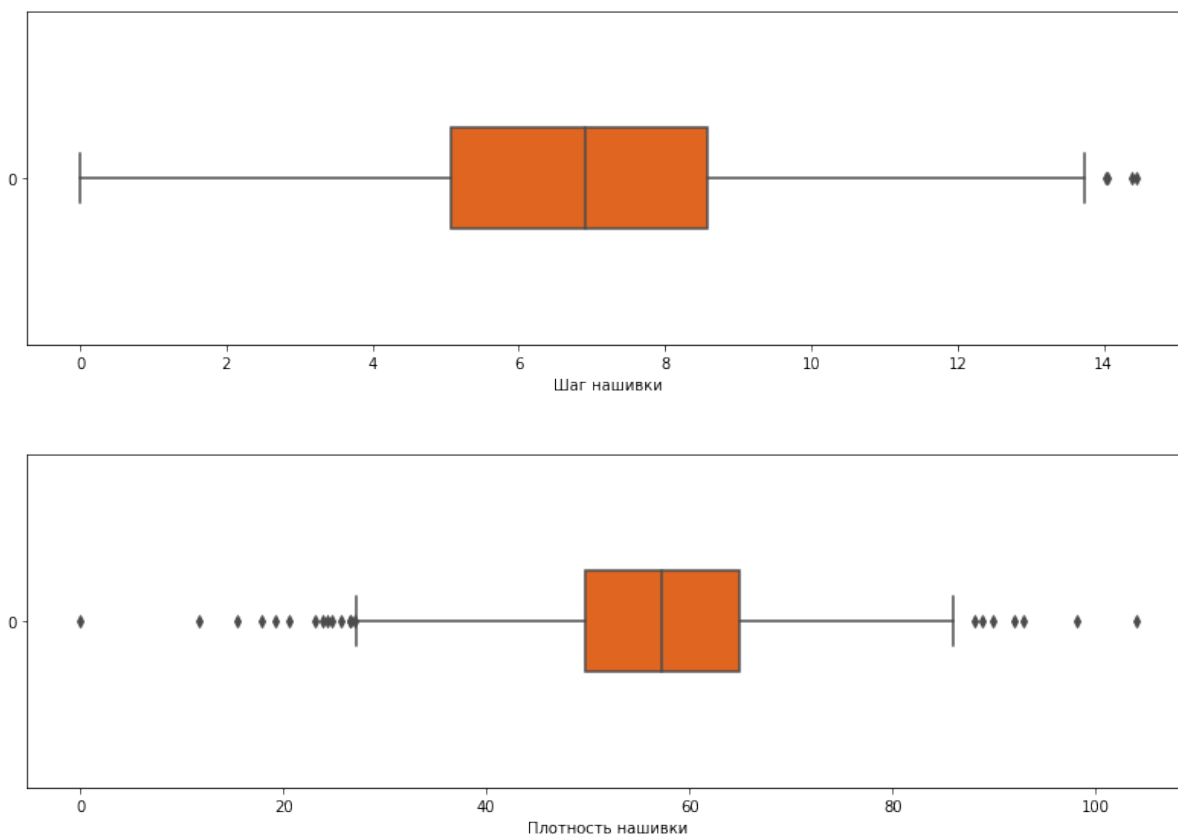


Рисунок 4. Диаграммы «ящик с усами»

Из полученных диаграмм «ящик с усами» следует, что выбросов нет только в характеристике «угол нашивки», что закономерно, так как данный параметр в датасете принимает дискретные значения – 0 либо 90 градусов.

Далее были получены попарные графики рассеяния точек для значений характеристик в датасете (Рисунок 5), а также построены тепловая карта для исследования зависимостей между значениями характеристик, имеющихся в датасете (Рисунок 6).

Согласно данным, полученным из построенных графиков, между значениями характеристик в датасете отсутствуют явные зависимости.

