



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

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА по курсу «Data Science»

Прогнозирование конечных свойств новых материалов (композиционных материалов)

Слушатель: Балакин Андрей Сергеевич

Постановка задачи

Необходимо спрогнозировать ряд конечных свойств получаемых композиционных материалов имея данные о начальных свойствах компонентов (количество связующего, наполнителя, температурный режим отверждения и т.д.).

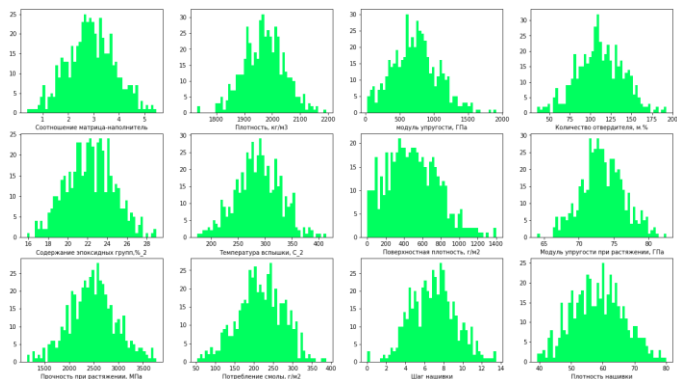
Для этого необходимо:

- 1) провести разведочный анализ предложенных данных;
- 2) провести предобработку данных;
- 3) обучить нескольких моделей для прогноза модуля упругости при растяжении и прочности при растяжении;
- 4) написать нейронную сеть, которая будет рекомендовать соотношение матрица-наполнитель;
- 5) разработать приложение с графическим интерфейсом или интерфейсом командной строки, которое будет выдавать прогноз.

