



ЦЕНТР
ДОПОЛНИТЕЛЬНОГО
ОБРАЗОВАНИЯ
МГТУ им. Н.Э. Баумана

Выпускная квалификационная работа по курсу "Data Science Pro"

Тема: Прогнозирование конечных свойств
новых материалов (композиционных материалов)

Слушатель: Нурукулиев Егор Игоревич



Постановка задачи

- изучить предметную область
- провести разведочный анализ данных
- разделить данные на тренировочную и тестовую выборки
- выполнить предобработку
- выбрать базовые модели
- сравнить модели с гиперпараметрами по умолчанию
- подобрать гиперпараметры с помощью поиска по сетке с перекрестной проверкой
- спроектировать нейронную сеть
- сравнить качество лучшей модели на тренировочной и тестовой выборке
- разработать приложение



анализ данных

X_br (матрица из базальтопластика):

- признаков: 10 и индекс
- строк: 1023

X_nir (наполнитель из углепластика):

- признаков: 3 и индекс
- строк: 1040

Объединение с типом INNER по индексу, получилось:

- признаков: 13
- строк: 1023



Разведочный анализ данных

1 Пропусков нет

2 Типы данных
int64 и float64

№	Параметр	Количество строк	Тип данных
1	2	3	4
1	Соотношение матрица-наполнитель	1023	float64
2	Плотность, кг/м ³	1023	float64
3	модуль упругости, ГПа	1023	float64
4	Количество отвердителя, м.%	1023	float64
5	Содержание эпоксидных групп, %_2	1023	float64
6	Температура вспышки, С_2	1023	float64
7	Поверхностная плотность, г/м ²	1023	float64

8	Модуль упругости при растяжении, ГПа	1023	float64
9	Прочность при растяжении, МПа	1023	float64
10	Потребление смолы, г/м ²	1023	float64
11	Угол нашивки, град	1023	int64
12	Шаг нашивки	1023	float64
13	Плотность нашивки	1023	float64

