



ЦЕНТР  
ДОПОЛНИТЕЛЬНОГО  
ОБРАЗОВАНИЯ  
МГТУ им. Н.Э. Баумана

# ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА по курсу «Data Science»

Дьяченко Юлия Анатольевна



## Цели и задачи

1

Разработать модели для прогноза модуля упругости при растяжении, прочности при растяжении и соотношения «матрица-наполнитель»

2

Провести разведочный анализ и предобработку данных

3

Обучить модели для прогноза модуля упругости при растяжении и прочности при растяжении

4

Написать нейронную сети, которая будет рекомендовать соотношение матрица-наполнитель и разработка приложения

5

Создать репозиторий в GitHub / GitLab и разместить там код исследования



Алгоритмы машинного обучения:

1. «Случайный лес» (RandomForest);
2. Линейная регрессия (Linear regression);
3. Градиентный бустинг (Gradient Boosting).

Библиотеки:

1. Pandas;
2. NumPy;
3. Matplotlib;
4. Seaborn;
5. Tensorflow.



## Библиотека Pandas Profiling раздел Overview (Обзор)

- количество наблюдений;
- количество переменных;
- тип данных;
- процент и количество пропущенных значений;
- процент и количество дубликатов.

Overview

Reproduction

### Dataset statistics

Number of variables	13
Number of observations	1023
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	144.2 KiB
Average record size in memory	144.3 B

### Variable types

Numeric	12
Categorical	1

**Раздел Overview из отчета  
Pandas Profiling**



## Блок Variables

- количество уникальных записей и их процент;
- количество пропущенных значений и их процент;
- количество значений NaN и их процент;
- среднее, минимальное и максимальное значение;
- количество и процент нулевых значений;
- график распределения значений

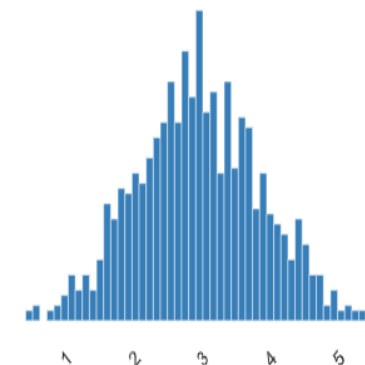
### Variables

Select Columns

Соотношение матрица-наполнитель

Real number ( $\mathbb{R}$ )

Distinct	1014	Minimum	0.38940261
Distinct (%)	99.1%	Maximum	5.5917416
Missing	0	Zeros	0
Missing (%)	0.0%	Zeros (%)	0.0%
Infinite	0	Negative	0
Infinite (%)	0.0%	Negative (%)	0.0%
Mean	2.9303658	Memory size	16.0 KiB

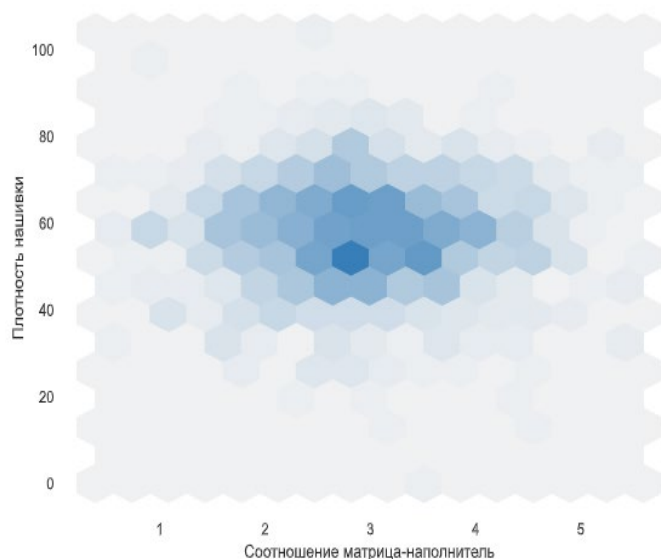


Раздел Variables, сведения о признаке  
«Соотношение матрица-наполнитель»