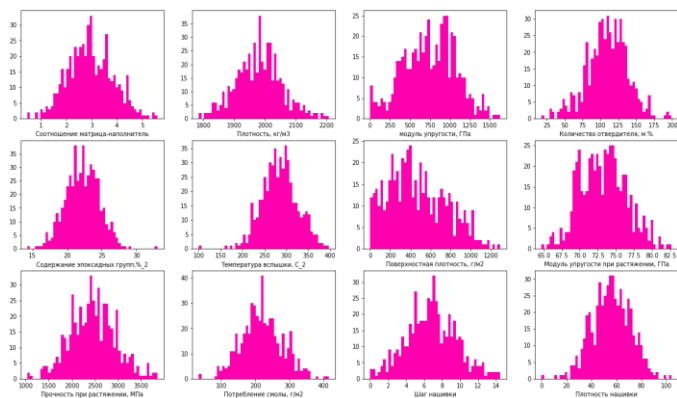
Статистические сведения о данных

	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

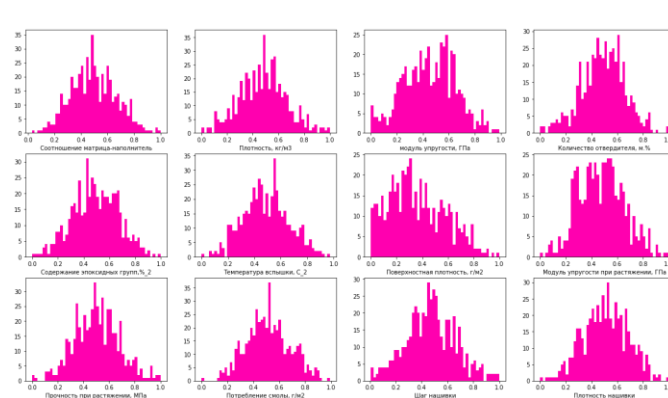
Гистограмма распределения данных



угол нашивки
90 градусов



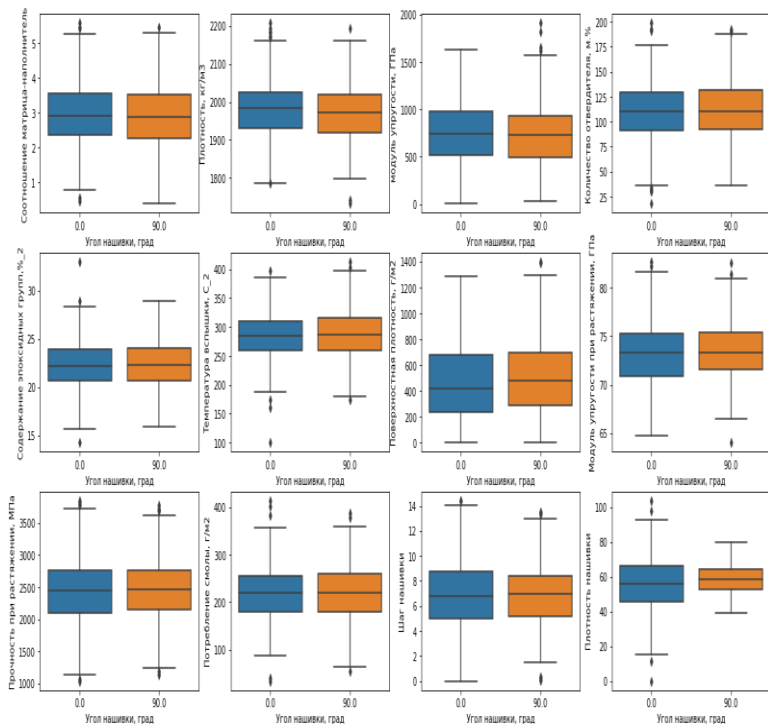
угол нашивки
0 градусов



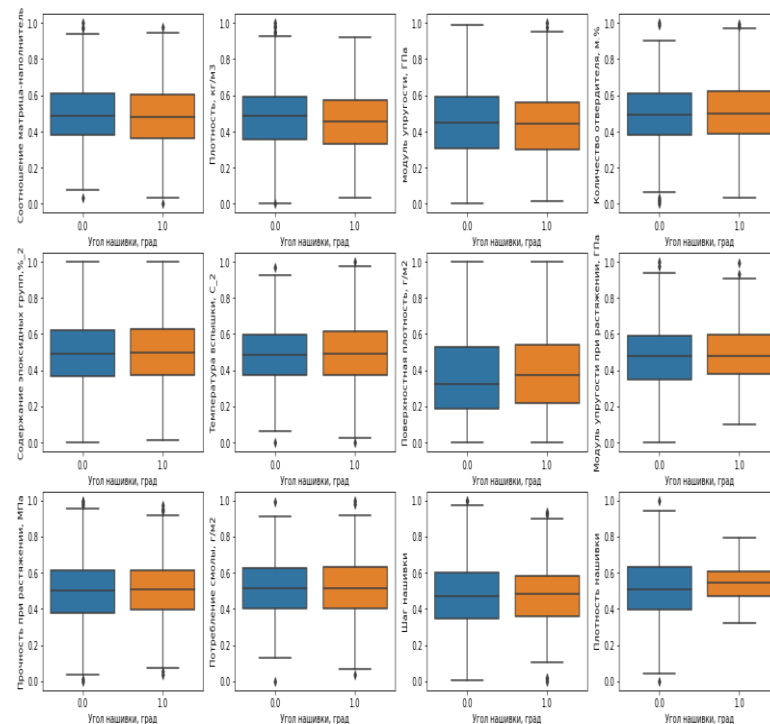
данные до предобработки

данные после предобработки

«Ящик с усами»

