



**ОБРАЗОВАТЕЛЬНЫЙ
ЦЕНТР** МГТУ им. Н. Э. Баумана

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА по курсу «Data Science»

Прогнозирование конечных свойств новых материалов (композиционных материалов)

Слушатель: Грицай Александр Николаевич
студент курса «Data Science»
Образовательного центра МГТУ им. Н. Э. Баумана

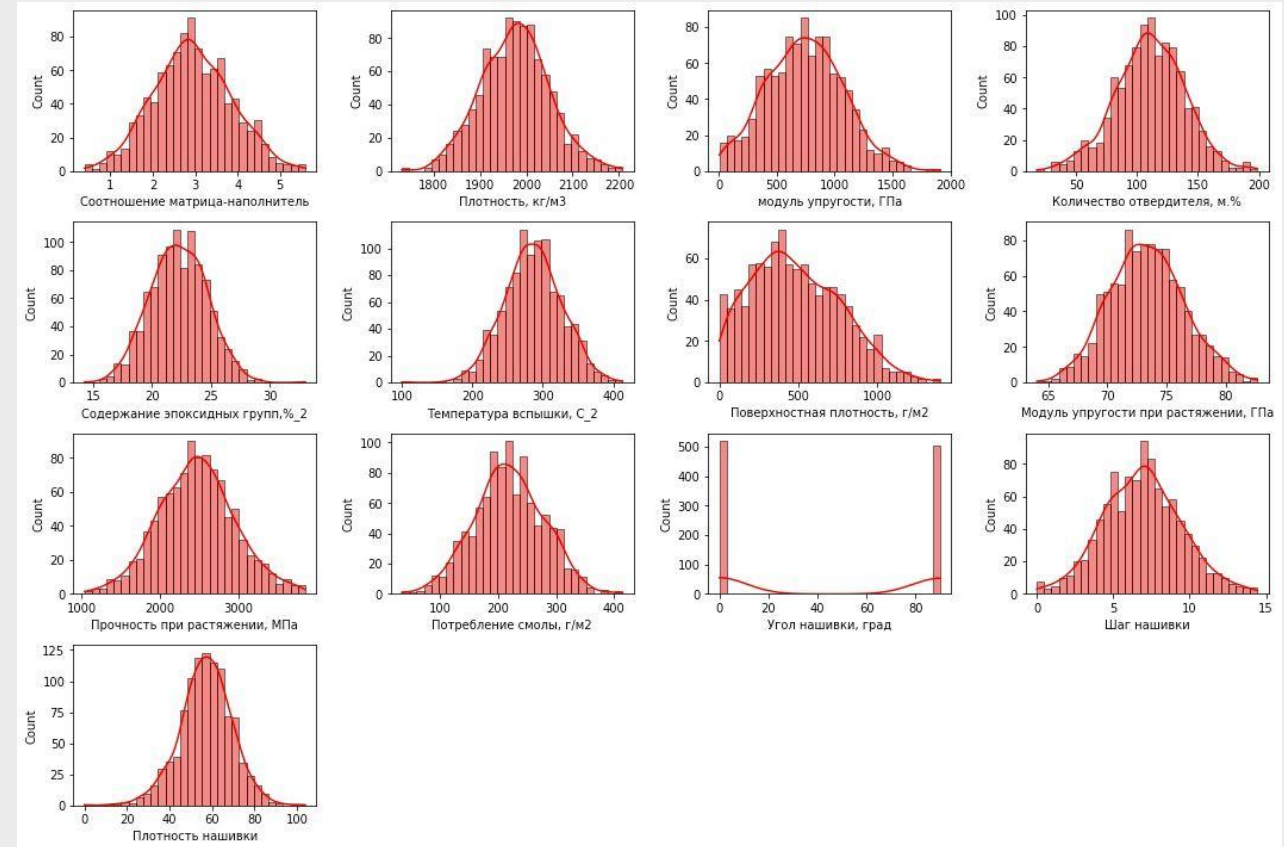
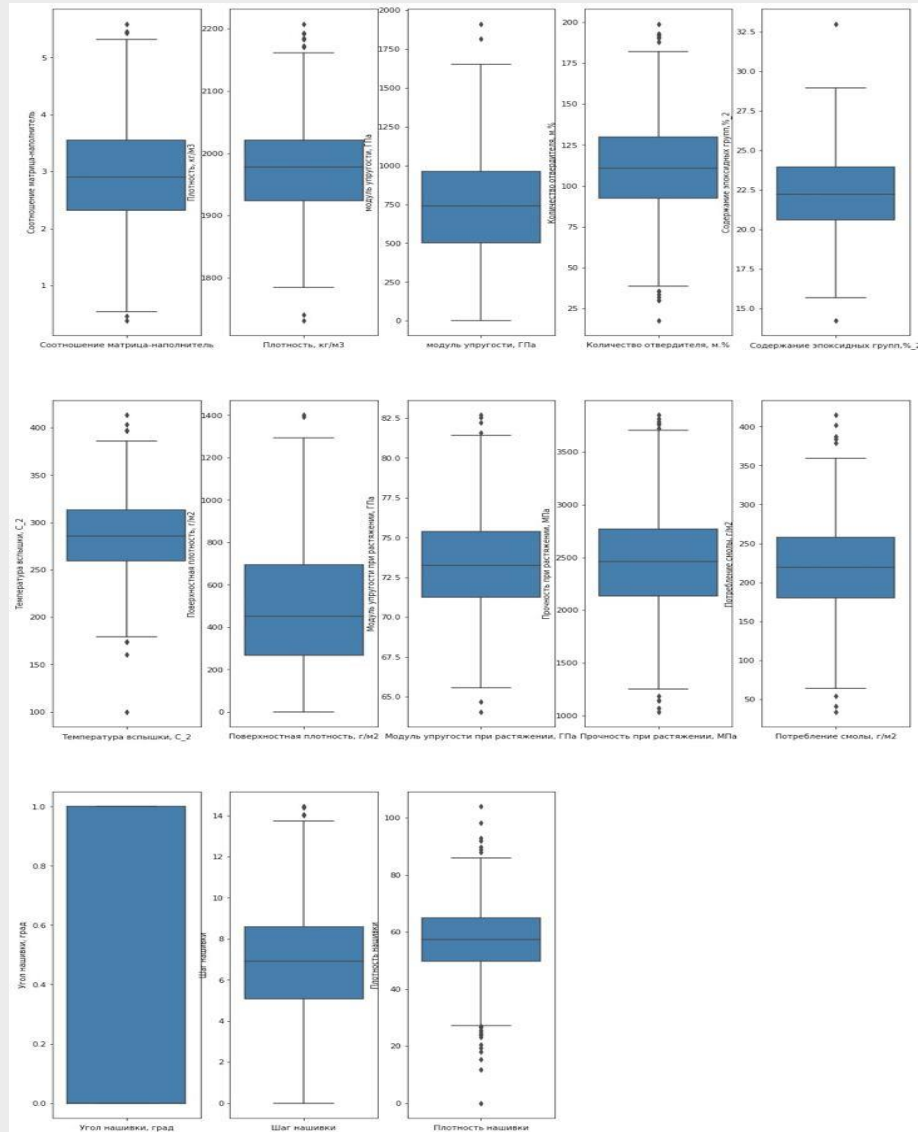
Задачи работы