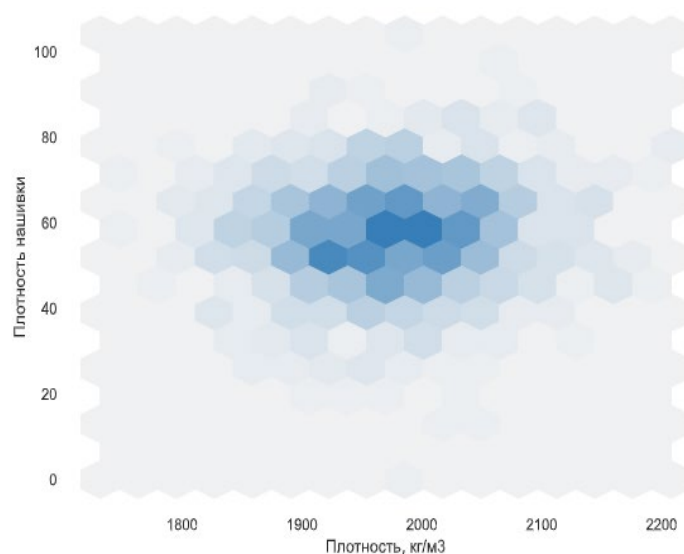


## Отчет Pandas Profiling

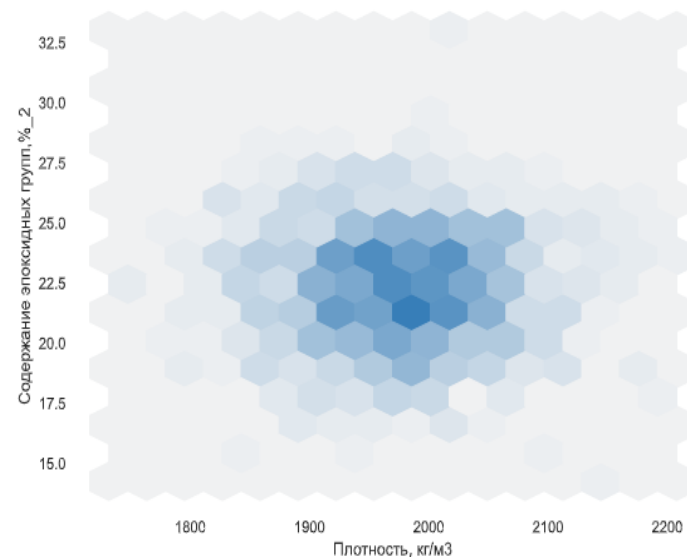
Блок Interactions, генерация графиков по парам переменных для визуализации зависимостей и распределения значений