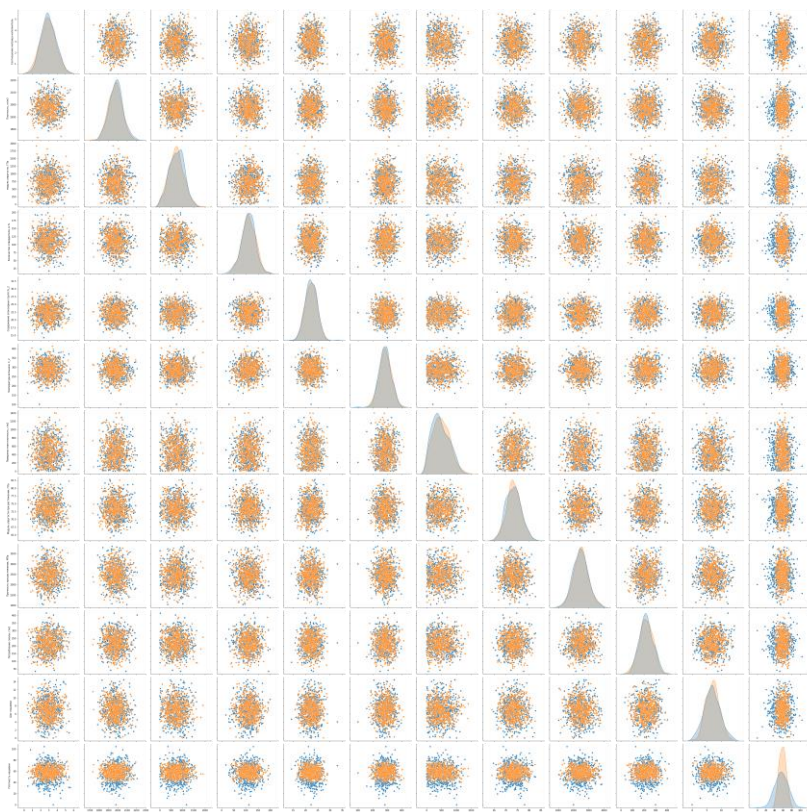


данные до предобработки

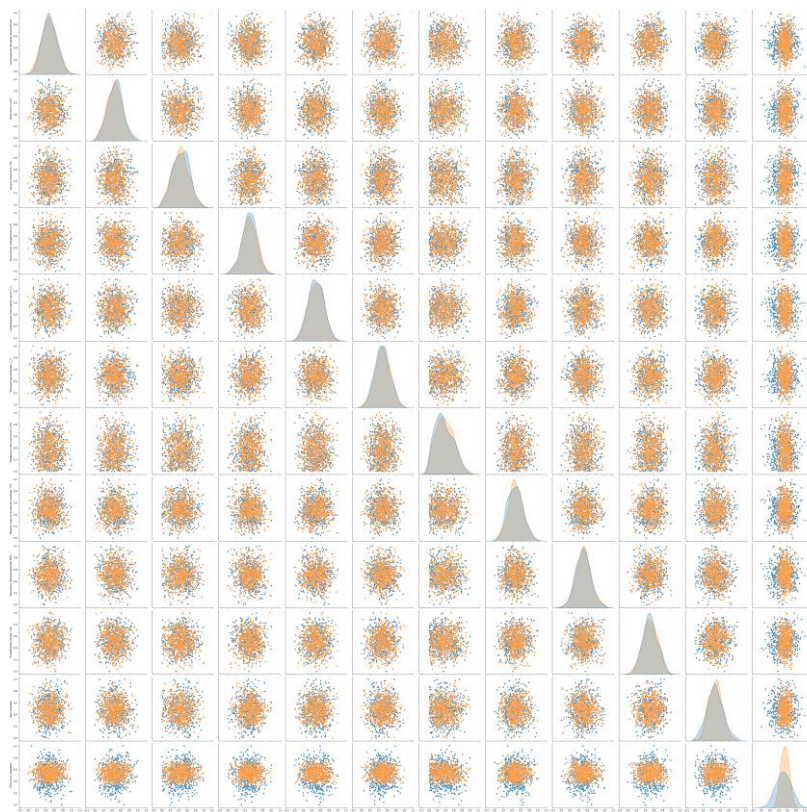


данные после предобработки

Попарная диаграмма рассеяния

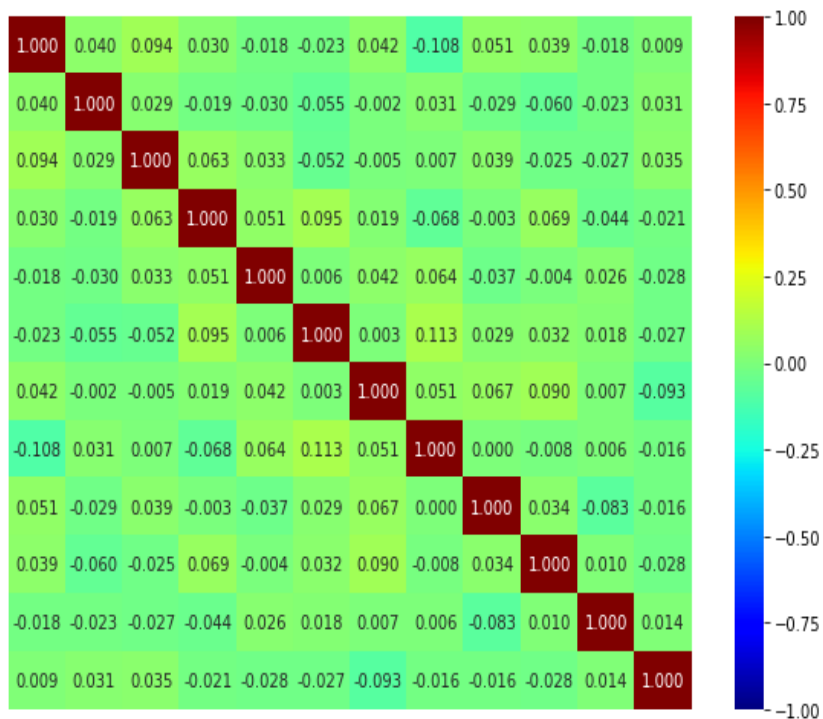


данные до предобработки

