

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА
по курсу «Data Science»

Тема:

**Прогнозирование конечных свойств
новых композиционных материалов**

Слушатель: Савичев П.М.

АКТУАЛЬНОСТЬ ТЕМЫ

Композиционные материалы — это искусственно созданные материалы, состоящие из нескольких других с четкой границей между ними, сами по себе композиты монолитны и неотделимы друг от друга (без разрушения конструкции в целом) и обладают теми свойствами, которые не наблюдаются у компонентов по отдельности.

Есть два пути для определения характеристик композита, состоящего из компонентов:

- физические испытания образцов материалов
- прогнозирование характеристик

Суть прогнозирования заключается в симуляции представительного элемента объема композита, на основе данных о характеристиках входящих компонентов.

На входе: имеются данные о начальных свойствах компонентов композиционных материалов

На выходе: необходимо спрогнозировать ряд конечных свойств получаемых композиционных материалов

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

ХАРАКТЕРИСТИКИ АНАЛИЗИРУЕМОГО ДАТАСЕТА

Датасет со свойствами композитов: «X_bp.xlsx» и «X_nup.xlsx» (две таблицы Microsoft Excel).
Объединение делаем по индексу, тип объединения INNER.

- объём объединенного датасета: 1023 записи по каждому показателю
- пропуски отсутствуют (пустых значений нет)

```
df_all.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
Int64Index: 1023 entries, 0 to 1022
```

```
Data columns (total 13 columns):
```

#	Column	Non-Null Count	Dtype
0	Соотношение матрица-наполнитель	1023 non-null	float64
1	Плотность, кг/м3	1023 non-null	float64
2	модуль упругости, ГПа	1023 non-null	float64
3	Количество отвердителя, м.%	1023 non-null	float64
4	Содержание эпоксидных групп,%_2	1023 non-null	float64
5	Температура вспышки, C_2	1023 non-null	float64
6	Поверхностная плотность, г/м2	1023 non-null	float64
7	Модуль упругости при растяжении, ГПа	1023 non-null	float64
8	Прочность при растяжении, МПа	1023 non-null	float64
9	Потребление смолы, г/м2	1023 non-null	float64
10	Угол нашивки, град	1023 non-null	int64
11	Шаг нашивки	1023 non-null	float64
12	Плотность нашивки	1023 non-null	float64

```
dtypes: float64(12), int64(1)
```

```
memory usage: 111.9 KB
```