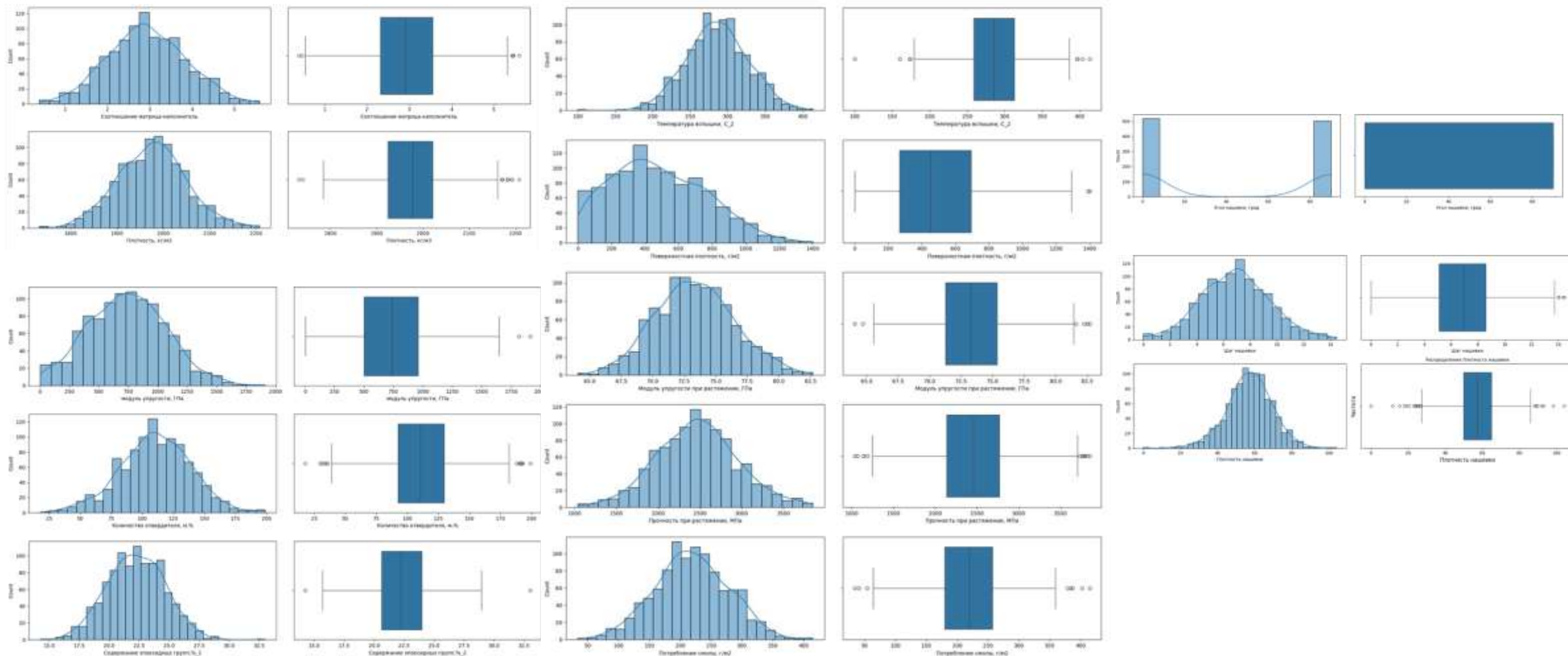


Статистические данные датасета

	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, C_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