**график зависимости  
между переменными  
«Плотность нашивки» и  
«Соотношение  
матрица/наполнитель»**



**график зависимости  
между переменными  
«Плотность нашивки» и  
«Плотность, кг/м3»**



**график зависимости  
между переменными  
«Содержание  
эпоксидных групп,%\_2»  
и «Плотность, кг/м3»**



## Отчет Pandas Profiling

### Блок Correlations, значения корреляции всех пар переменных

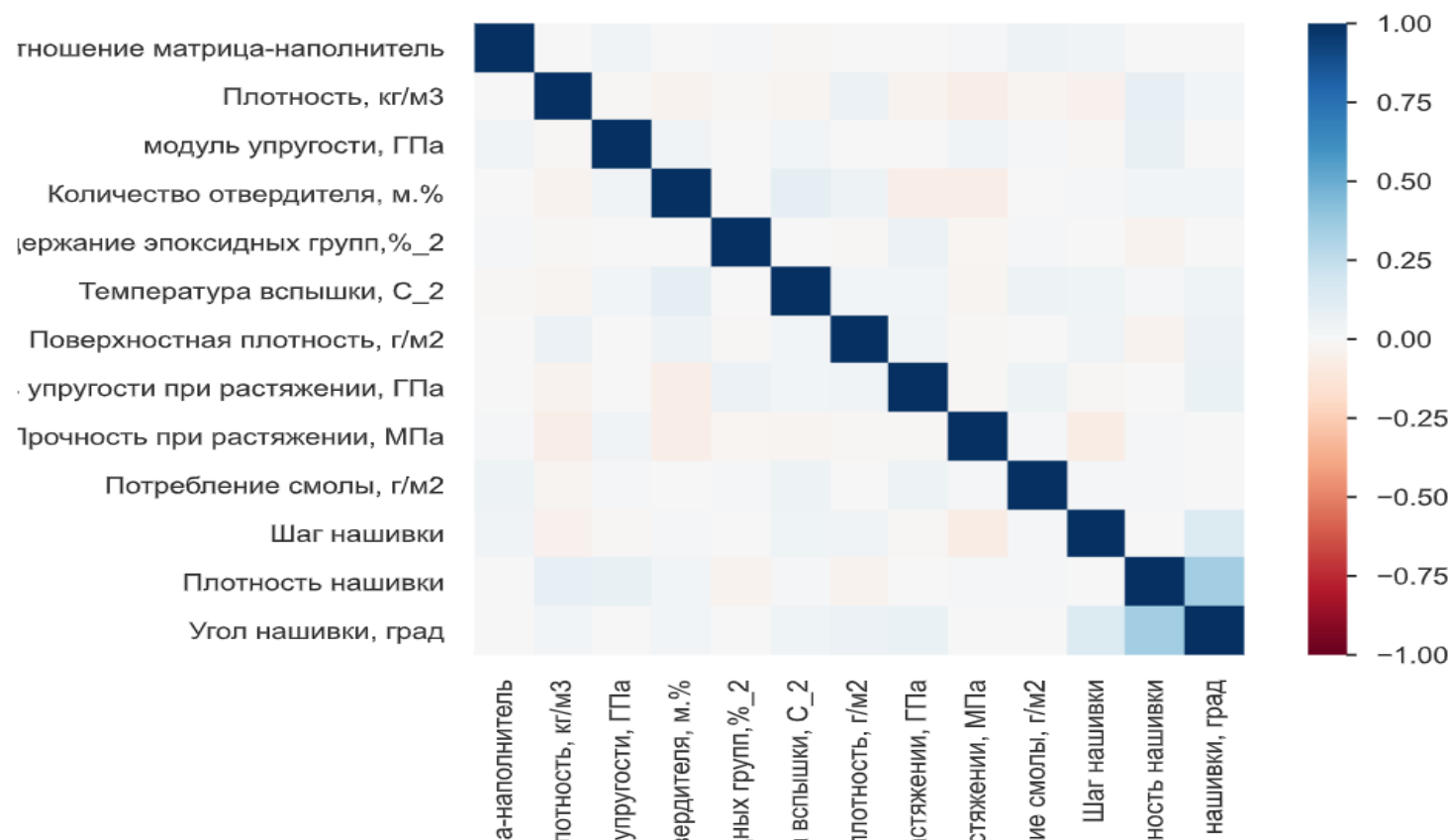


график корреляции всех пар переменных



# Отчет Pandas Profiling

## Блок Missing values, анализ пропущенных значений

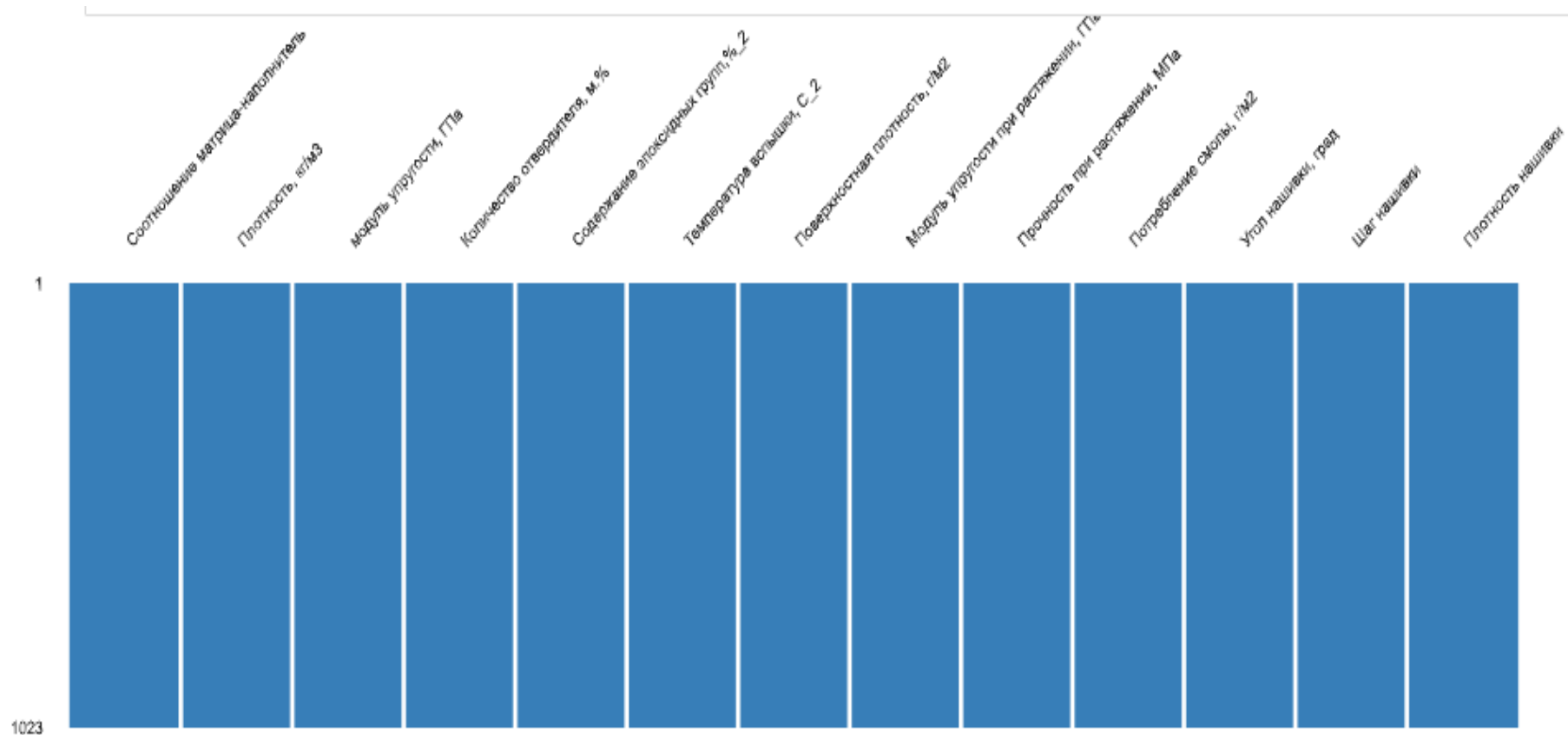


график пропущенных значений





# Отчет Pandas Profiling

## Блок Sample, выборкой первых и последних строк

First rows

Last rows

	Соотношение матрица-наполнитель	Плотность, кг/м3	модуль упругости, ГПа	Количество отвердителя, м.%	Содержание эпоксидных групп,%_2
0	1.857143	2030.0	738.736842	30.00	22.267857
1	1.857143	2030.0	738.736842	50.00	23.750000
2	1.857143	2030.0	738.736842	49.90	33.000000
3	1.857143	2030.0	738.736842	129.00	21.250000
4	2.771331	2030.0	753.000000	111.86	22.267857
5	2.767918	2000.0	748.000000	111.86	22.267857
6	2.569620	1910.0	807.000000	111.86	22.267857
7	2.561475	1900.0	535.000000	111.86	22.267857
8	3.557018	1930.0	889.000000	129.00	21.250000
9	3.532338	2100.0	1421.000000	129.00	21.250000