X_bp.xlsx

X_nup.xlsx

АНАЛИЗ МОЩНОСТИ ВЫБОРКИ - МАТРИЦА

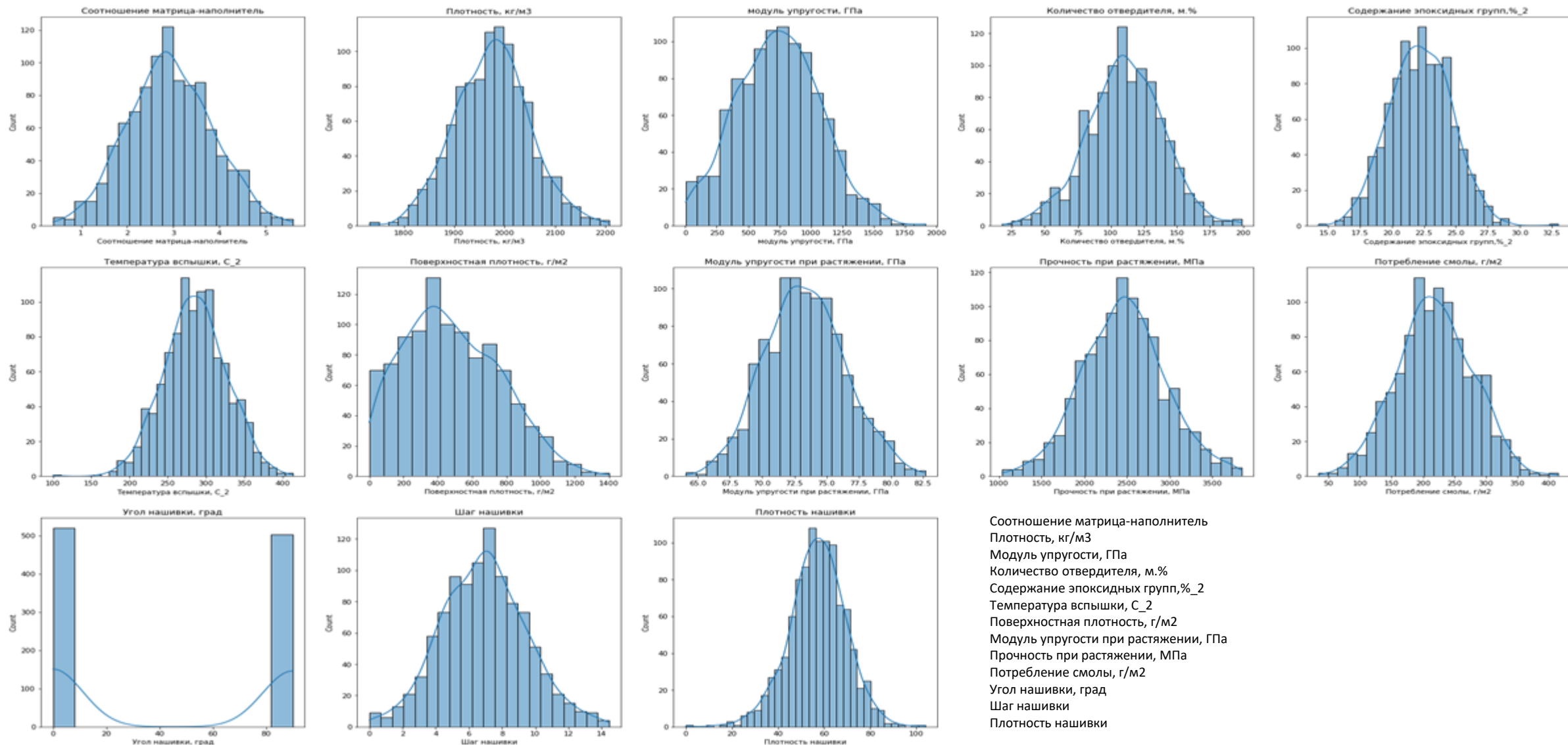
Представлены основные характеристики параметров датасета:

- количество элементов,
- средние значения параметров,
- минимальные и максимальные значения, а также медианные значения.