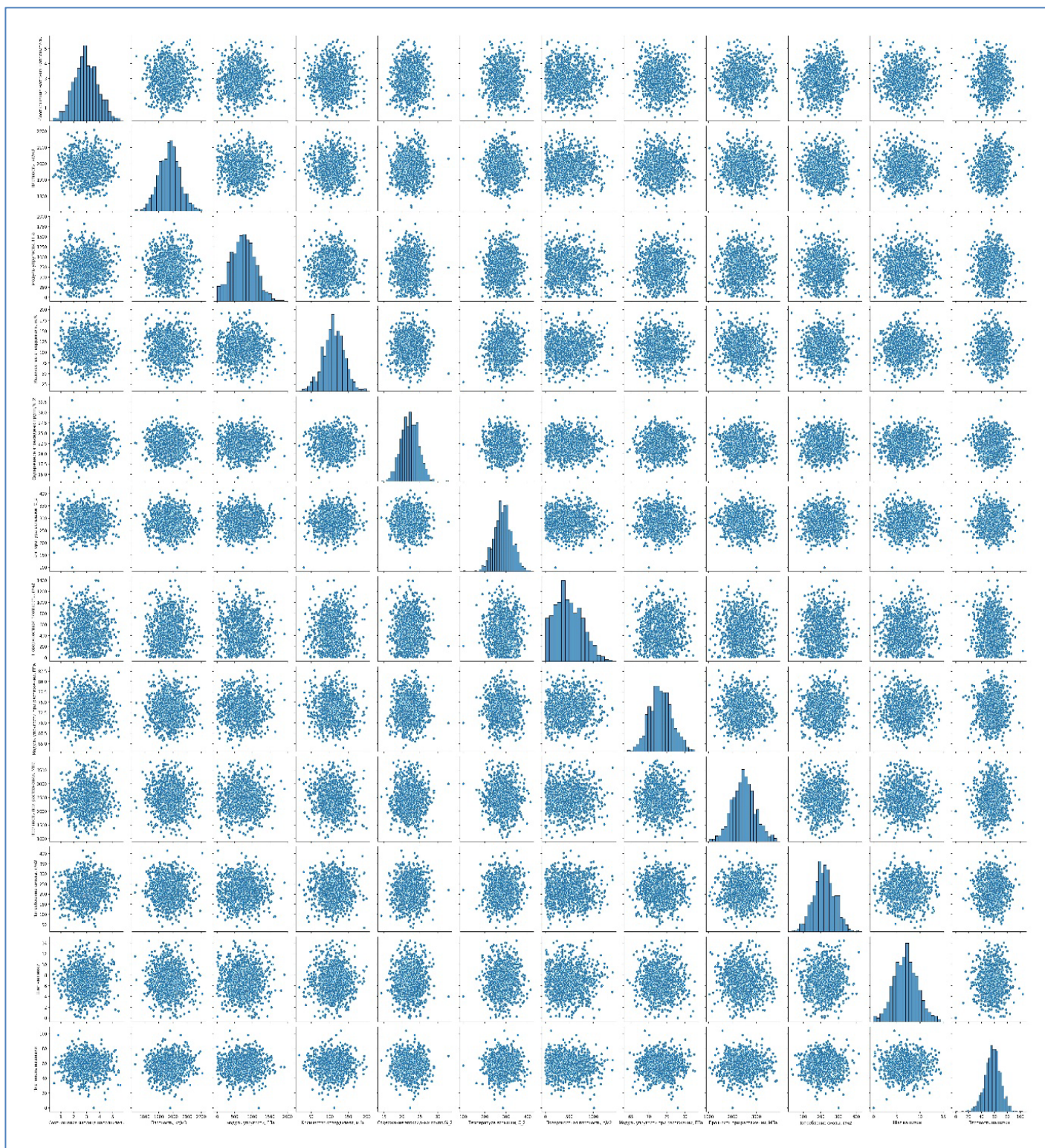


Рисунок 5. Попарные графики рассеяния точек.



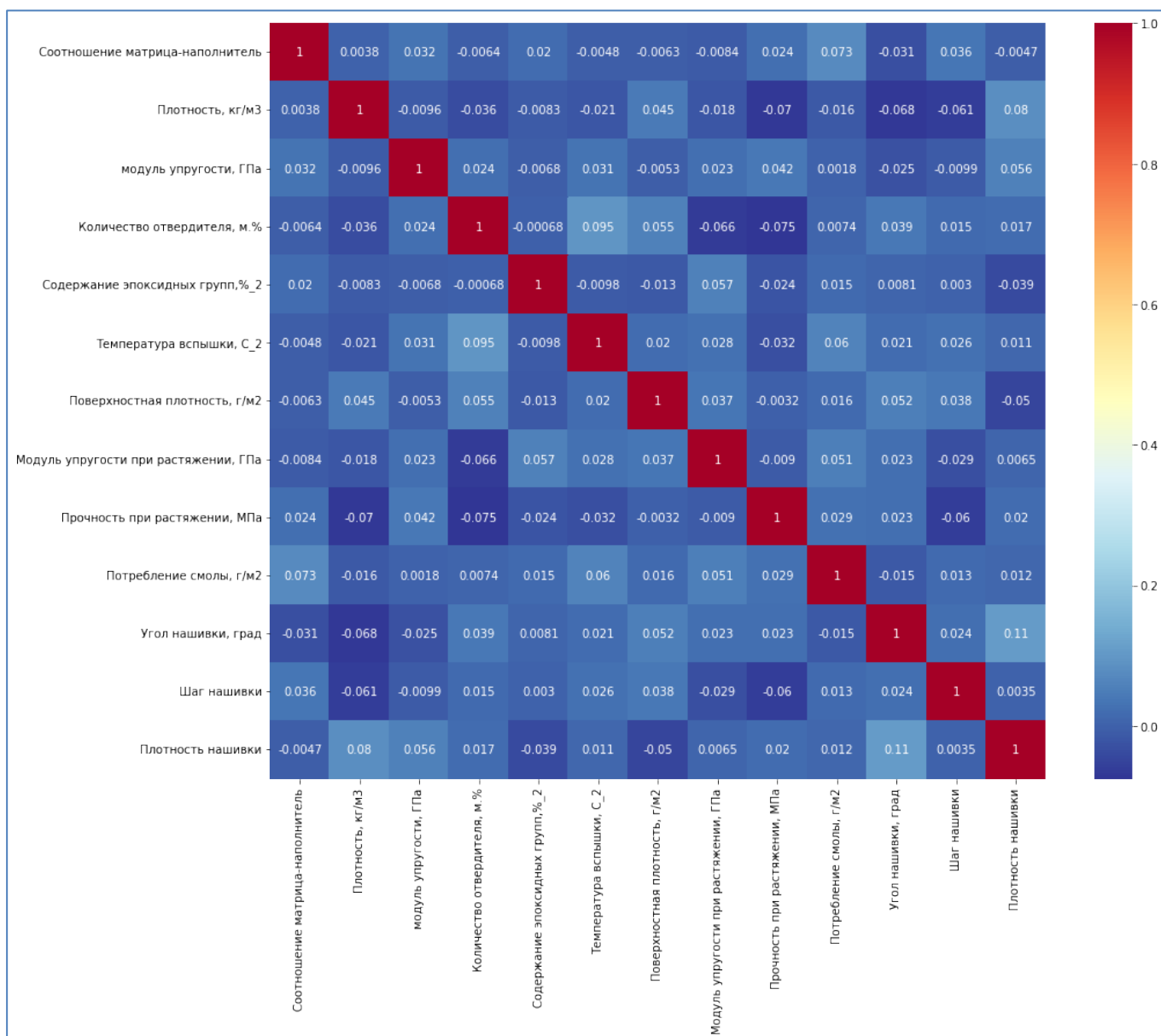


Рисунок 6. Тепловая карта зависимости между характеристиками

## 2. Практическая часть.

### 2.1. Предобработка данных.

В целях проведения предобработки данных датасета произведена оценка выбросов с применением метода 3-х сигм и их удаление.