- Описать методы, которые используются для решений
- Провести разведочный анализ данных предложенных датасетов:
 - 1. построить гистограммы распределения каждой из переменных
 - 2. построить диаграммы «ящики с усами»
 - 3. построить попарные графики рассеяния точек
 - 4. получить среднее и медианное значения
 - 5. исключить выбросы, проверить отсутствие пропусков.
- Провести предобработку данных: удаление шумов, нормализацию
- Обучить нескольких моделей для прогноза модуля упругости при растяжении и прочности при растяжении.
- Написать нейронную сеть, предназначенную для рекомендаций соотношения матрица-наполнитель.
- Разработать приложение с графическим интерфейсом, которое будет выдавать прогноз.
- Оценить точность модели на тренировочном и тестовом датасете.
- Создать репозиторий в GitHub и разместить там код исследования.

Статистические характеристики до предобработки

	count	mean	std	min	25%	50%	75%	max
Соотношение матрица-наполнитель	1023.0	2.930366	0.913222	0.389403	2.317887	2.906878	3.552660	5.591742
Плотность, кг/м3	1023.0	1975.734888	73.729231	1731.764635	1924.155467	1977.621657	2021.374375	2207.773481
модуль упругости, ГПа	1023.0	739.923233	330.231581	2.436909	500.047452	739.664328	961.812526	1911.536477
Количество отвердителя, м.%	1023.0	110.570769	28.295911	17.740275	92.443497	110.564840	129.730366	198.953207
Содержание эпоксидных групп,%_2	1023.0	22.244390	2.406301	14.254985	20.608034	22.230744	23.961934	33.000000
Температура вспышки, С_2	1023.0	285.882151	40.943260	100.000000	259.066528	285.896812	313.002106	413.273418
Поверхностная плотность, г/м2	1023.0	482.731833	281.314690	0.603740	266.816645	451.864365	693.225017	1399.542362
Модуль упругости при растяжении, ГПа	1023.0	73.328571	3.118983	64.054061	71.245018	73.268805	75.356612	82.682051
Прочность при растяжении, МПа	1023.0	2466.922843	485.628006	1036.856605	2135.850448	2459.524526	2767.193119	3848.436732
Потребление смолы, г/м2	1023.0	218.423144	59.735931	33.803026	179.627520	219.198882	257.481724	414.590628
Угол нашивки, град	1023.0	44.252199	45.015793	0.000000	0.000000	0.000000	90.000000	90.000000
Шаг нашивки	1023.0	6.899222	2.563467	0.000000	5.080033	6.916144	8.586293	14.440522
Плотность нашивки	1023.0	57.153929	12.350969	0.000000	49.799212	57.341920	64.944961	103.988901

Гистограммы и ящики с усами до предобработки



Распределение выбросов по характеристикам

0 -> выбросов в признаке: 'Соотношение матрица-наполнитель'

3 -> выбросов в признаке: 'Плотность, кг/м3'

2 -> выбросов в признаке: 'модуль упругости, ГПа'

2 -> выбросов в признаке: 'Количество отвердителя, м.%'

2 -> выбросов в признаке: 'Содержание эпоксидных групп, %_2'

3 -> выбросов в признаке: 'Температура вспышки, С_2'

2 -> выбросов в признаке: 'Поверхностная плотность, г/м2'

0 -> выбросов в признаке: 'Модуль упругости при растяжении, ГПа'

0 -> выбросов в признаке: 'Прочность при растяжении, МПа'

3 -> выбросов в признаке: 'Потребление смолы, г/м2'