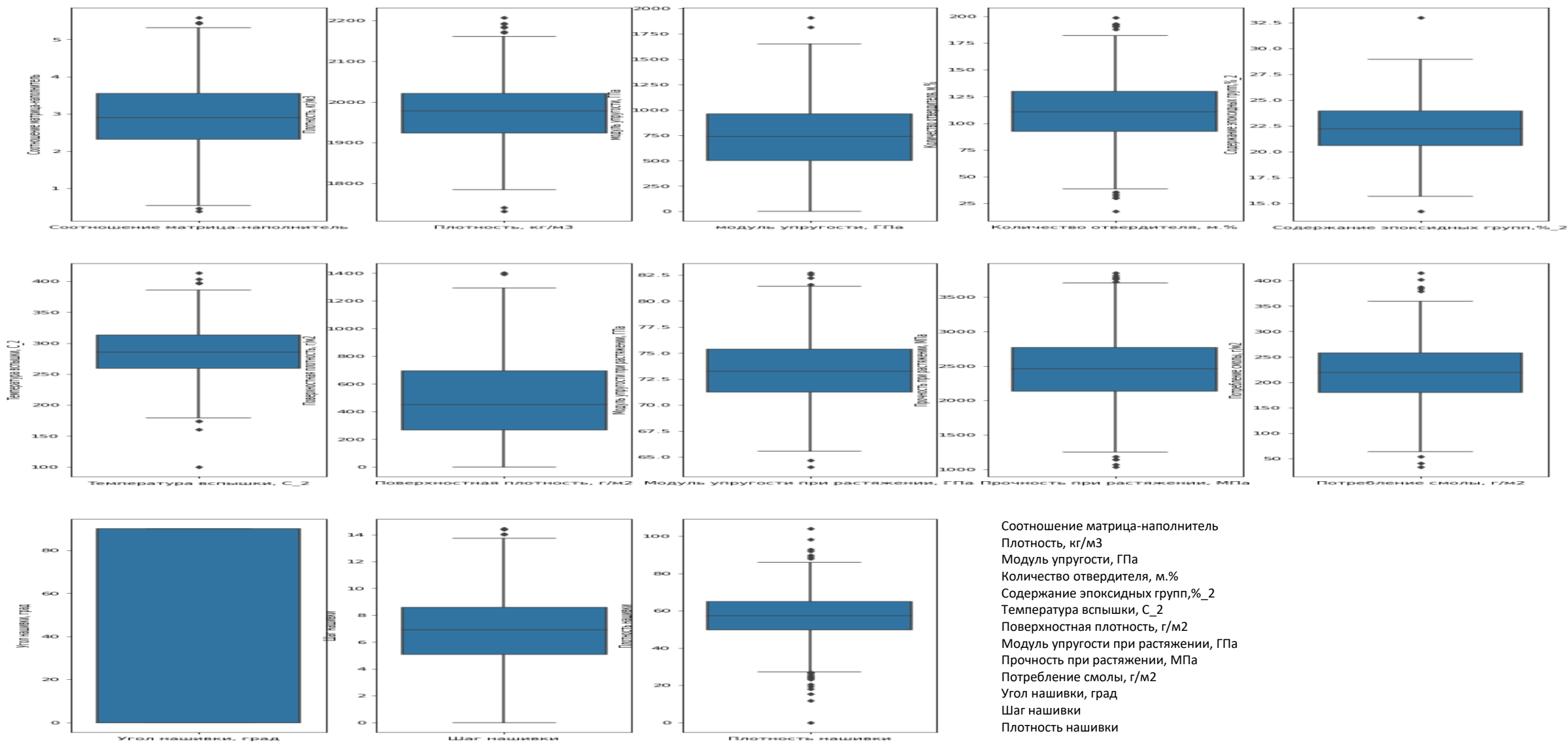
```
df_all.describe()
```

	Соотношение матрица-наполнитель	Плотность, кг/м3	модуль упругости, ГПа	Количество отвердителя, м.%	Содержание эпоксидных групп,%_2	Температура вспышки, С_2	Поверхностная плотность, г/м2	Модуль упругости при растяжении, ГПа	Прочность при растяжении, МПа	Потребление смолы, г/м2	Уг нашивк гр
count	936.000000	936.000000	936.000000	936.000000	936.000000	936.000000	936.000000	936.000000	936.000000	936.000000	936.0000
mean	2.925683	1974.040023	738.247627	110.916216	22.209030	286.040414	482.993901	73.305127	2467.488822	217.613374	46.0576
std	0.893712	70.808120	328.708665	27.037891	2.394871	39.400677	280.190377	3.037381	463.838911	57.827255	45.0116
min	0.547391	1784.482245	2.436909	38.668500	15.695894	179.374391	0.603740	65.553336	1250.392802	63.685698	0.0000
25%	2.321931	1923.443748	498.577158	92.523816	20.571516	259.184486	266.004099	71.248823	2146.936034	179.489091	0.0000
50%	2.904731	1977.258043	738.736842	111.113175	22.184713	286.024118	457.732246	73.259230	2457.959767	218.388715	90.0000
75%	3.546650	2020.158764	958.418993	130.001450	23.961818	312.991425	695.900862	75.310788	2755.169485	256.396777	90.0000
max	5.314144	2161.565216	1649.415706	181.828448	28.955094	386.067992	1291.340115	81.417126	3705.672523	359.052220	90.0000

ГИСТОГРАММЫ РАСПРЕДЕЛЕНИЯ



ЯЩИК С УСАМИ



МАТРИЦА КОРРЕЛЯЦИЯ



МОДЕЛИ МАШИННОГО ОБУЧЕНИЯ

- `sklearn.linear_model.LinearRegression`
- `sklearn.linear_model.SGDRegressor`
- `sklearn.neighbors.KNeighborsRegressor`
- `sklearn.ensemble.RandomForestRegressor`
- `sklearn.tree.DecisionTreeRegressor`
- `sklearn.ensemble.HistGradientBoostingRegressor`

Используемые библиотеки:

NumPy

- Pandas
- Matplotlib

Seaborn

- Sklearn
- Keras

ЭТАПЫ ОБРАБОТКИ ДАННЫХ

Шаг 1. Анализ всех признаков для определения их максимальных, минимальных, средних и медианных значений, а также выполнена визуализация данных.