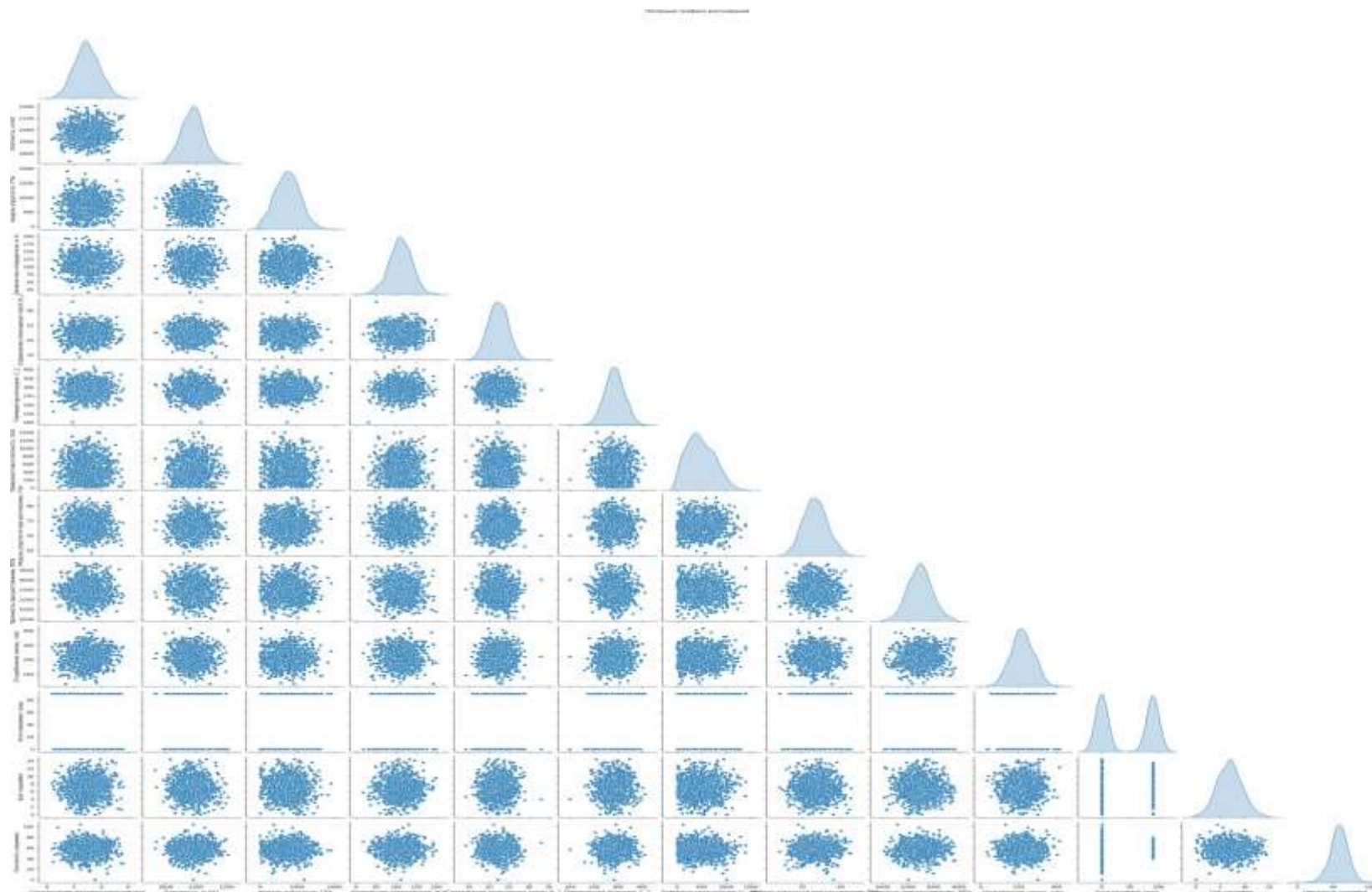


Гистограммы распределения и диаграммы “ящик с усами”