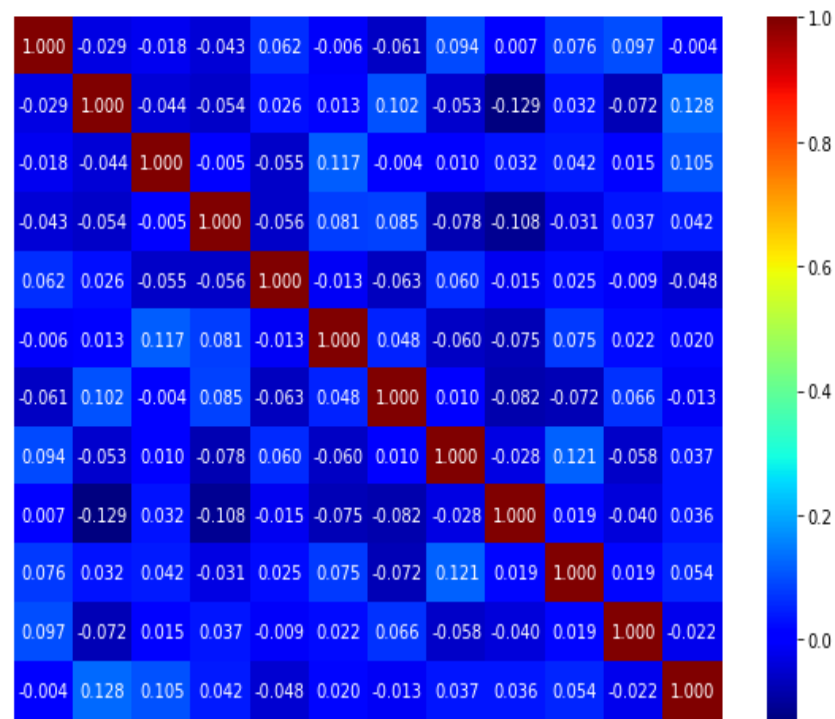


данные после предобработки

Корреляционная матрица

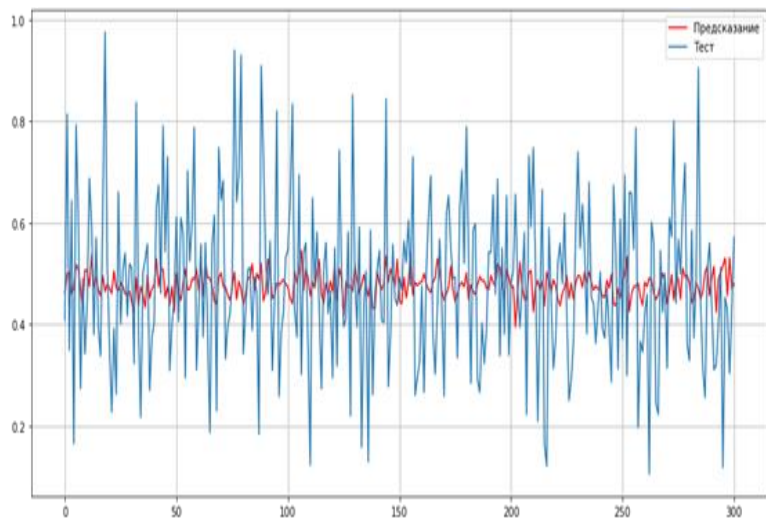


угол нашивки 90 градусов

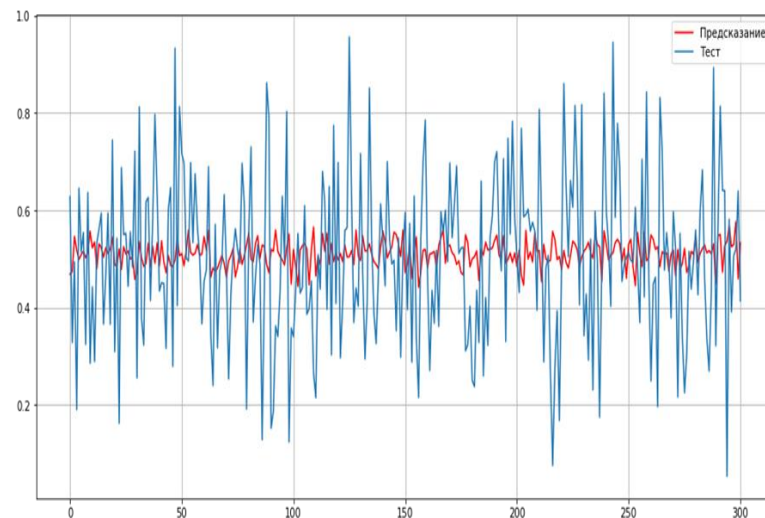


угол нашивки 0 градусов

Результаты применения модели линейной регрессии для прогноза