Выборка строк датасета



# Оценка выбросов с помощью Boxplot («ящик с усами»)

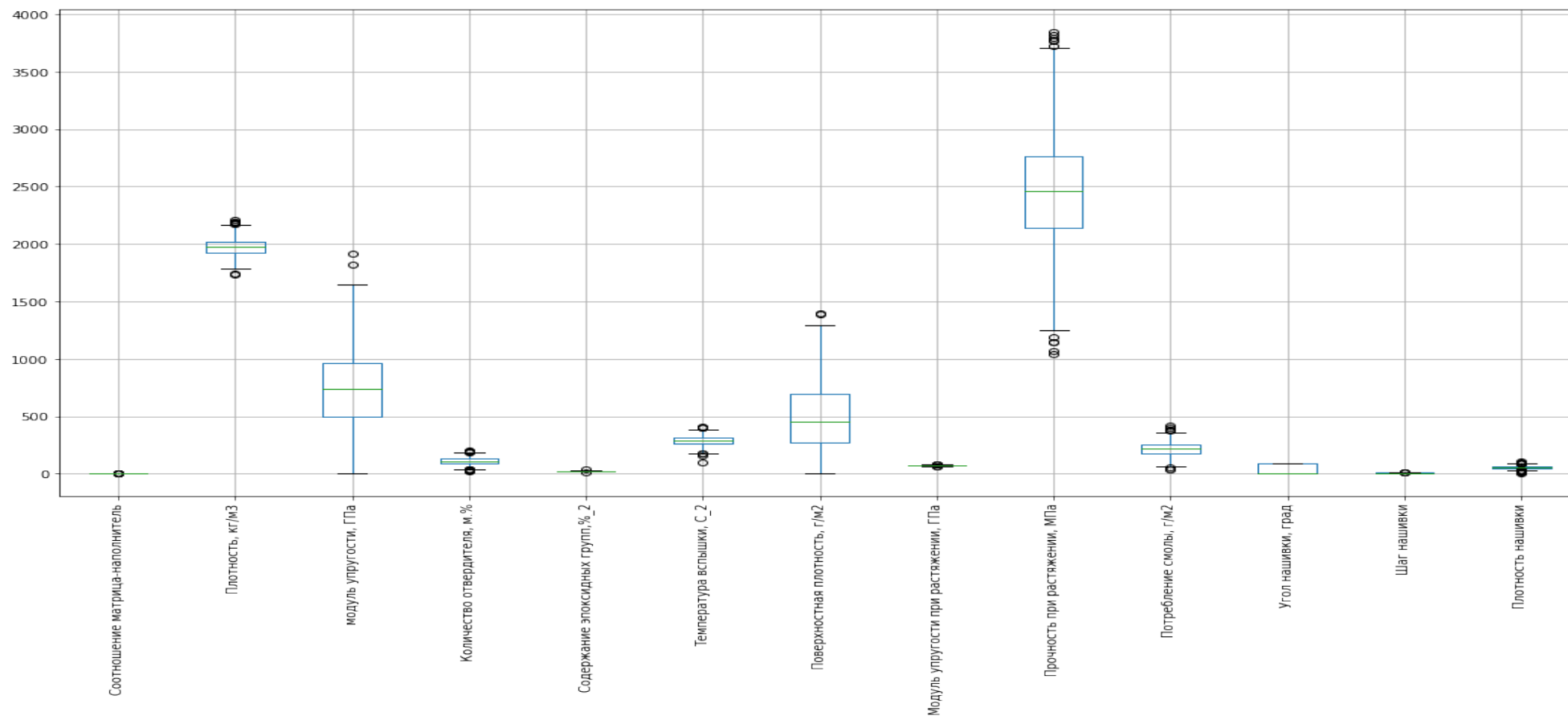
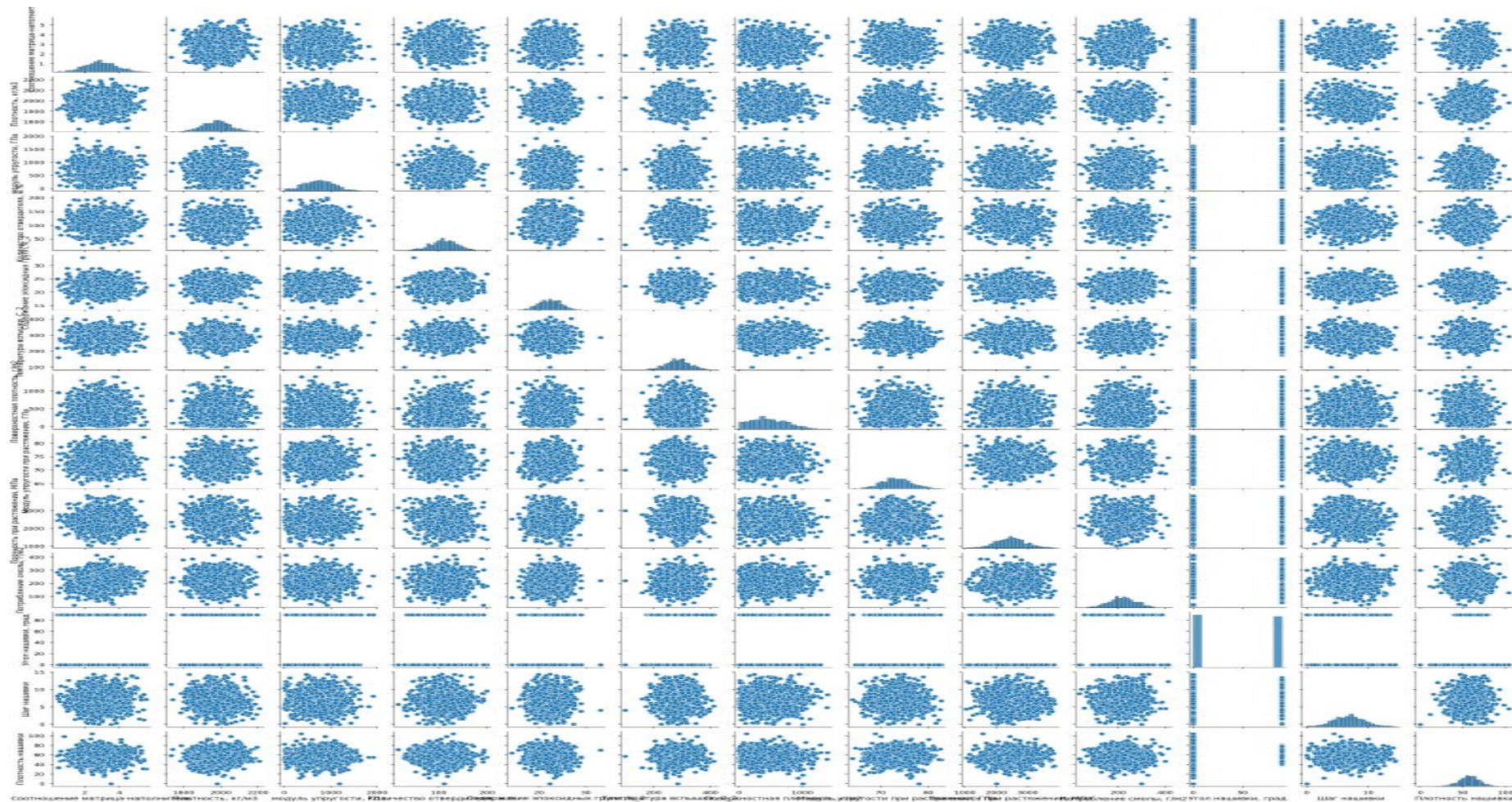