Попарный график рассеивания



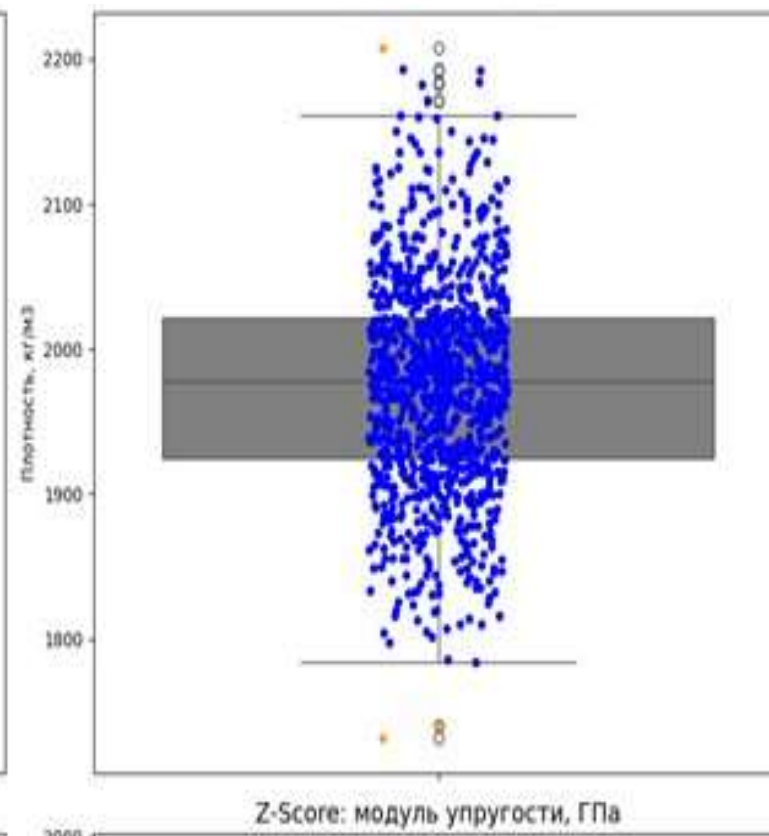
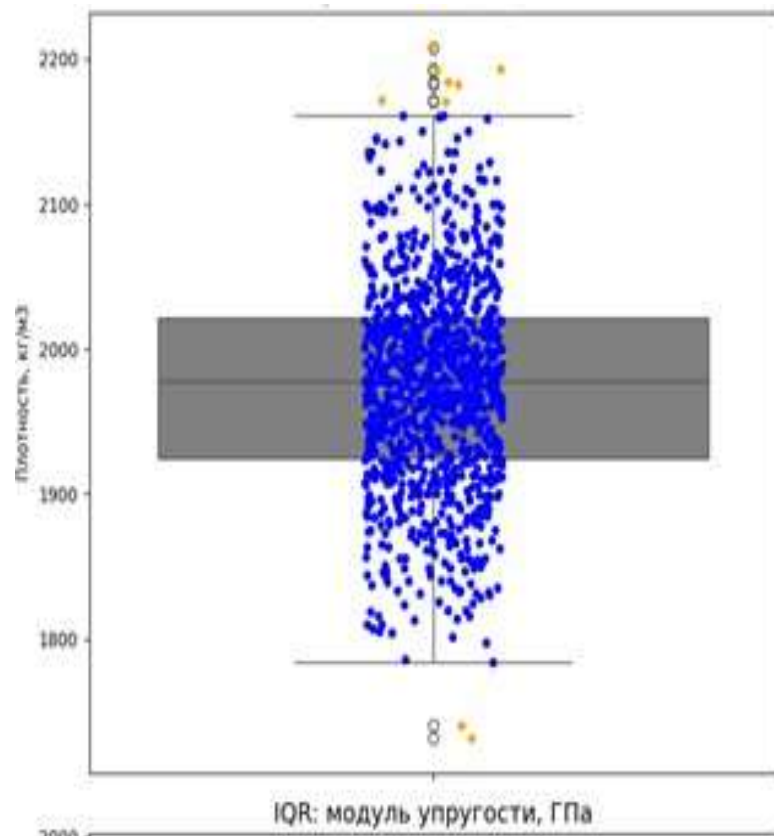