Согласно оценке по методу 3-х сигм, количество выбросов в датасете несущественное – общее число выбросов 24.

количество выбросов в параметре Соотношение матрица-наполнитель - 0  
 количество выбросов в параметре Плотность, кг/м3 - 3  
 количество выбросов в параметре модуль упругости, ГПа - 2  
 количество выбросов в параметре Количество отвердителя, м.% - 2  
 количество выбросов в параметре Содержание эпоксидных групп,%\_2 - 2  
 количество выбросов в параметре Температура вспышки, С\_2 - 3  
 количество выбросов в параметре Поверхностная плотность, г/м2 - 2  
 количество выбросов в параметре Модуль упругости при растяжении, ГПа - 0  
 количество выбросов в параметре Прочность при растяжении, МПа - 0  
 количество выбросов в параметре Потребление смолы, г/м2 - 3  
 количество выбросов в параметре Угол нашивки, град - 0  
 количество выбросов в параметре Шаг нашивки - 0  
 количество выбросов в параметре Плотность нашивки - 7  
 Всего выбросов 24

Рисунок 7. Поиск выбросов в датасете.

Был проведен поиск и удаление выбросов в датасете.

Индекс	Соотношение матрица-наполнитель	Плотность, кг/м3	модуль упругости, ГПа	Количество отвердителя, м.%	Содержание эпоксидных групп
1	1.857143	2030.000000	738.736842	50.000000	23.75
3	1.857143	2030.000000	738.736842	129.000000	21.25
4	2.771331	2030.000000	753.000000	111.860000	22.26
5	2.767918	2000.000000	748.000000	111.860000	22.26
6	2.569620	1910.000000	807.000000	111.860000	22.26
...	...	...	...	...	...
1018	2.271346	1952.087902	912.855545	86.992183	20.12
1019	3.444022	2050.089171	444.732634	145.981978	19.59
1020	3.280604	1972.372865	416.836524	110.533477	23.95
1021	3.705351	2066.799773	741.475517	141.397963	19.24
1022	3.808020	1890.413468	417.316232	129.183416	27.47

999 rows × 6 columns

Рисунок 8. Датасет после удаления выбросов.



После нахождения и удаления строк датасета с выбросами общее количество строк датасета составило 999, таким образом, существенной потери данных при очистке от выбросов путем удаления строк датасета не произошло.

Также в процессе предобработки данных был построен график распределения для каждой из характеристик в датасете, что наглядно показало большой разброс значений в столбцах датасета (Рисунок 9).

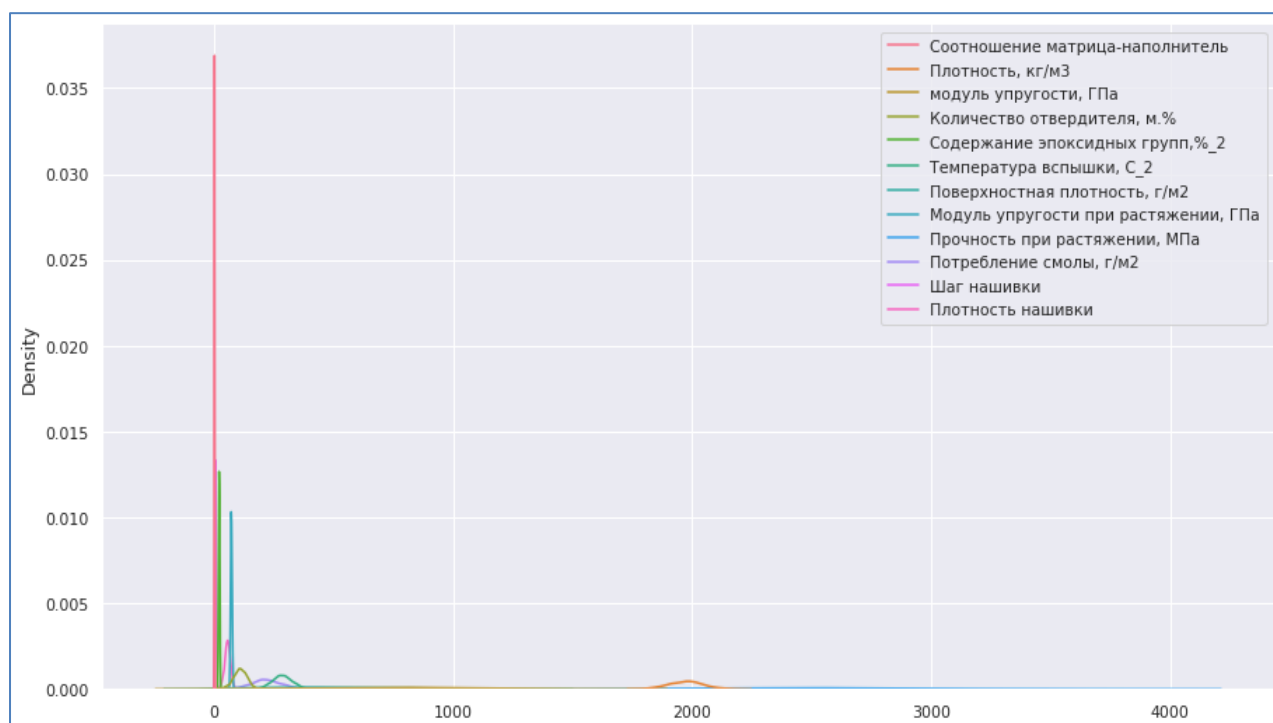


Рисунок 9. График распределения характеристик до нормализации.