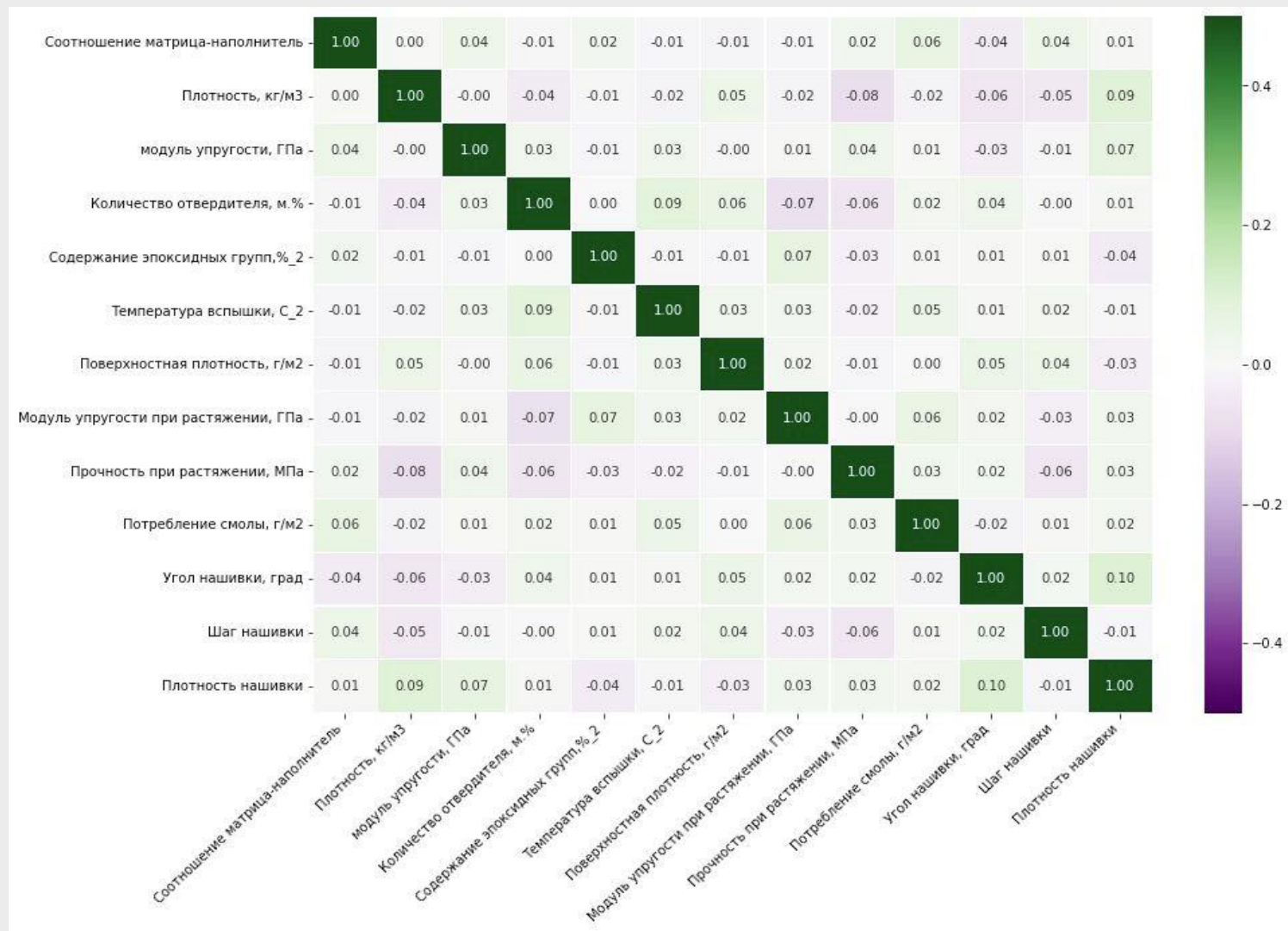
0 -> выбросов в признаке: 'Угол нашивки, град'

0 -> выбросов в признаке: 'Шаг нашивки'

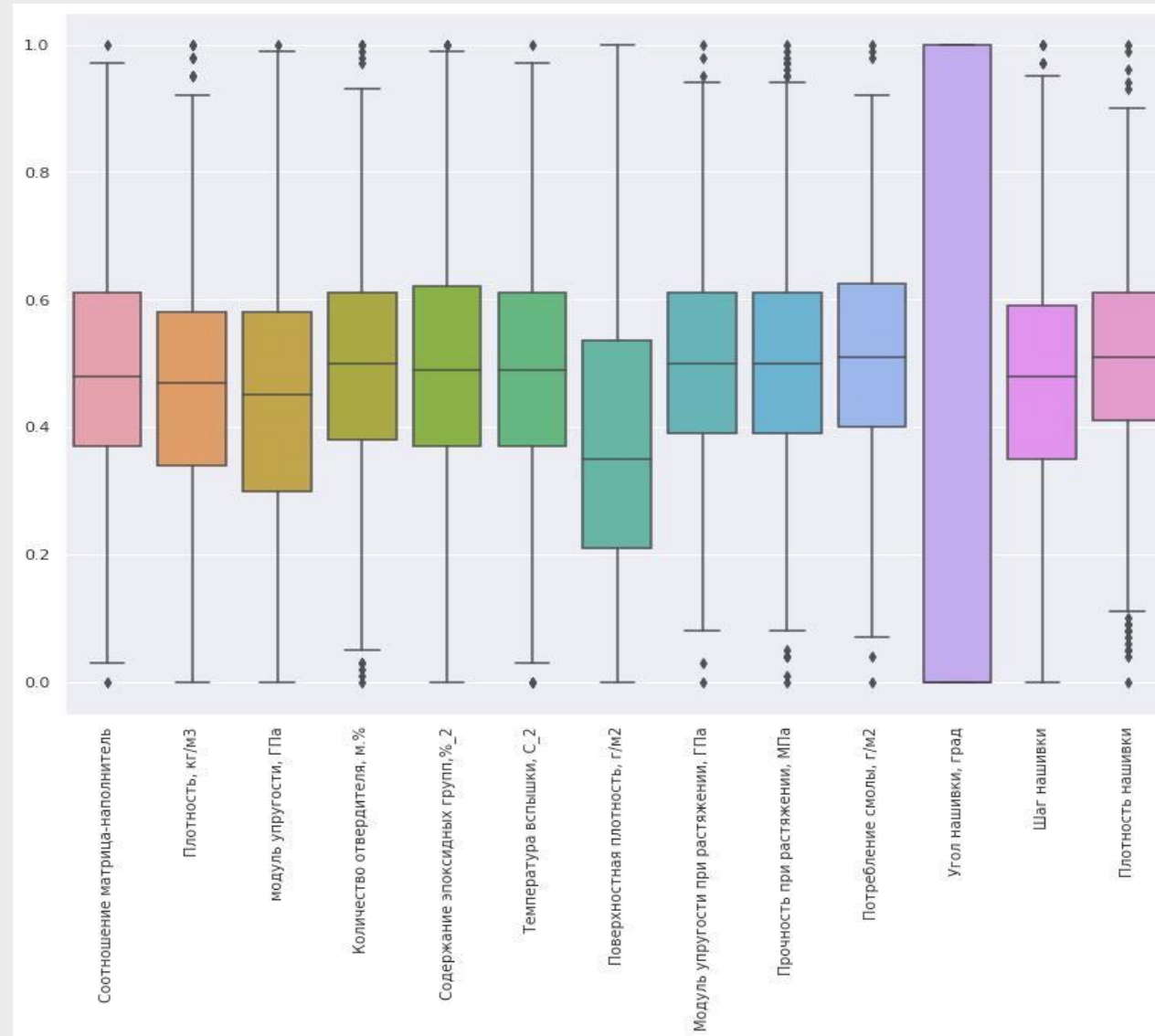
7 -> выбросов в признаке: 'Плотность нашивки'