Шаг 2. Исключение выбросов данных, то есть точек данных, которые лежали вдали от обычного распределения данных. Визуализацию таких значений произвели с помощью диаграммы ящиков с усами.

Шаг 3. Нормализация данных.

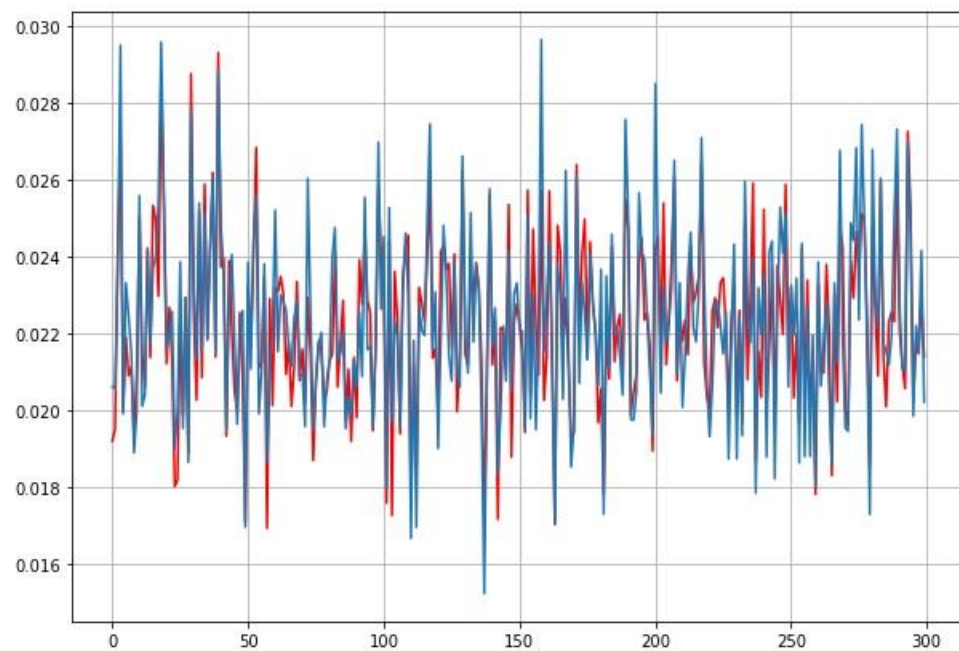
После нормализации данных был также проведен анализ взаимосвязи переменных друг с другом.

Вывод: по результатам предобработки данных между параметрами не наблюдается корреляций и очевидных связей. Число выбросов получилось незначительным.

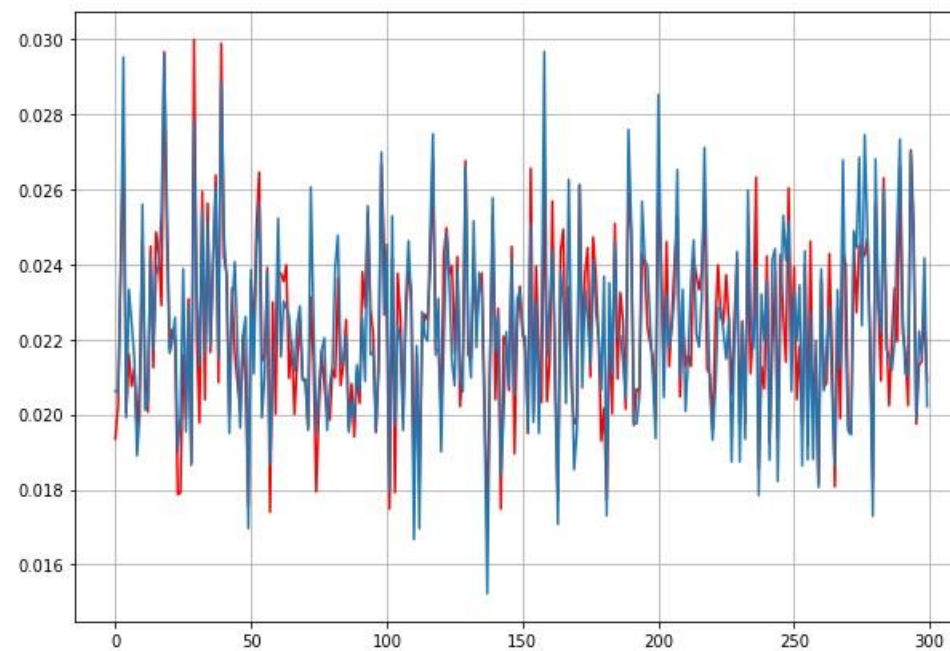
Для рекомендации соотношения «матрица-наполнитель» была разработана простая модель глубокого обучения с помощью Keras.