В целях обеспечения корректной обработки алгоритмами имеющегося датасета необходимо исключить влияние разницы в значениях величины различных характеристик на работу моделей обработки данных, для этого проведена нормализация датасета.

После проведения нормализации датасета для сравнения был повторно построен график распределения для каждой из характеристик (Рисунок 10).

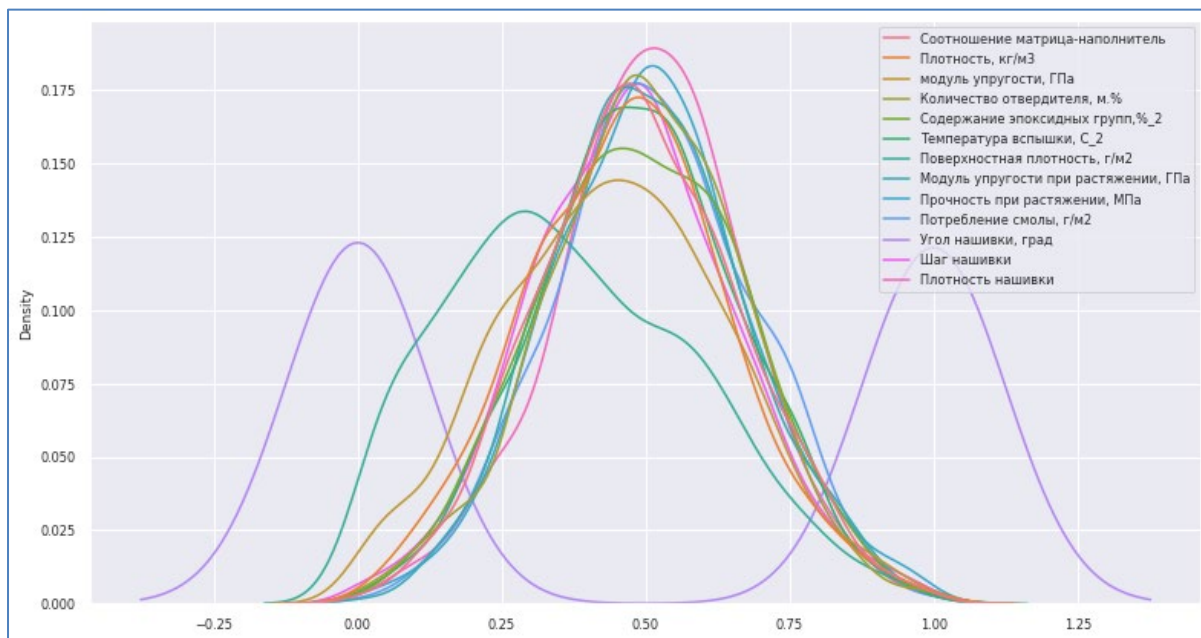


Рисунок 10. График распределения характеристик после нормализации.

## 2.2. Разработка и обучение моделей.

В целях решения поставленной задачи необходимо разработать и обучить несколько моделей для прогноза модуля упругости при растяжении и прочности при растяжении. При построении моделей 30% данных оставлено на тестирование модели, 70% данных использовано при обучении модели.

Для получения прогнозов требуемых характеристик были использованы линейная регрессия, метод к-ближайших соседей и случайный лес.

Проведена оценка работы моделей для каждой из характеристик.

Характеристики моделей приведены далее:

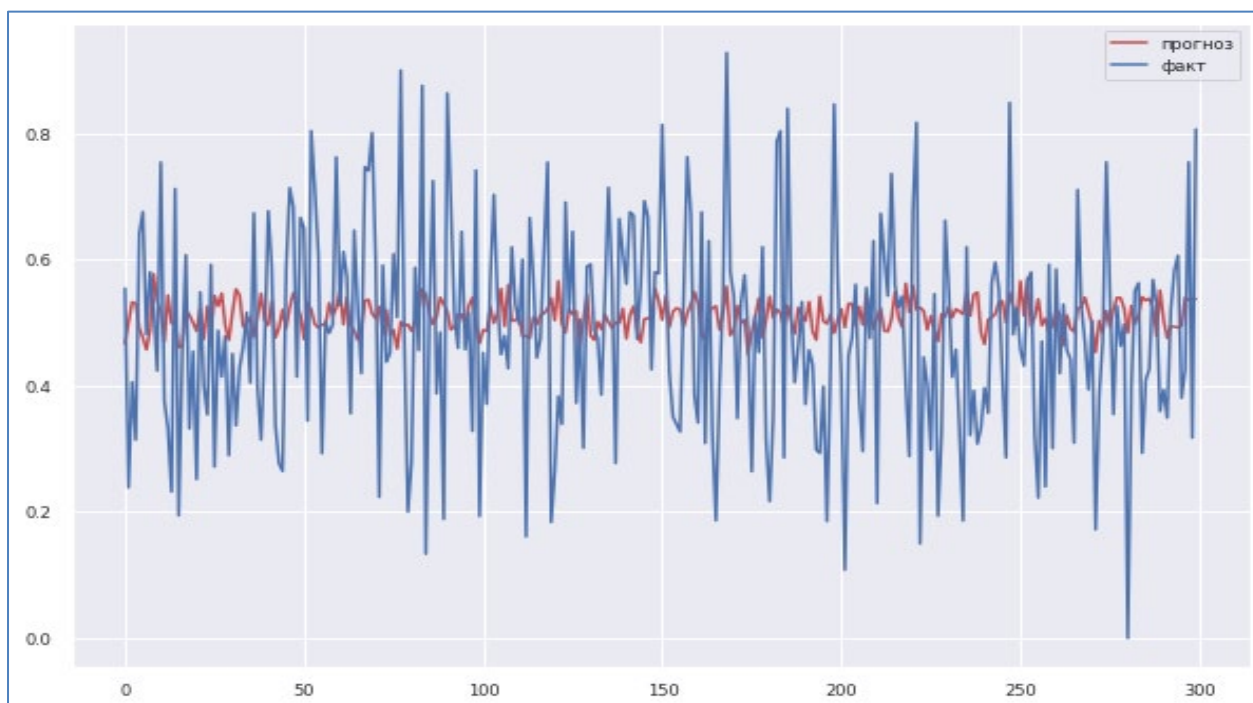


Рисунок 10. Оценка модели линейной регрессии для прогноза модуля упругости.



Рисунок 11. Оценка модели k-ближайших соседей для прогноза модуля упругости.

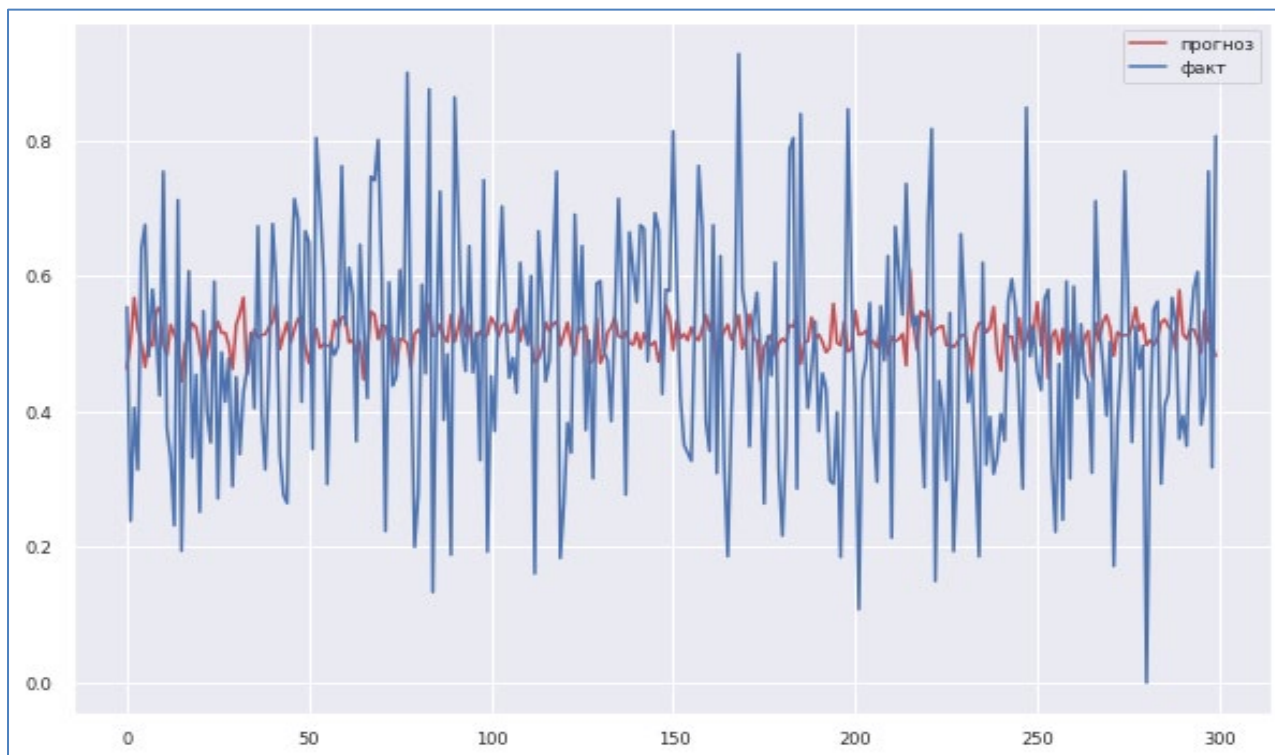


Рисунок 12. Оценка модели случайный лес для прогноза модуля упругости.