Всего - 24 выброса

Тепловая карта и корреляция переменных



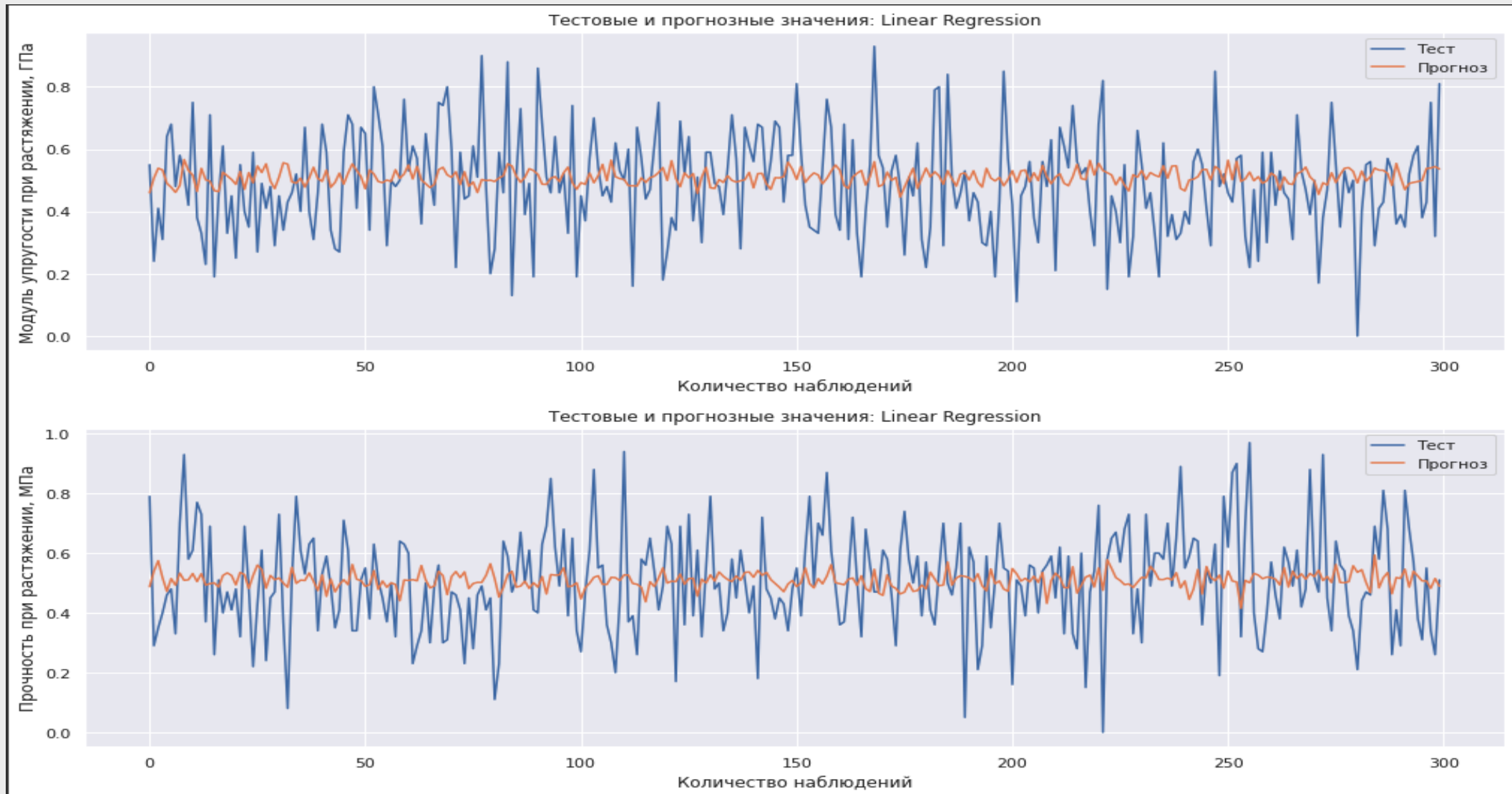
«Ящики с усами» после нормализации



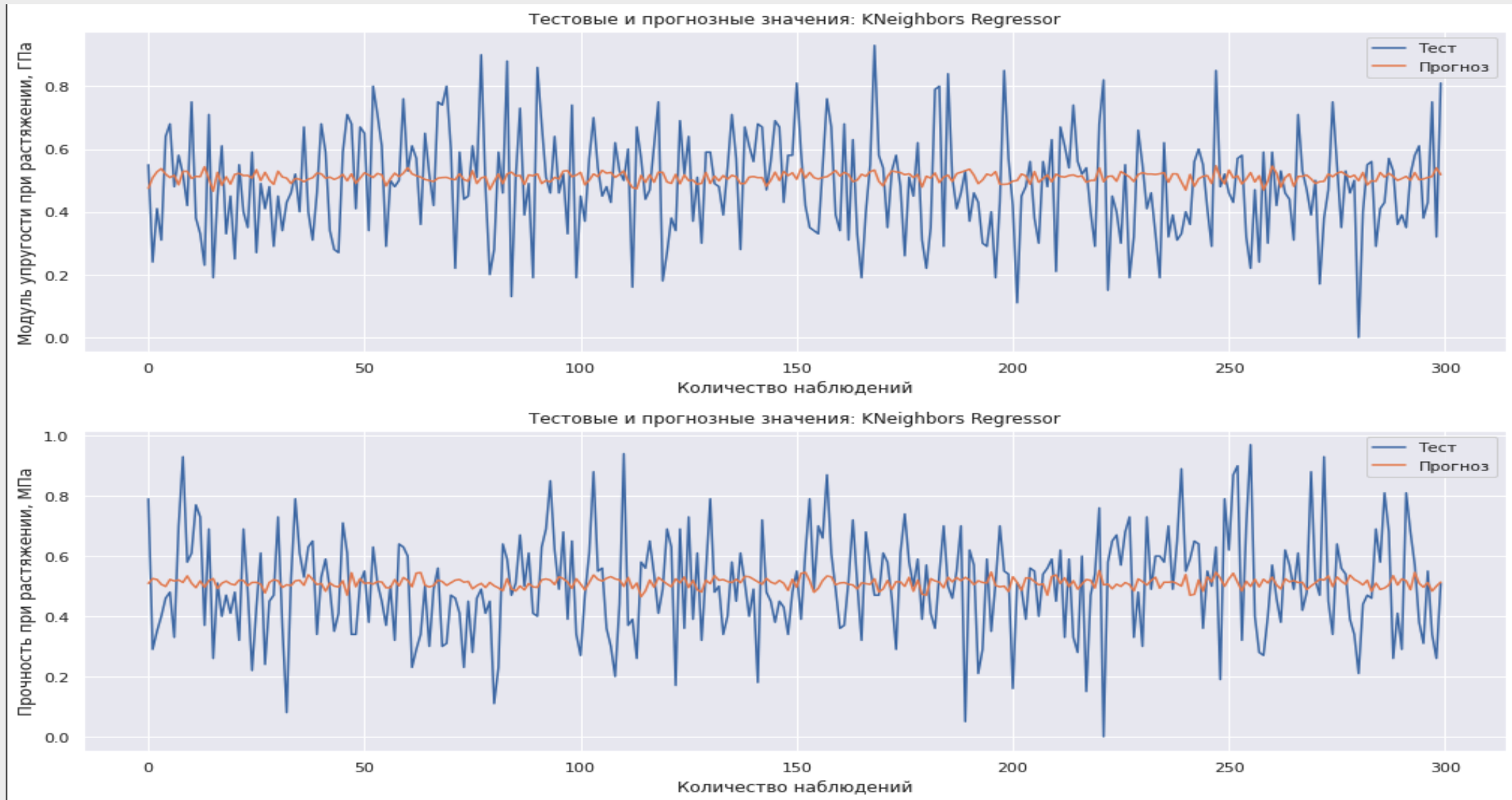
Описательная статистика характеристик после нормализации и удаления выбросов

index	count	mean	std	min	25%	50%	75%	max
Соотношение матрица-наполнитель	999.0	0.49	0.17	0.0	0.37	0.48	0.61	1.0
Плотность, кг/м3	999.0	0.47	0.18	0.0	0.34	0.47	0.58	1.0
модуль упругости, ГПа	999.0	0.45	0.2	0.0	0.3	0.45	0.58	1.0
Количество отвердителя, м. %	999.0	0.5	0.17	0.0	0.38	0.5	0.61	1.0
Содержание эпоксидных групп, %_2	999.0	0.49	0.18	0.0	0.37	0.49	0.62	1.0
Температура вспышки, С_2	999.0	0.49	0.17	0.0	0.37	0.49	0.61	1.0
Поверхностная плотность, г/м2	999.0	0.37	0.22	0.0	0.21	0.35	0.54	1.0
Модуль упругости при растяжении, ГПа	999.0	0.5	0.17	0.0	0.39	0.5	0.61	1.0
Прочность при растяжении, МПа	999.0	0.51	0.17	0.0	0.39	0.5	0.61	1.0
Потребление смолы, г/м2	999.0	0.51	0.17	0.0	0.4	0.51	0.62	1.0
Угол нашивки, град	999.0	0.5	0.5	0.0	0.0	0.0	1.0	1.0
Шаг нашивки	999.0	0.48	0.18	0.0	0.35	0.48	0.59	1.0
Плотность нашивки	999.0	0.51	0.16	0.0	0.41	0.51	0.61	1.0