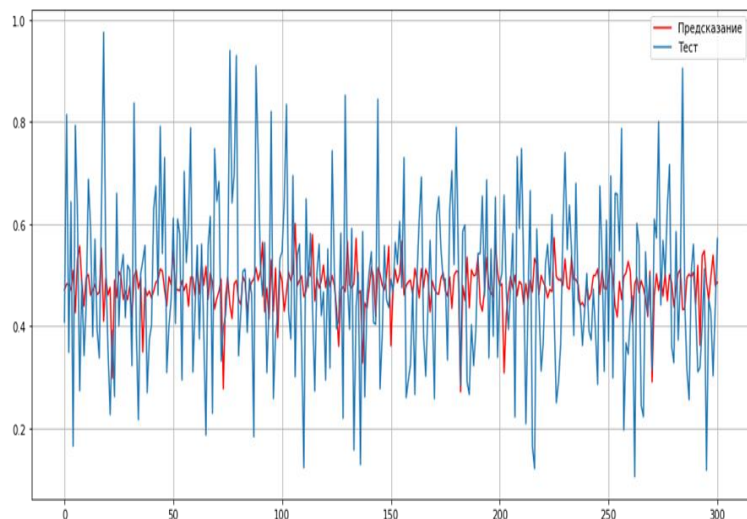


модуля упругости при растяжении

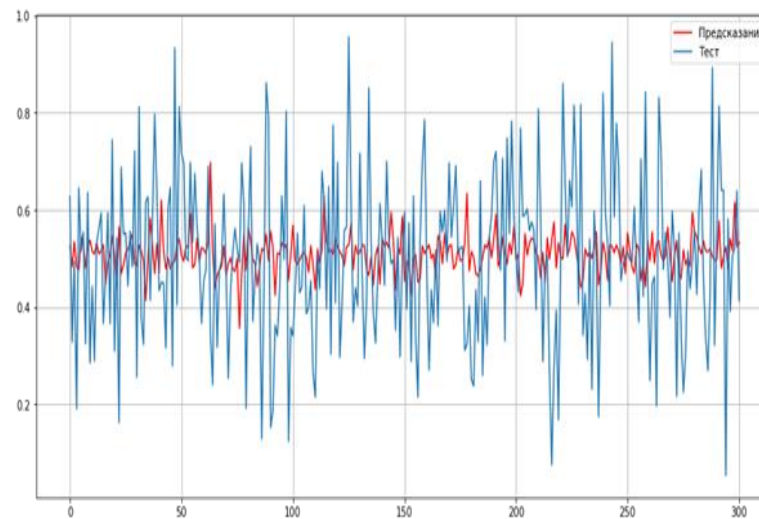


прочности при растяжении

Результаты применения модели градиентного бустинга для прогноза

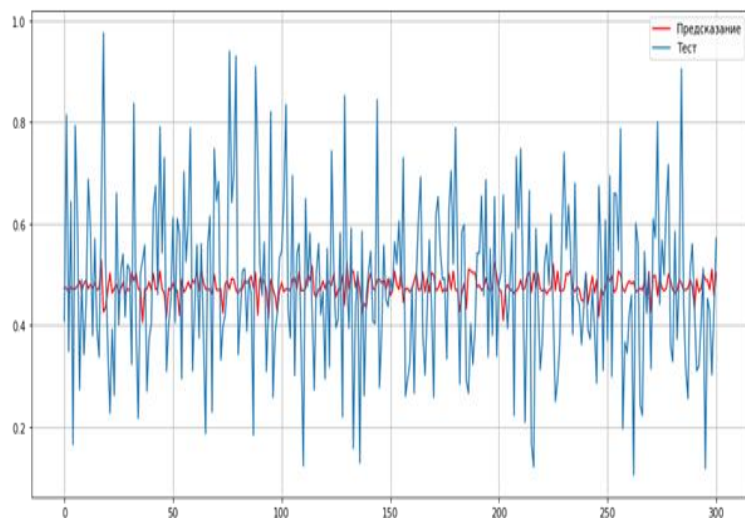