Диаграмма Boxplot или «ящик с усами»



# Попарные графики рассеяния точек



Попарные графики рассеяния точек





## Предобработка данных

К задачам предварительной обработки данных относятся:

- Очистка данных;
- Редактирование данных;
- Заполнение пропусков.

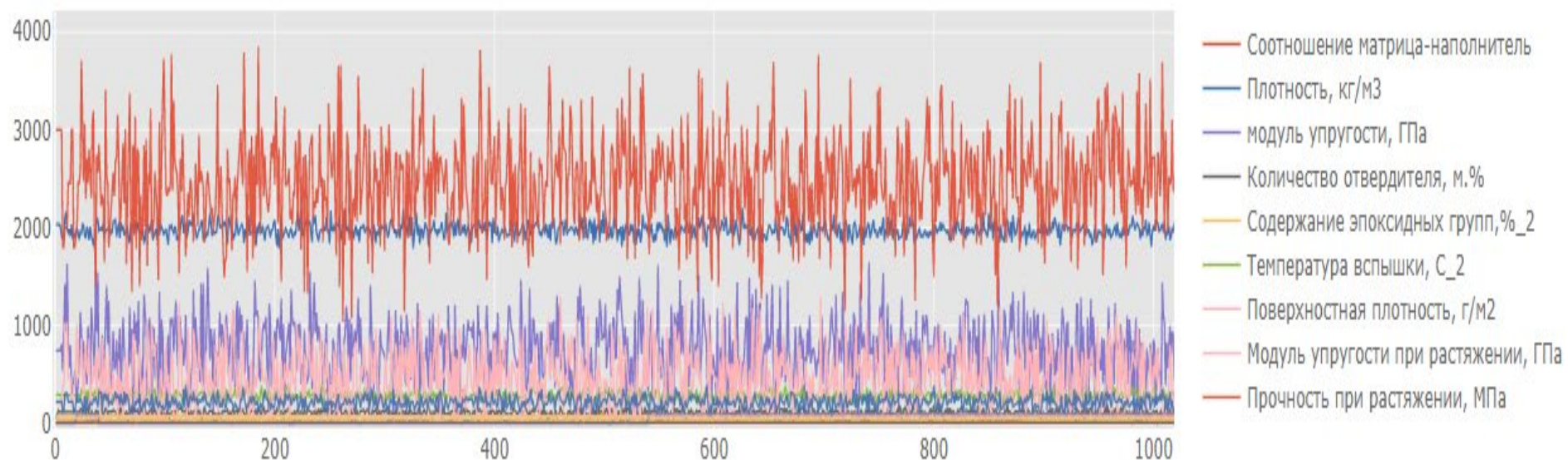


График распределения значений до нормализации данных



# Нормализация данных

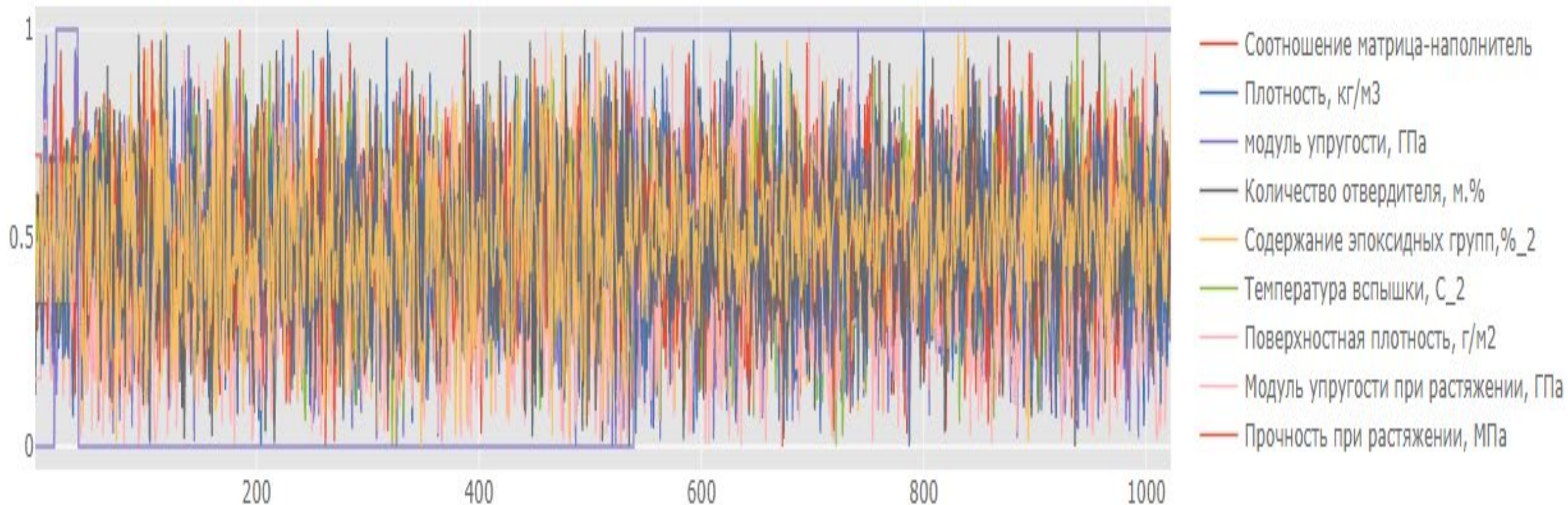


График распределения значений после нормализации



## Разработка и обучение моделей

- Разделение на обучающую и тестовую выборки в соотношении 70/30;
- Подбор гиперпараметров для моделей с помощью поиска по сетке с перекрестной проверкой;
- Построение модели при помощи линейной регрессии;
- Построение модели при помощи «случайного леса»;
- Построение модели при помощи градиентного бустинга.



# Тестирование моделей

	Model	MAE	MSE	R2 score
Прочность при растяжении	LinearRegression()	0.13712	0.029412	-0.017691

	Model	MAE	MSE	R2 score
Прочность при растяжении	RandomForestRegressor(random_state=42)	0.135316	0.028723	0.006151

	Model	MAE	MSE	R2 score
Прочность при растяжении	GradientBoostingRegressor()	0.13497	0.028699	0.006997

Ошибки модели предсказания для параметра  
«Прочность при растяжении»

	Model	MAE	MSE	R2 score
Модуль упругости при растяжении	LinearRegression()	0.138812	0.029498	0.00418

	Model	MAE	MSE	R2 score
Модуль упругости при растяжении	RandomForestRegressor(random_state=42)	0.138678	0.029595	0.000933

	Model	MAE	MSE	R2 score
Модуль упругости при растяжении	GradientBoostingRegressor()	0.13886	0.029657	-0.001164

Ошибки модели предсказания для параметра  
«Модуль упругости при растяжении»