Визуализация выбросов

Найдено:

- методом Z-Score — 24 выброса
- методом IQR — 93 выброса

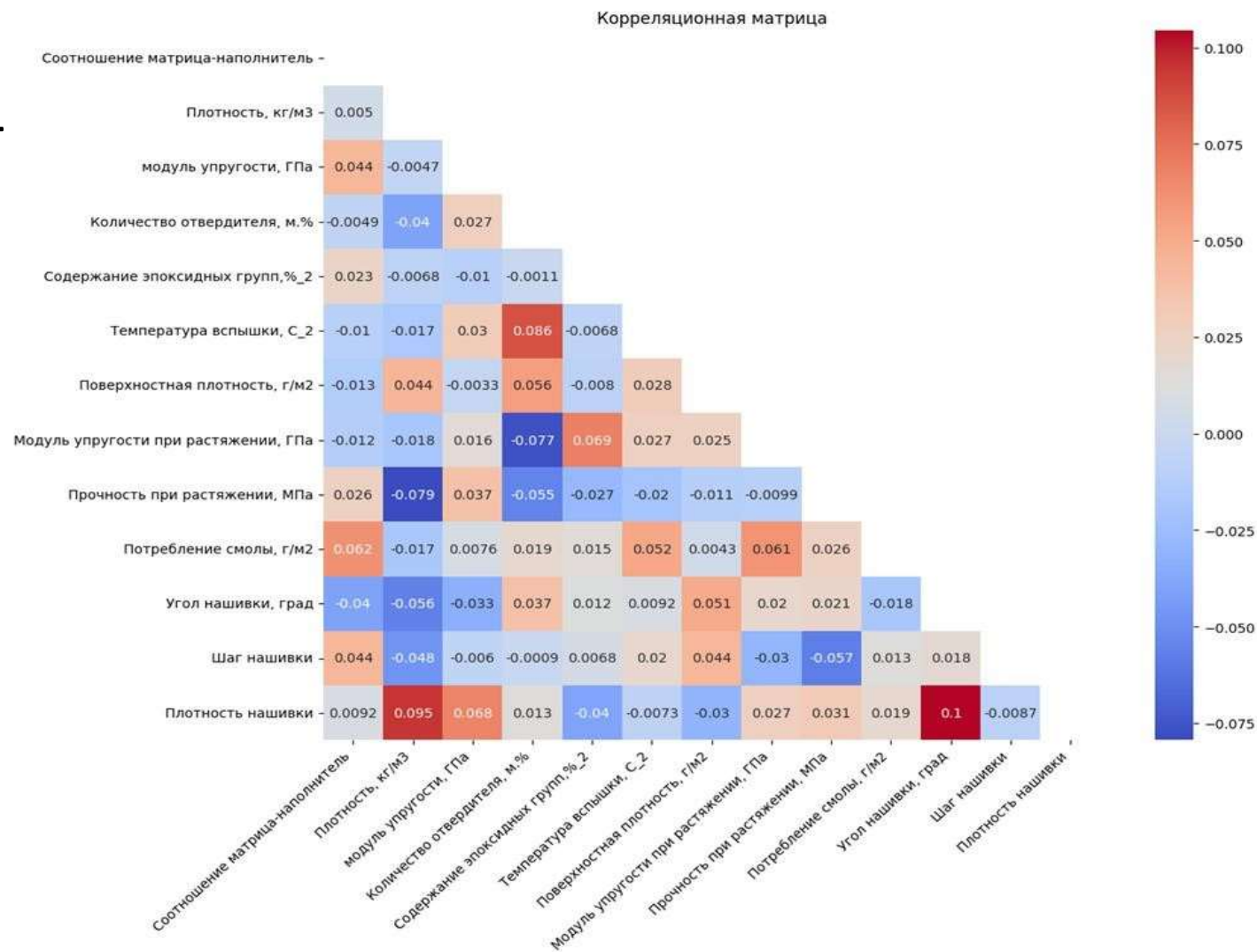
**Удалено 24 строки
осталось 1000 строк**