модуля упругости при растяжении

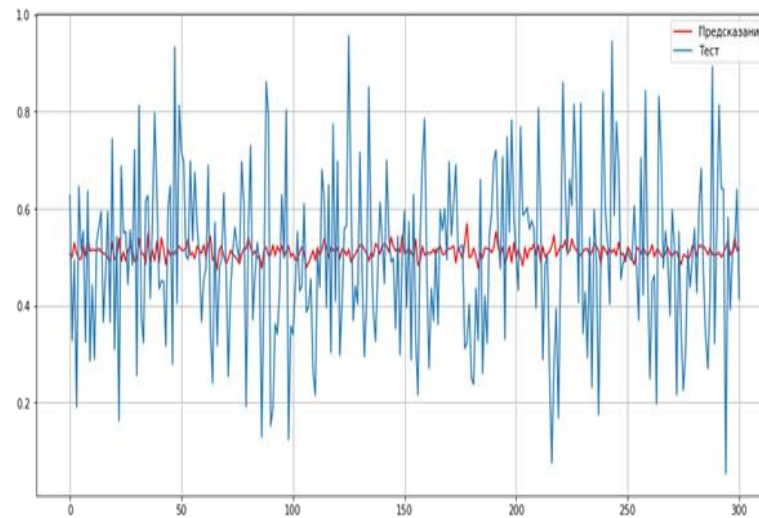


прочности при растяжении

Результаты применения модели случайного леса для прогноза

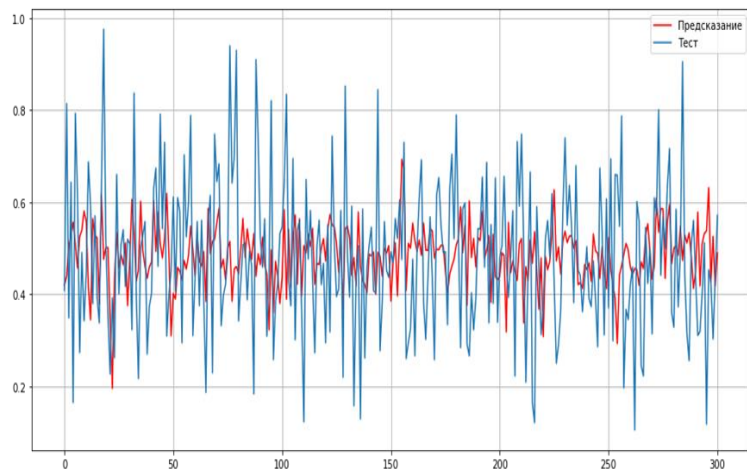
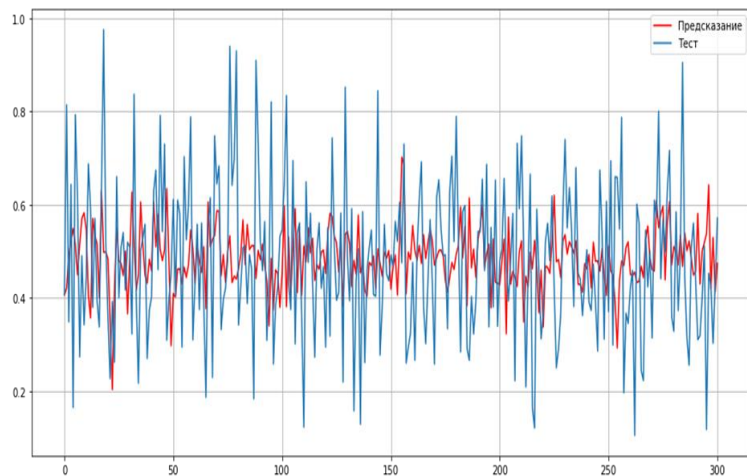