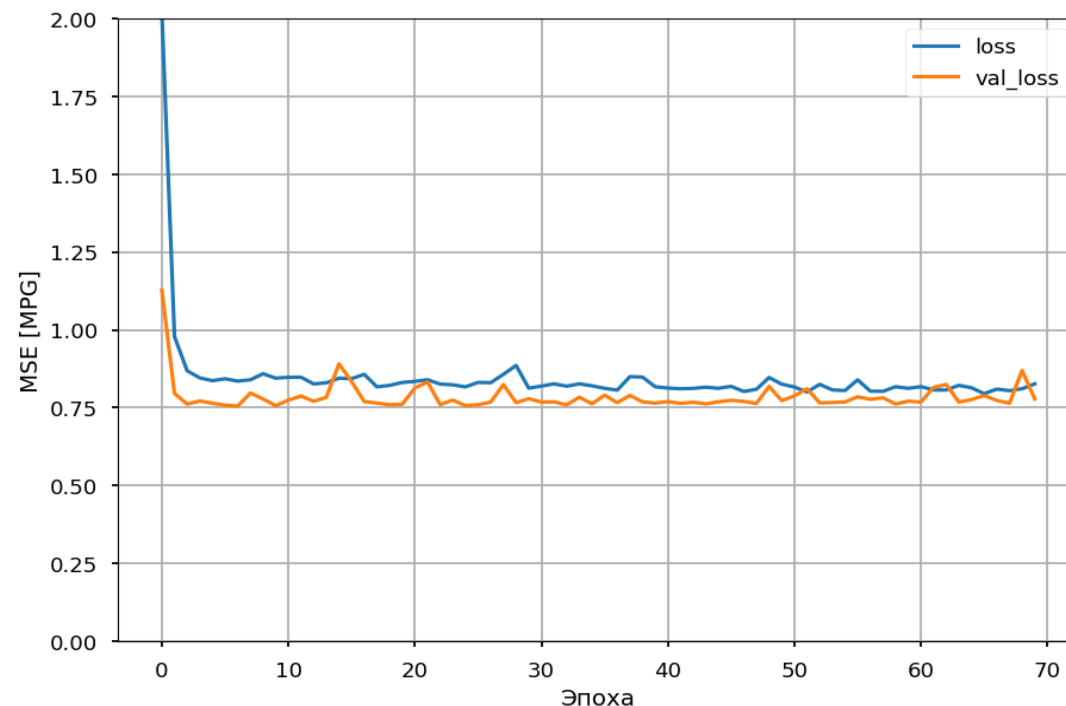
# Нейронная сеть

Model: "sequential"

Layer (type)	Output Shape	Param #
normalization (Normalization)	(None, 12)	3
dense (Dense)	(None, 1024)	13312
dense_1 (Dense)	(None, 1024)	1049600
dense_2 (Dense)	(None, 1)	1025

=====  
Total params: 1,063,940  
Trainable params: 1,063,937  
Non-trainable params: 3

Информация о модели нейронной сети



Визуализация ошибки модели нейронной сети





ЦЕНТР  
ДОПОЛНИТЕЛЬНОГО  
ОБРАЗОВАНИЯ  
МГТУ им. Н.Э. Баумана

# Разработка приложения

## Прогнозирование конечных свойств новых материалов (композиционных материалов)

Спрогнозировать значение матрица-наполнитель

### Композиционные материалы

Композиционные материалы — это искусственно созданные материалы, состоящие из нескольких других с четкой границей между ними. Композиты обладают теми свойствами, которые не наблюдаются у компонентов по отдельности. При этом композиты являются монолитным материалом, т. е. компоненты материала неотделимы друг от друга без разрушения конструкции в целом.

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

В данном веб-приложении с помощью нейронной сети прогнозируются соотношение "Матрица-наполнитель" в композиционных материалах, на основе введенных пользователем значений.

Ссылка на репозиторий в GitHub:  
[https://github.com/shish27/VKR\\_Dyachenko](https://github.com/shish27/VKR_Dyachenko)

## Расчет соотношения матрица-наполнитель

> Введите параметры для модели

Введите Плотность, кг/м<sup>3</sup>

Введите Модуль упругости, ГПа

Введите Количество отвердителя, м. %

Введите Содержание эпоксидных групп, %<sub>2</sub>

Введите Температура вспышки, С<sub>2</sub>

Введите Поверхностная плотность, г/м<sup>2</sup>

Введите Модуль упругости при растяжении, ГПа

Введите Прочность при растяжении, МПа

Введите Потребление смолы, г/м<sup>2</sup>

Введите Угол нашивки, град

Введите Шаг нашивки

Введите Плотность нашивки

Рассчитать

Сбросить

Спрогнозированное Соотношение матрица-наполнитель для введенных параметров: [[2.8133044]]

[Вернуться на главную страницу](#)



ЦЕНТР  
ДОПОЛНИТЕЛЬНОГО  
ОБРАЗОВАНИЯ  
МГТУ им. Н.Э. Баумана



[do.bmstu.ru](https://do.bmstu.ru)