ПРЕДСКАЗАНИЕ ДЛЯ ПРИЗНАКА «МОДУЛЬ УПРУГОСТИ ПРИ РАСТЯЖЕНИИ»

- Линейная регрессия

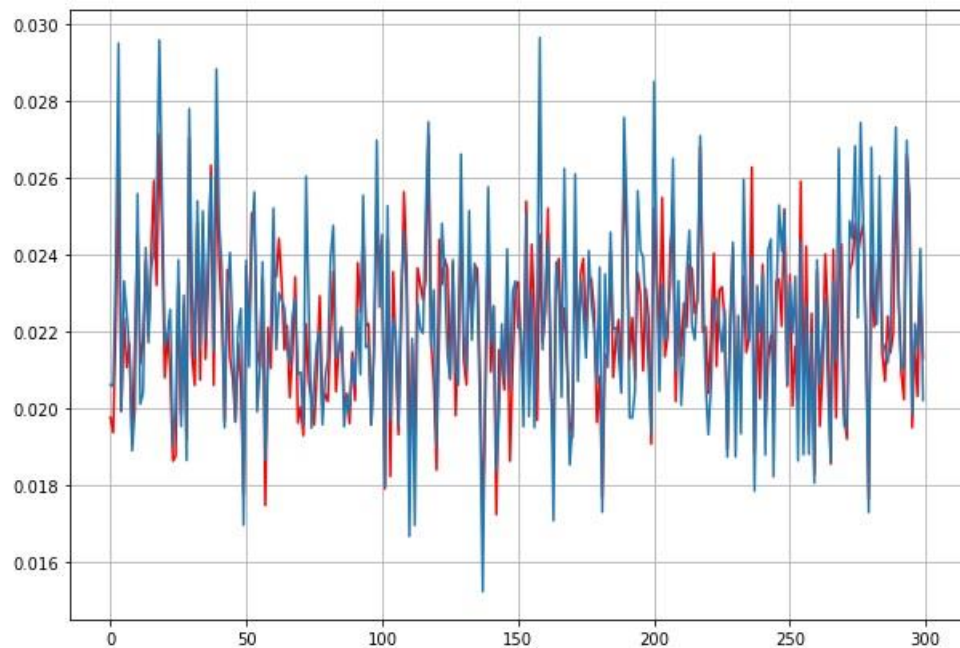


- Случайный лес

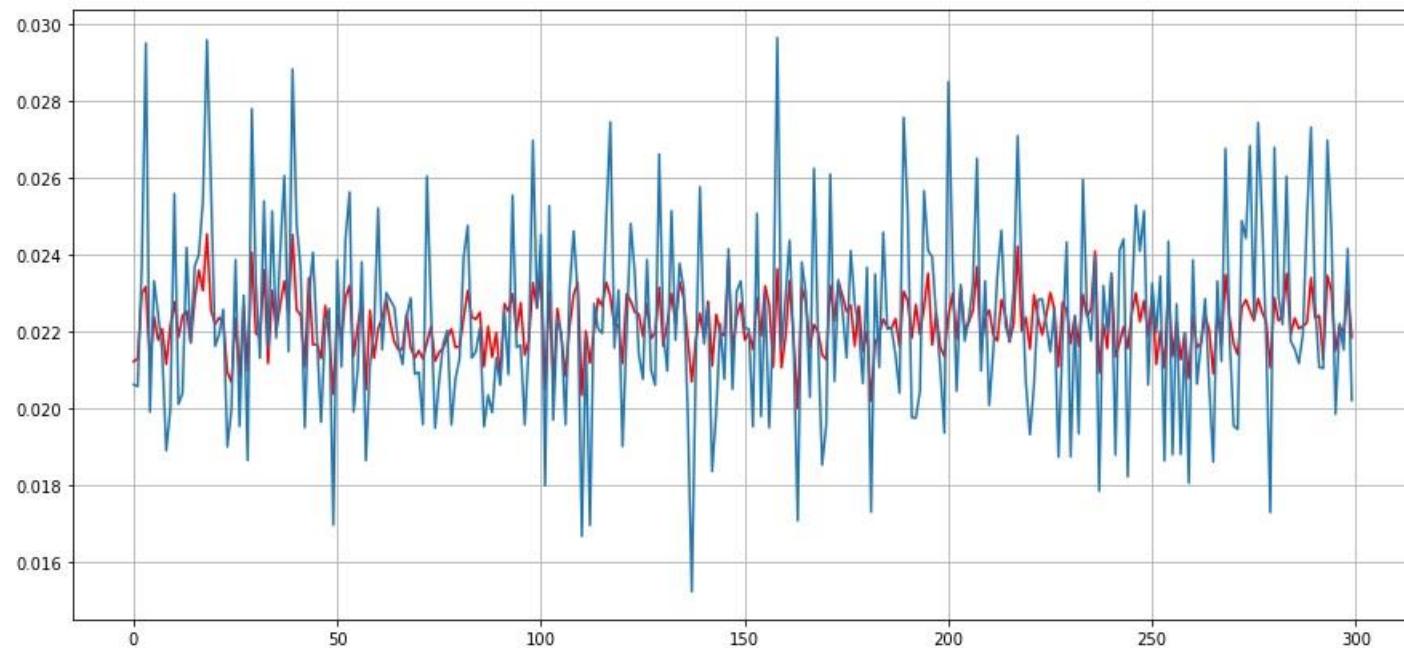


ПРЕДСКАЗАНИЕ ДЛЯ ПРИЗНАКА «МОДУЛЬ УПРУГОСТИ ПРИ РАСТЯЖЕНИИ»