модуля упругости при растяжении

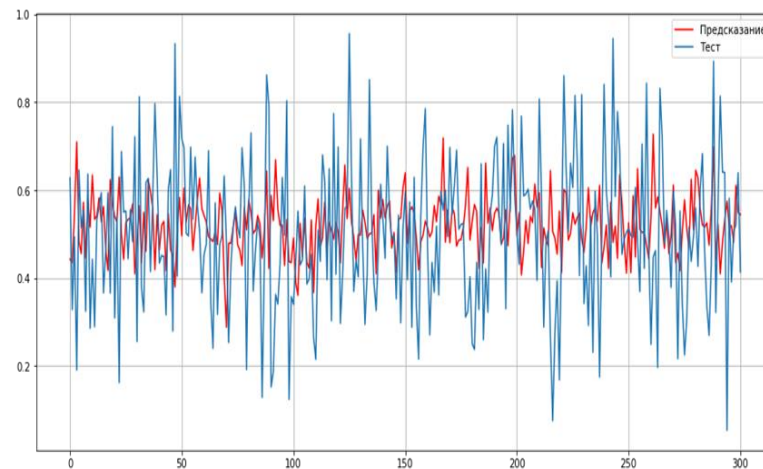
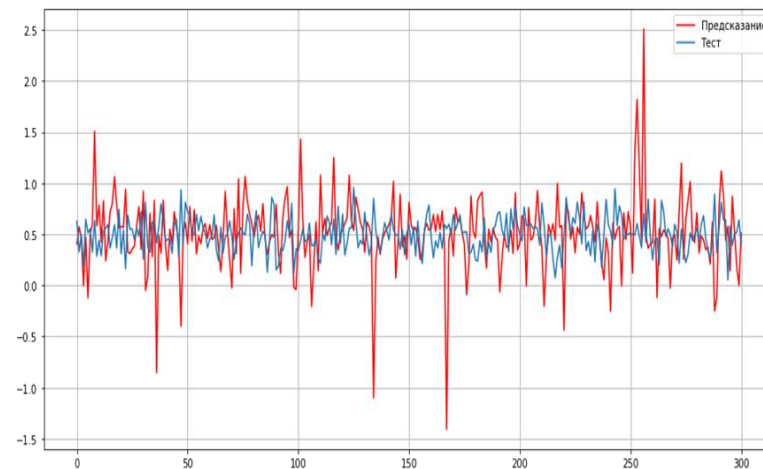


прочности при растяжении

Результаты применения модели полиномиальной регрессии для прогноза



модуля упругости при растяжении



прочности при растяжении

Результаты тестирования моделей:

Модель	Прогноз модуля упругости		Прогноз прочности при растяжении	
	MSE	R ²	MSE	R ²
1	2	3	4	5
Линейная регрессия	0.027345286816019446	0.00031270815893347237	0.029649401268432553	-0.029467652046270665
Градиентный бустинг	0.029467297660744712	-0.07726363210377118	0.02983422117720099	-0.035884851361989156
Случайный лес	0.027802607282353512	-0.016405984958784092	0.028720337795920776	0.0027906989062314036
Полиномиальная регрессия (преобразование в функции взаимодействия)	0.029053128732401397	-0.06212247022362982		
Полиномиальная регрессия (преобразование во 2-ю степень)	0.029346061183560812	-0.07283147652377098	0.03298482318652948	-0.14527805035818364
Полиномиальная регрессия (преобразование в 3-ю степень)			0.15914794481839756	-4.525833712352958

Построение нейронной сети

Построена нейронная сеть для расчета рекомендованных соотношений «матрица-наполнитель» со следующими параметрами:

- 1) входной слой нормализации 12 признаков;
- 2) выходной слой для 1 признака;
- 3) скрытых слоев: 2;
- 4) нейронов в скрытых слоях: 64 в каждом слое;
- 5) активационная функция скрытых слоев: relu;
- 6) оптимизатор: Adam;
- 7) loss-функция: MeanAbsoluteError.