Матрица корреляции

Линейной зависимости нет





Модели

1. **RandomForestRegressor**
2. **К-ближайших соседей**
3. **GradientBoostingRegressor**
4. **Support Vector Regression, SVR**
5. **Нейросеть, рекомендующая в соотношение матрицы при помощи библиотеки TensorFlow**



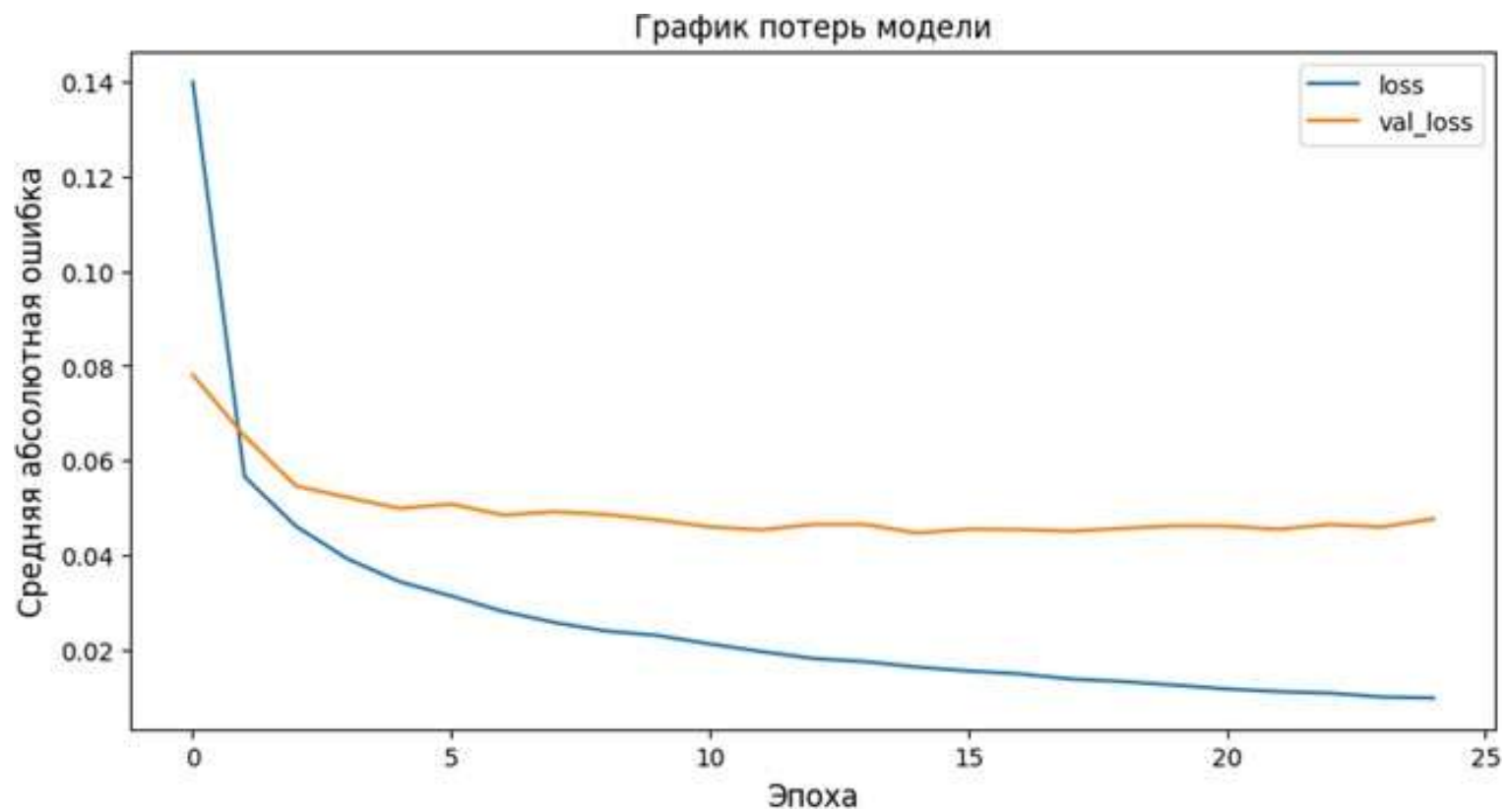
Метрики

Сравнение моделей

	MAE	MSE	R2
Случайный лес (без Модуль упругости)	0.138948	0.030037	-0.105227
Случайный лес (без Прочности)	0.134053	0.028761	-0.054115
GradientBoostingRegressor (без Модуль упругости)	0.141053	0.031211	-0.148409
GradientBoostingRegressor (без Прочность)	0.137829	0.029994	-0.099316
SVR (без Модуль упругости)	0.154055	0.037427	-0.377129
SVR (без Прочность)	0.155312	0.037279	-0.366320
KNN (без Модуль упругости)	0.144251	0.032005	-0.177618
KNN (без Прочность)	0.144932	0.033238	-0.218219