- К-ближайших соседей

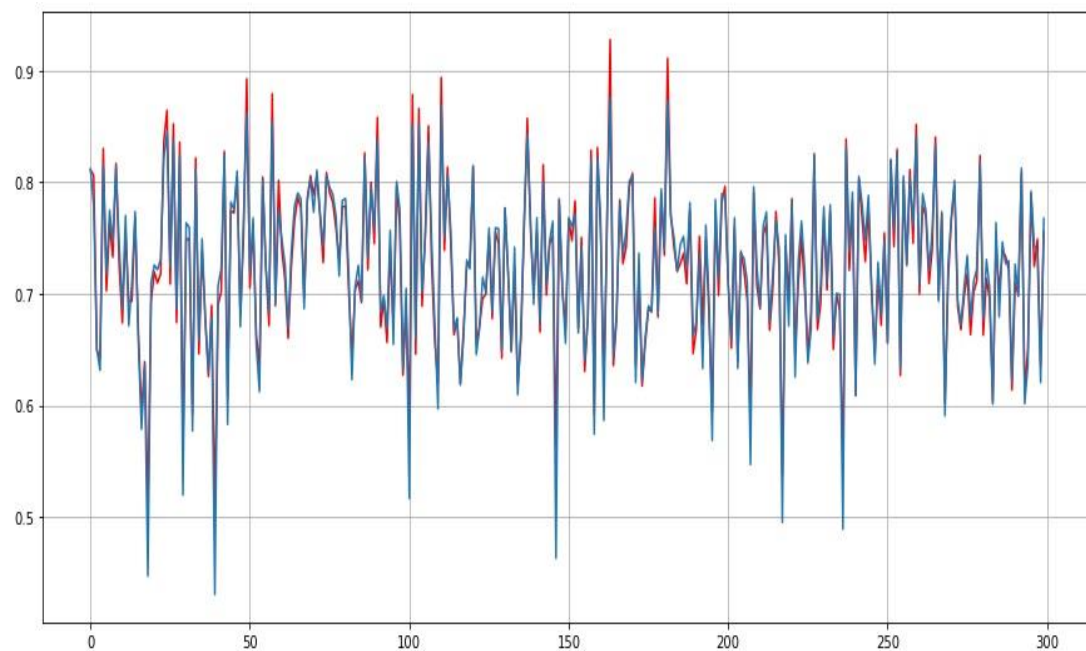


- Стохастический градиентный спуск

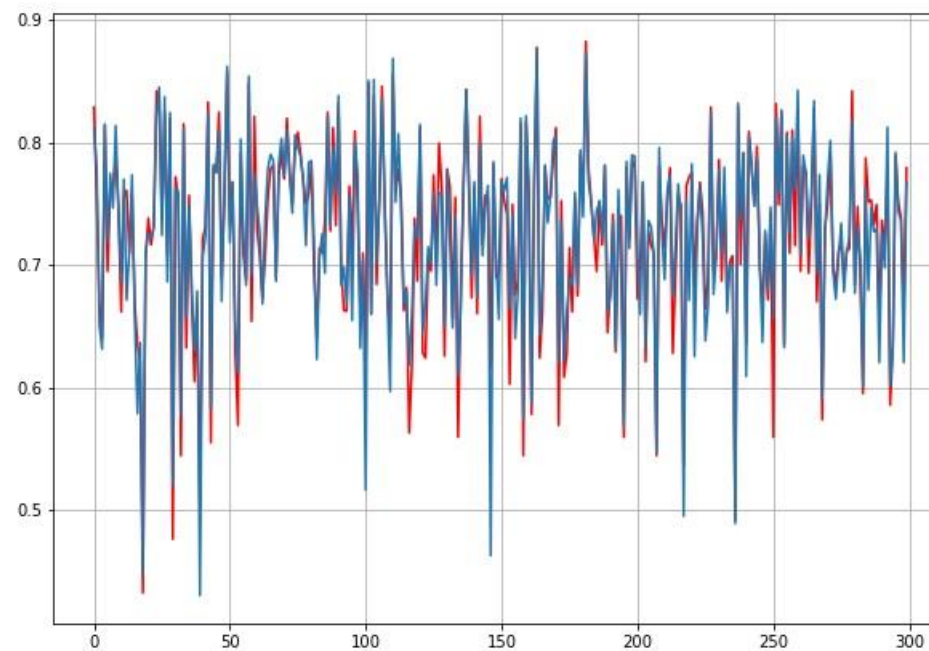


ПРЕДСКАЗАНИЕ ДЛЯ ПРИЗНАКА «ПРОЧНОСТЬ ПРИ РАСТЯЖЕНИИ»