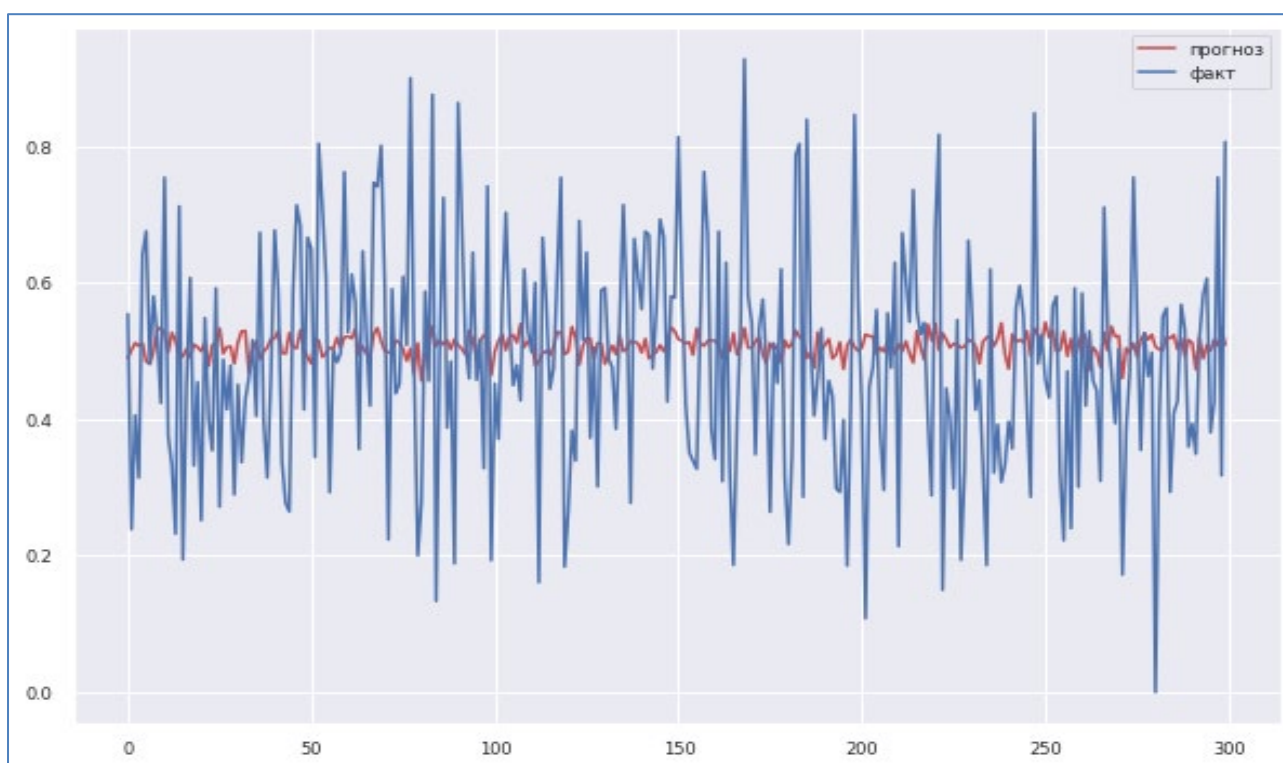


Рисунок 13. Оценка модели случайный лес для прогноза модуля упругости после подбора параметров.

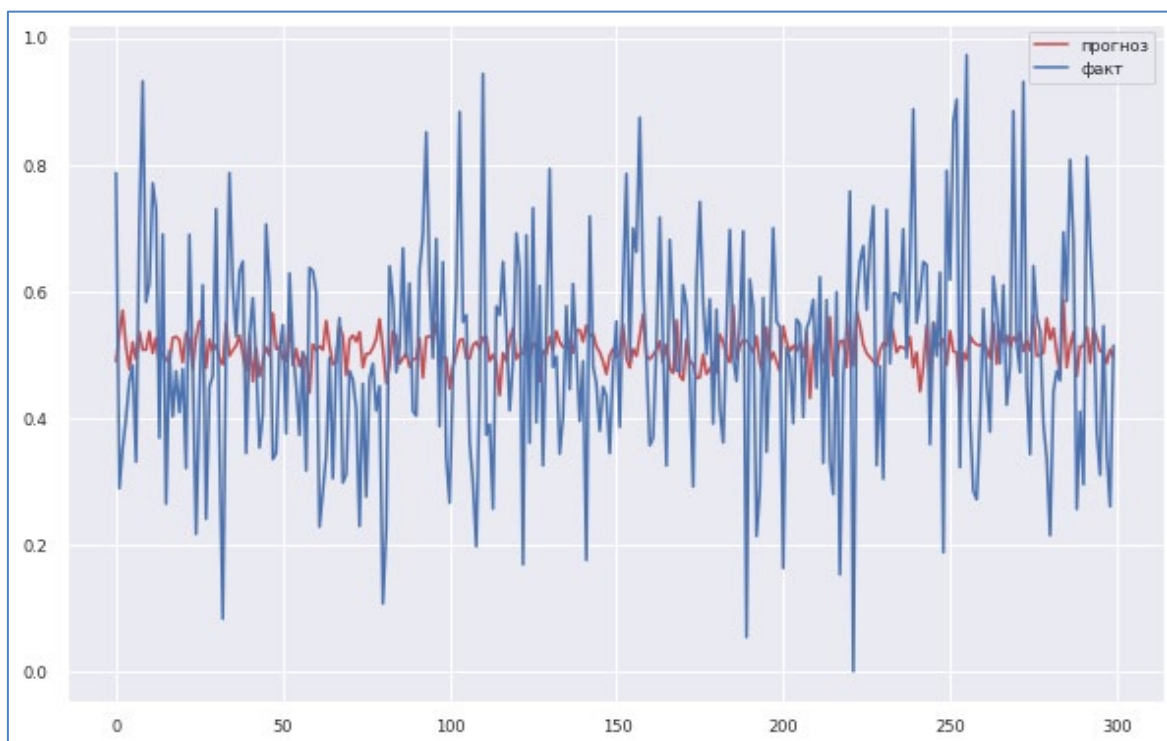


Рисунок 14. Оценка модели линейной регрессии для прогноза прочности при растяжении.