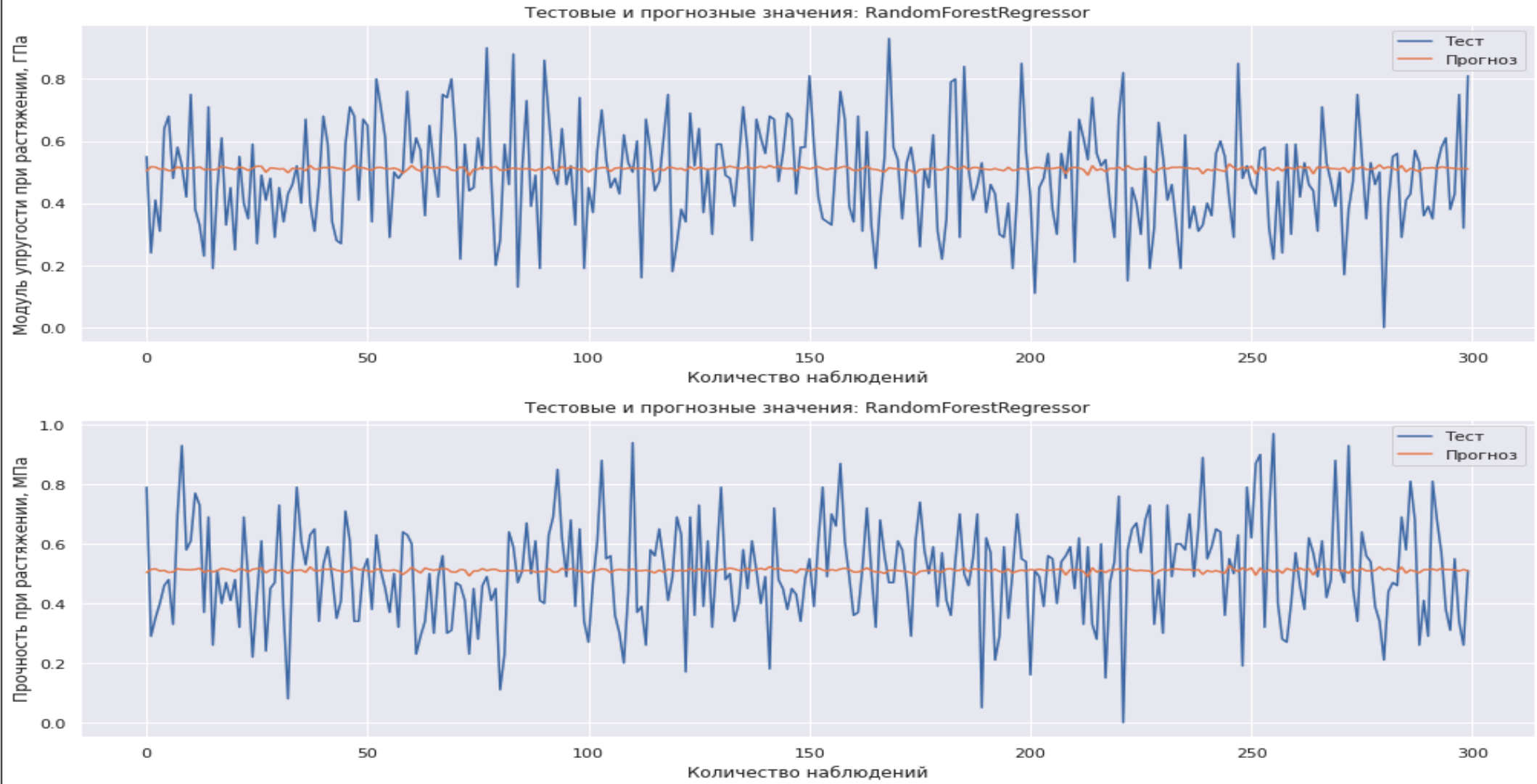
Линейная регрессия



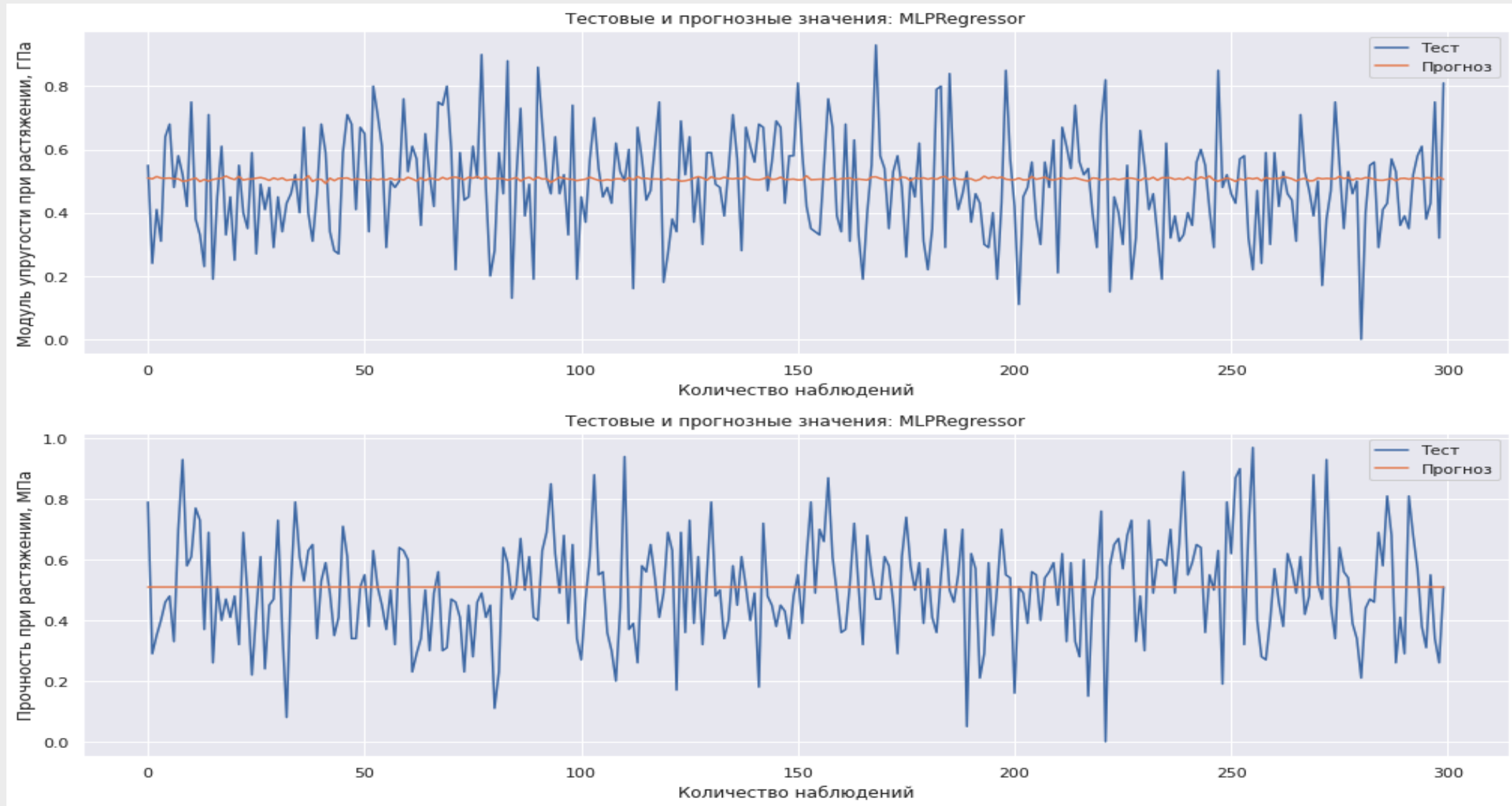
Регрессия k-ближайших соседей



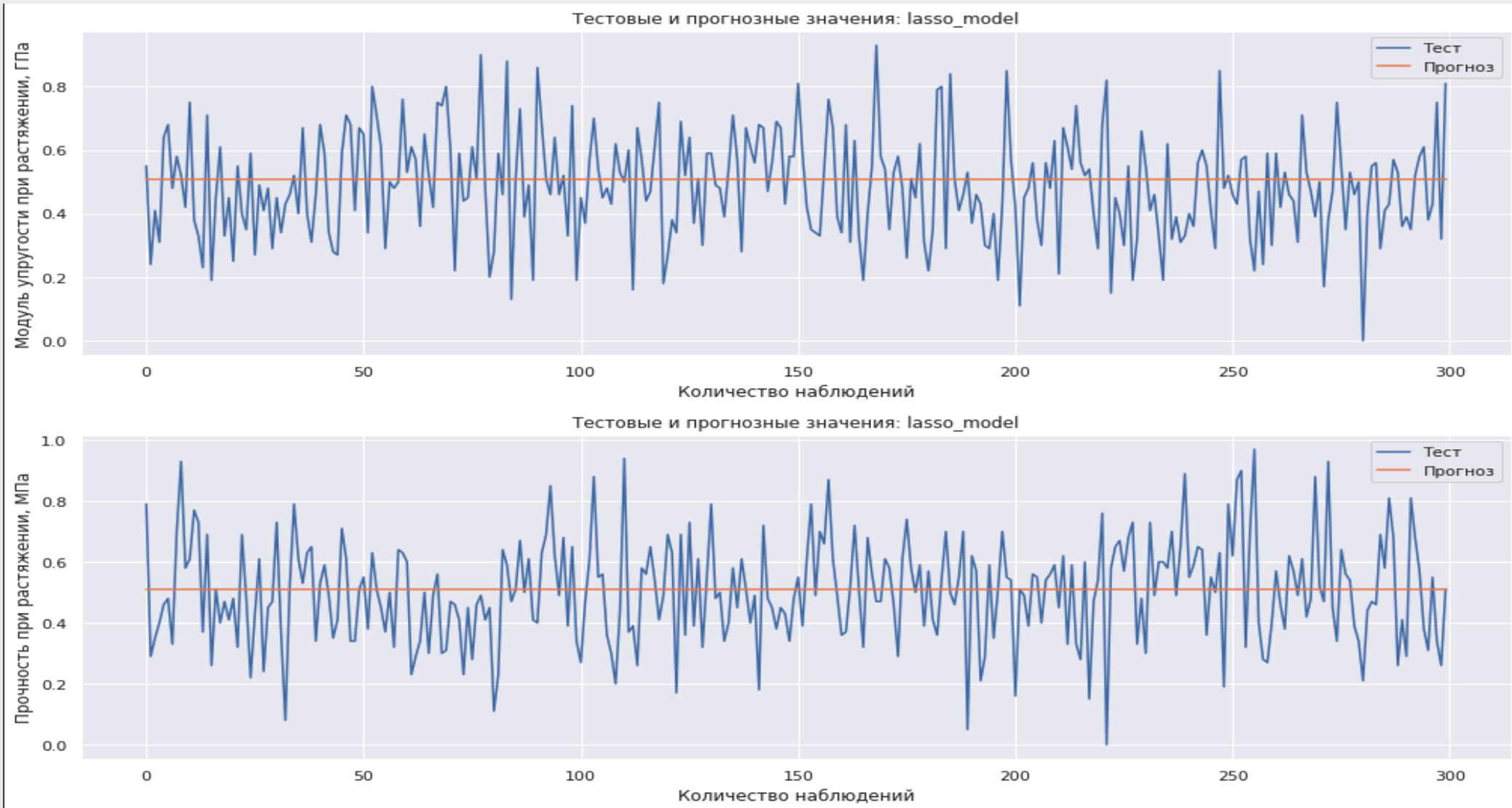
Случайный лес



Многослойный перцептрон



Лассо регрессия



Результаты обучения

index	target_var	model_name	MSE	R2
0	Модуль упругости при растяжении, ГПа	Linear Regression	0.02798367226727433	-0.027799467409495682
1	Прочность при растяжении, МПа	Linear Regression	0.02882712765798571	-0.012278080179114692
2	Модуль упругости при растяжении, ГПа	KNeighborsRegressor	0.027689151925078044	-0.016982164806770284
3	Прочность при растяжении, МПа	KNeighborsRegressor	0.02888706698854738	-0.014382877826249807
4	Модуль упругости при растяжении, ГПа	RandomForestRegressor	0.027616678542793847	-0.014320323179923822
5	Прочность при растяжении, МПа	RandomForestRegressor	0.028468337706038273	0.00032099691430820254
6	Модуль упругости при растяжении, ГПа	MLPRegressor	0.027745752151894395	-0.01906100858472337
7	Прочность при растяжении, МПа	MLPRegressor	0.028503449010552653	-0.0009119529774772595
8	Модуль упругости при растяжении, ГПа	lasso_model	0.027706436085886032	-0.017616986823816516
9	Прочность при растяжении, МПа	lasso_model	0.028506689898301473	-0.001025758267666932