- Линейная регрессия

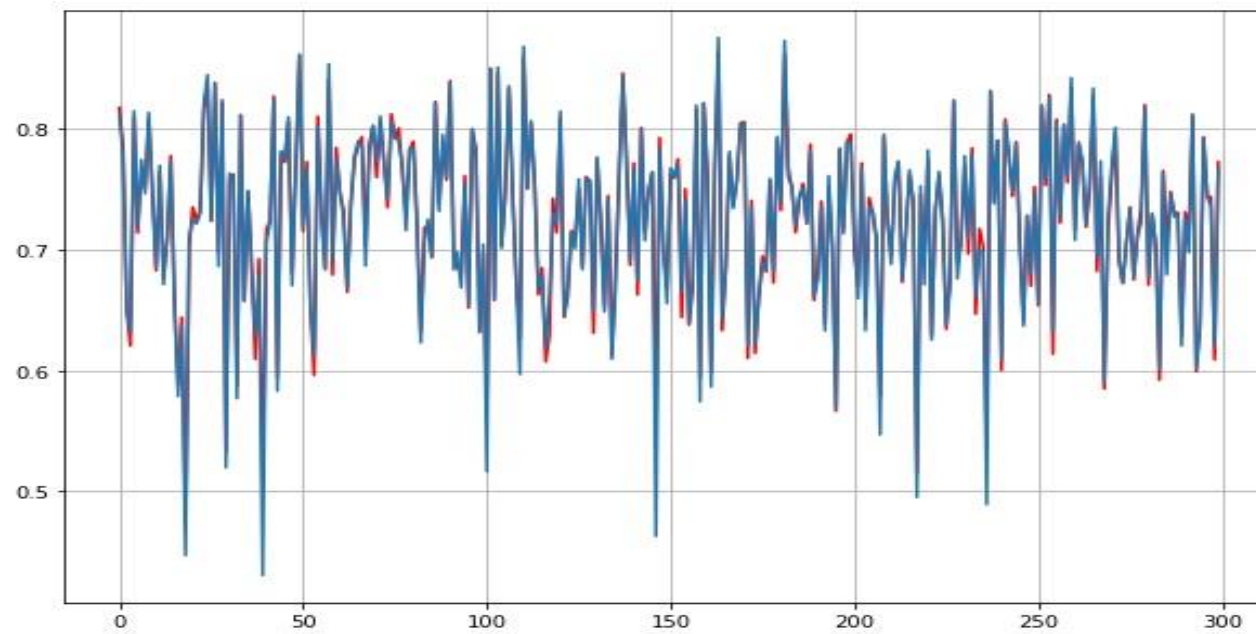


- Деревья решений

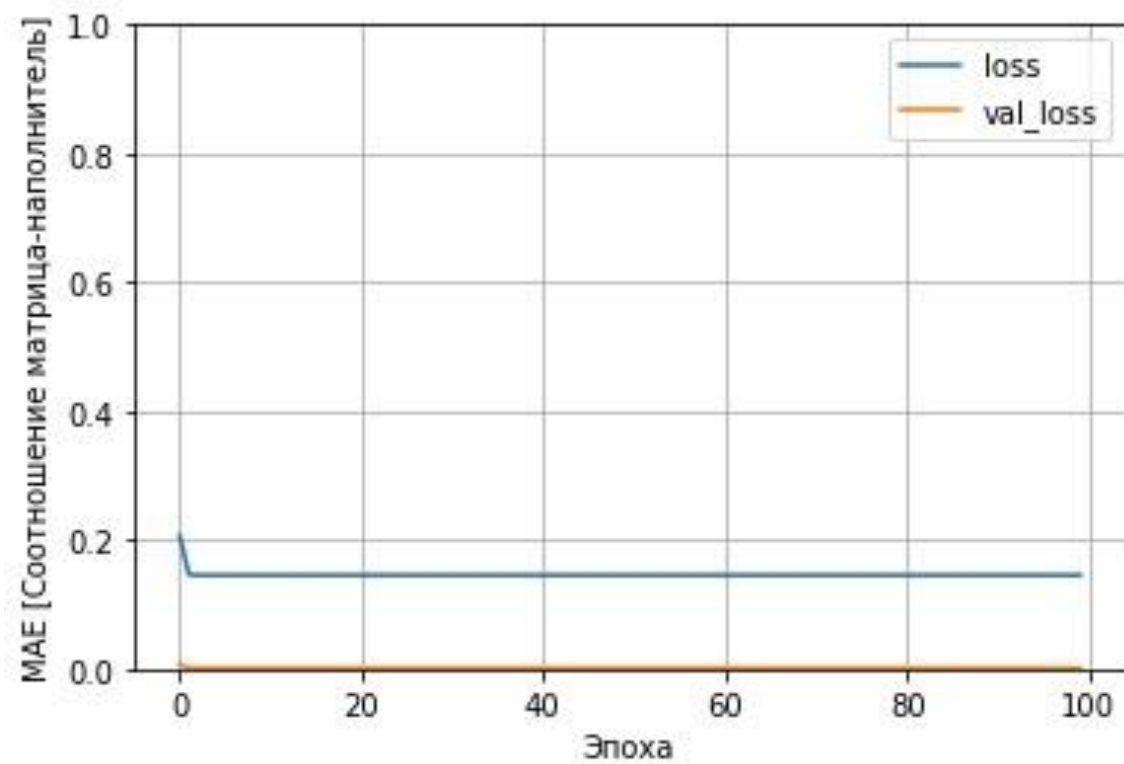


ПРЕДСКАЗАНИЕ ДЛЯ ПРИЗНАКА «ПРОЧНОСТЬ ПРИ РАСТЯЖЕНИИ»

- Градиентный бустинг



ПРЕДСКАЗАНИЕ ДЛЯ ПРИЗНАКА «СООТНОШЕНИЯ МАТРИЦА-НАПОЛНИТЕЛЬ»



УДАЛЁННЫЙ РЕПОЗИТАРИЙ

Репозиторий проекта находится на сервис GitHub.

Для просмотра репозитория с ВКР необходимо перейти по ссылке:

- https://github.com/crtman/bmstu_final_qualifying_work

ЗАКЛЮЧЕНИЕ

Теоретически разработанный метод определения надёжности изделий из композиционных материалов, основанный на использовании статистически достоверных характеристик материалов, полученных физическим и вычислительным экспериментом, позволяет оценивать уровень надёжности изделий как в отдельных точках, так и по всему объёму в целом.

СПАСИБО ЗА ВНИМАНИЕ