Структура модели:

Model: "sequential"

Layer (type)	Output Shape	Param #
normalization (Normalization)	(None, 12)	25
dense (Dense)	(None, 64)	832
dense_1 (Dense)	(None, 64)	4160
dense_2 (Dense)	(None, 1)	65

Total params: 5,082

Trainable params: 5,057

Non-trainable params: 25

Визуализация результатов нейросети

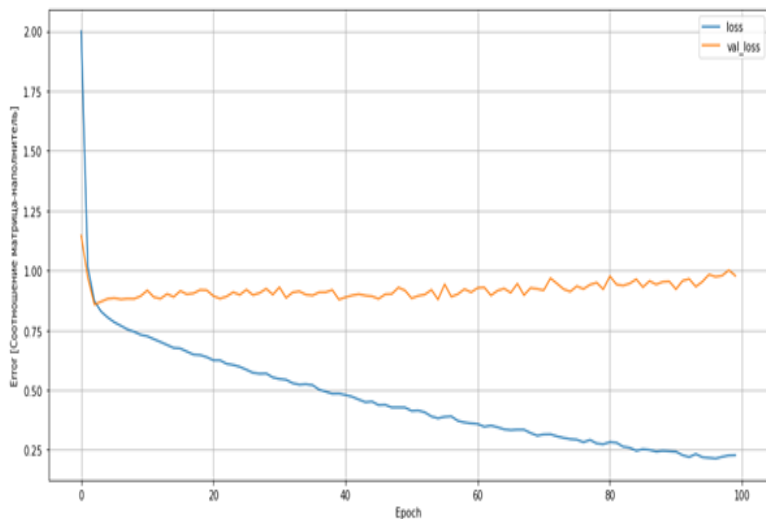


График ошибки обучения сети

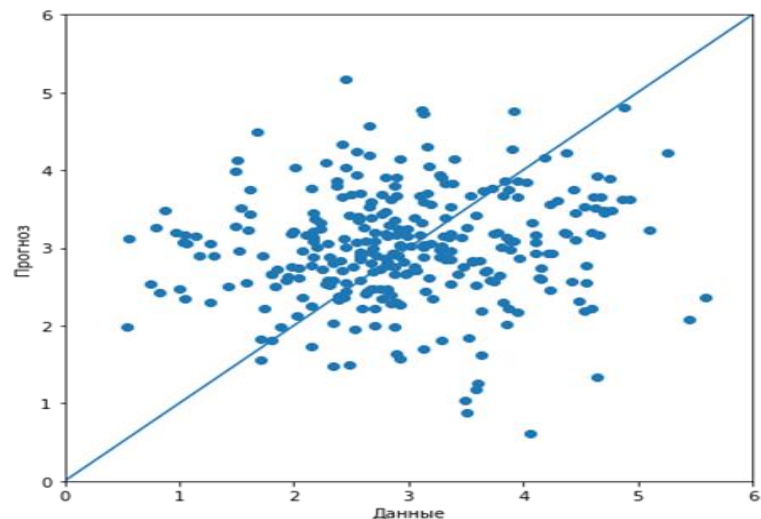


График сравнения прогноза с данными валидации

Заключение

На основании проведенного исследования можно сделать следующие основные выводы:

- коэффициенты корреляции между парами признаков стремятся к нулю;
- примененные модели не показали высокой эффективности в прогнозировании свойств композитов;
- необходимы дополнительные входные данные для улучшения моделей.

Исходя из основных выводов, требуется вернуться на этап сбора данных, более тщательно оценить точность и качество входных данных, кол-во входных параметров и исследовать их содержательно, в том числе выявив наиболее важные переменные, после чего разработать новые гипотезы, выбрать и построить оптимальные модели.



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

СПАСИБО ЗА ВНИМАНИЕ!