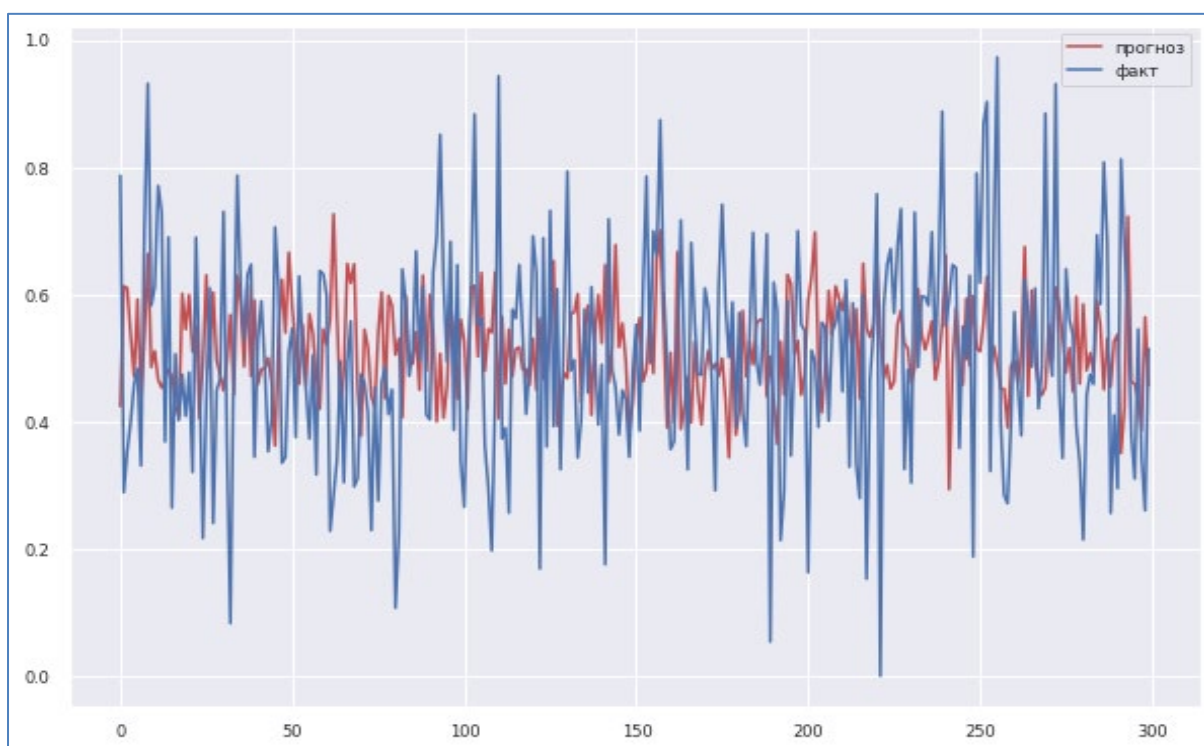


Рисунок 15. Оценка модели k-ближайших соседей для прогноза прочности при растяжении.





Рисунок 16. Оценка модели случайный лес для прогноза  
прочности при растяжении.

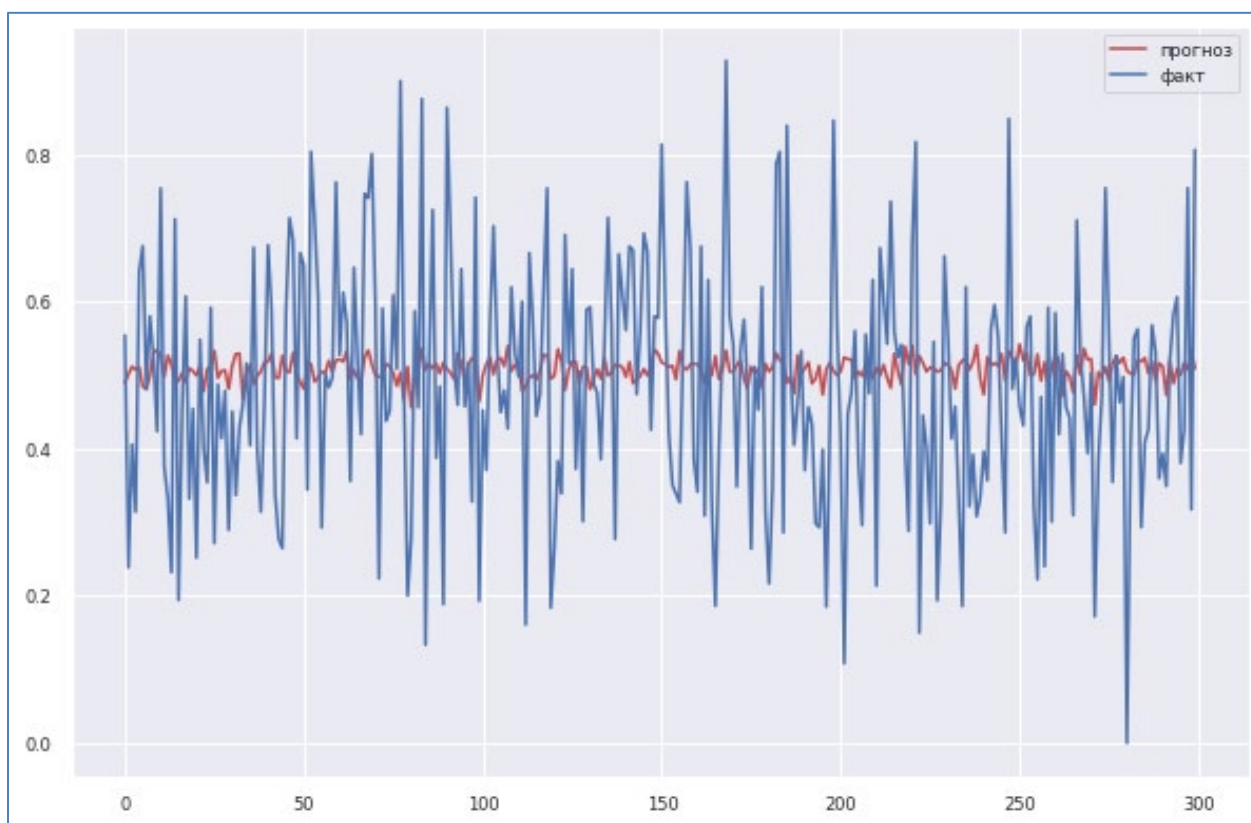


Рисунок 17. Оценка модели случайный лес для прогноза прочности  
при растяжении после подбора параметров.