Модели показали неудовлетворительный результат.
Если результат отрицательный, наша модель не так хороша, как догадки.

Нейронная сеть, которая будет рекомендовать соотношение матрица-наполнитель

```
[160] model = tf.keras.Sequential([x_train_norm, layers.Dense(128, activation='relu'),  
                                  layers.Dense(128, activation='relu'),  
                                  layers.Dense(128, activation='relu'),  
                                  layers.Dense(64, activation='relu'),  
                                  layers.Dense(32, activation='relu'),  
                                  layers.Dense(16, activation='relu'),  
                                  layers.Dense(1)  
                                  ])
```

```
model.compile(optimizer=tf.keras.optimizers.Adam(0.001), loss='mean_squared_error')
```

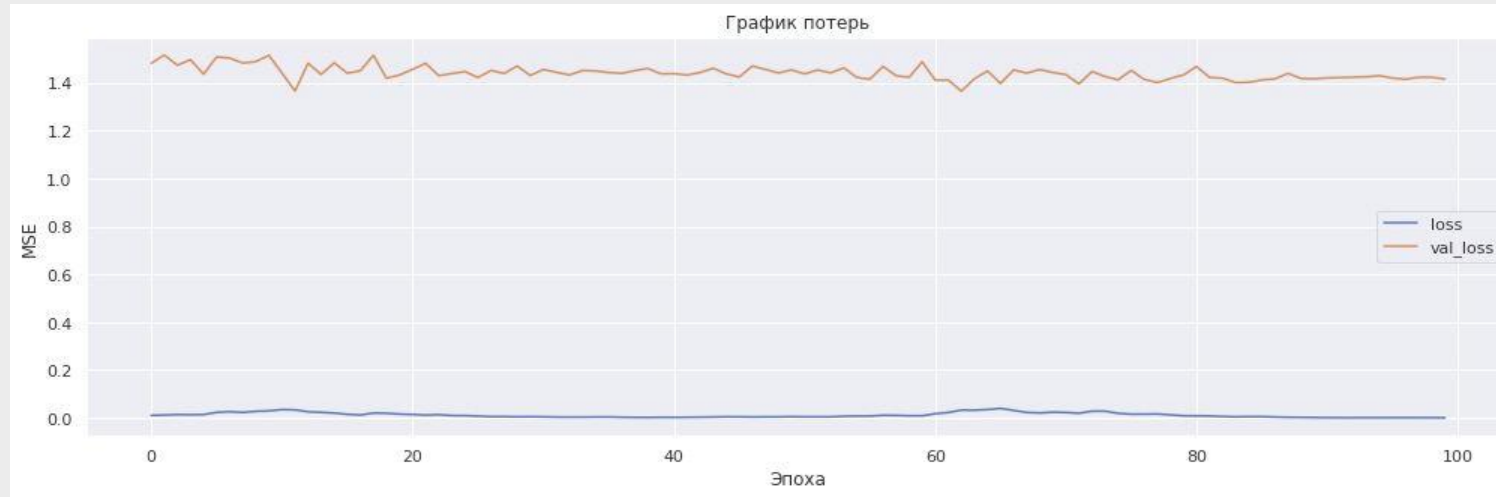
```
[161] model.summary()
```

Model: "sequential"

Layer (type)	Output Shape	Param #
=====		
normalization (Normalization)	(None, 12)	25
dense (Dense)	(None, 128)	1664
dense_1 (Dense)	(None, 128)	16512
dense_2 (Dense)	(None, 128)	16512
dense_3 (Dense)	(None, 64)	8256
dense_4 (Dense)	(None, 32)	2080
dense_5 (Dense)	(None, 16)	528
dense_6 (Dense)	(None, 1)	17

```
=====
Total params: 45,594
Trainable params: 45,569
Non-trainable params: 25
```

Визуализация тест/прогноз и график потерь модели (MSE) показаны ниже.



Ошибки модели: $MSE = 1.1775$,
 $R^2 = -0.5459$. Результаты
неудовлетворительны.

Заключительные выводы

- В ходе выполнения данной работы было выполнено:
 - изучение теоретических методов анализа данных и машинного обучения;
 - разведочный анализ данных;
 - предобработка данных;
 - построение регрессионных моделей;
 - визуализация модели и оценка качества прогноза;
- Использованные при разработке моделей подходы не позволили получить достоверных прогнозов. Возможные причины неудовлетворительной работы моделей и пути решения:
 - Необходима дополнительная информации о зависимости признаков с точки зрения физики процесса.
 - Возможно, исследование предварительно обработанных данных, не позволяет построить качественные модели на этом датасете.
- На основании проведенного исследования можно сделать следующие основные выводы по теме:
 - распределение полученных данных близко к нормальному;
 - коэффициенты корреляции между парами признаков стремятся к нулю.
- Считаю, что для неудовлетворительных результатов моделей нет смысла разрабатывать приложение.



edu.bmstu.ru

+7 495 182-83-85

edu@bmstu.ru

Москва, Госпитальный переулок , д.
4-6, с.3