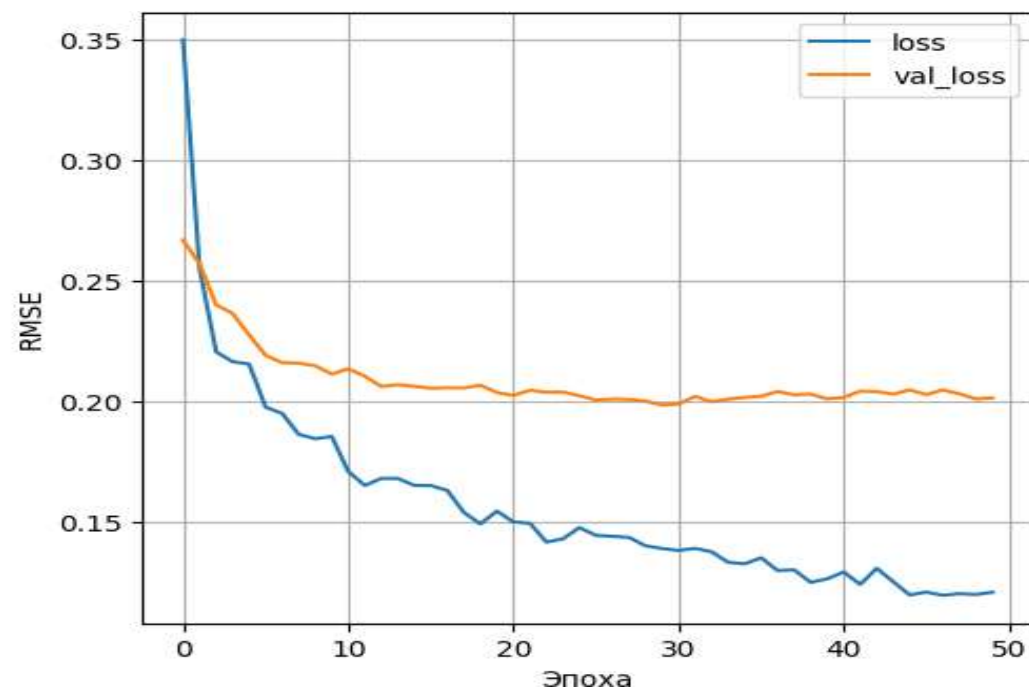
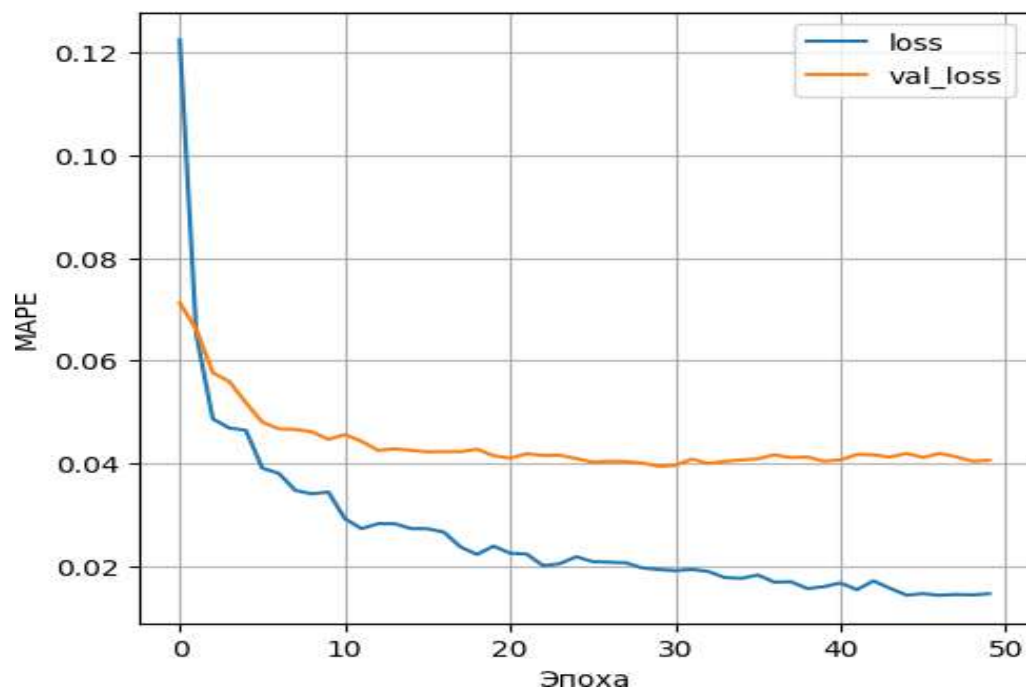
График обучения нейронной сети



Плохая работа модели



График обучения нейронной сети с Dropout



Переобучение модели примерно на 10 эпохе



ЦЕНТР
ДОПОЛНИТЕЛЬНОГО
ОБРАЗОВАНИЯ
МГТУ им. Н.Э. Баумана

Приложение Flask

Приложение размещено на Github:
<https://github.com/SamsungPipidon/-DS>

Прогнозирование

Плотность, кг/м^3

Модуль упругости, ГПа

Количество отвердителя, м.%

Содержание эпоксидных групп, %

Температура вспышки, $^{\circ}\text{C}$

Поверхностная плотность, г/м^2

Модуль упругости при растяжении, ГПа

Прочность при растяжении, МПа

Потребление смолы, г/м^2

Угол нашивки, град

Шаг нашивки

Плотность нашивки

Прогнозировать



ЦЕНТР
ДОПОЛНИТЕЛЬНОГО
ОБРАЗОВАНИЯ
МГТУ им. Н.Э. Баумана



do.bmstu.ru