Как видно из оценки построенных моделей, предсказать значения модуля упругости при растяжении и прочности при растяжении в зависимости от остальных характеристик не удалось.

### 2.3. Нейронная сеть.

Для рекомендации соотношения матрица-наполнитель была разработана нейронная сеть следующей архитектуры:

Model: "sequential"

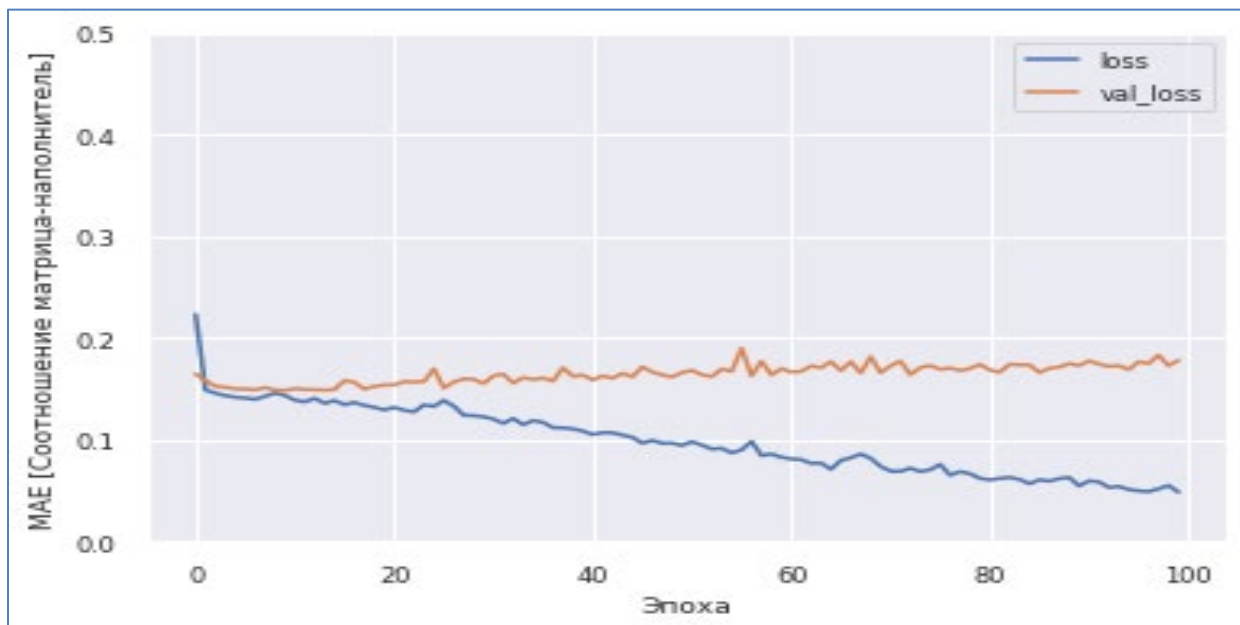
Layer (type)	Output Shape	Param #
dense (Dense)	(None, 128)	1664
dense_1 (Dense)	(None, 128)	16512
dense_2 (Dense)	(None, 64)	8256
dense_3 (Dense)	(None, 64)	4160
dense_4 (Dense)	(None, 1)	65

=====  
Total params: 30,657

Trainable params: 30,657

Non-trainable params: 0  
=====

Указанная сеть имеет 4 скрытых уровня с уменьшением числа нейронов к последнему уровню для сжатия предыдущей обработанной информации.



Построенная нейронная сеть не выявила закономерностей между соотношением матрица-наполнитель и остальными характеристиками, включенными в датасет.

### 3. Выводы.

В процессе выполнения выпускной квалификационной работы не найдены алгоритмы, позволяющие выявить зависимость величины модуля упругости при растяжении и величины прочности при растяжении от остальных характеристик на имеющемся наборе данных. Разработанная нейронная сеть также не позволяет предсказать соотношение матрица-наполнитель от входных параметров, имеющихся в датасете.

Материалы к данной ВКР размещены на GitHub

<https://github.com/ALOrlyanskii/BMSTU/tree/main>



## Список использованной литературы

1. Информация о композиционных материалах: – Режим доступа: <https://dic.academic.ru/dic.nsf/ruwiki/7078>
2. Документация по библиотеке pandas: – Режим доступа: <https://pandas.pydata.org/docs/index.html>
3. Документация по библиотеке matplotlib: – Режим доступа: <https://matplotlib.org/stable/index.html>
4. Документация по библиотеке seaborn: – Режим доступа: <https://seaborn.pydata.org/index.html>
5. Документация по библиотеке sklearn: – Режим доступа: <https://scikit-learn.org/stable/index.html>
6. Документация по библиотеке keras: – Режим доступа: <https://ru-keras.com/home/>
7. Seaborn Heatmaps: 13 способов настроить визуализацию матрицы корреляции: – Режим доступа: <https://datastart.ru/blog/read/seaborn-heatmaps-13-sposobov-nastroit-vizualizaciyu-matricy-korrelyacii>
8. Свёрточная нейронная сеть с нуля.: – Режим доступа: <https://habr.com/ru/company/ruvds/blog/488342/>
9. Основы работы с python: – Режим доступа: <https://pythonru.com/osnovy>