



ЦЕНТР
ДОПОЛНИТЕЛЬНОГО
ОБРАЗОВАНИЯ
МГТУ им. Н.Э. Баумана

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА по курсу «Data Science Pro»

Прогнозирование конечных свойств новых материалов (композиционных материалов)

Калашникова Елена Александровна



ЦЕНТР
ДОПОЛНИТЕЛЬНОГО
ОБРАЗОВАНИЯ
МГТУ им. Н.Э. Баумана

Содержание

1 Разведочный анализ данных

2 Обучение моделей

3 Тестирование моделей

4 Нейронная сеть по рекомендации соотношения матрица-наполнитель

5 Разработка приложения



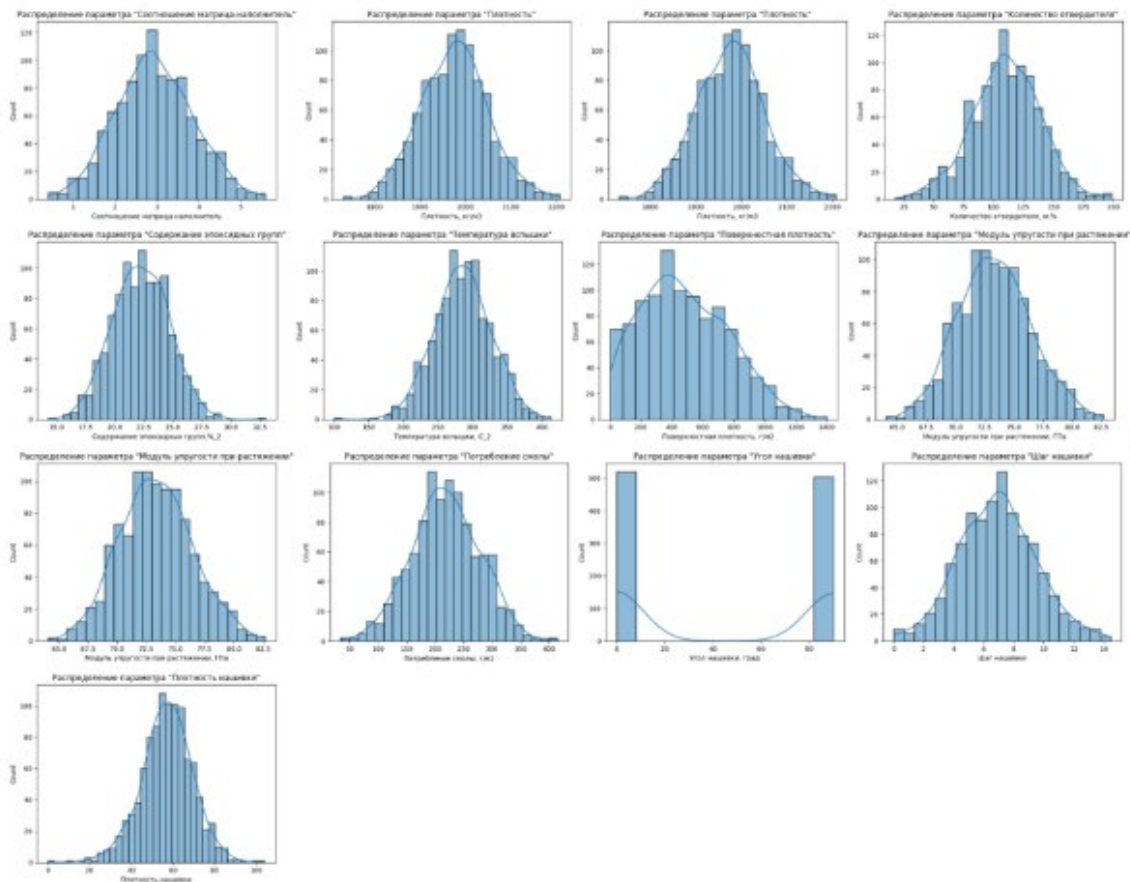
Анализ исходного датасета

	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	Угол нашивки, град	Шаг нашивки	Плотность нашивки
0	1.857143	2030.0	738.736842	30.00	22.267857	100.000000	210.0	70.0	3000.0	220.0	0	4.0	57.0
1	1.857143	2030.0	738.736842	50.00	23.750000	284.615385	210.0	70.0	3000.0	220.0	0	4.0	60.0
2	1.857143	2030.0	738.736842	49.90	33.000000	284.615385	210.0	70.0	3000.0	220.0	0	4.0	70.0
3	1.857143	2030.0	738.736842	129.00	21.250000	300.000000	210.0	70.0	3000.0	220.0	0	5.0	47.0
4	2.771331	2030.0	753.000000	111.86	22.267857	284.615385	210.0	70.0	3000.0	220.0	0	5.0	57.0
5	2.767918	2000.0	748.000000	111.86	22.267857	284.615385	210.0	70.0	3000.0	220.0	0	5.0	60.0
6	2.569620	1910.0	807.000000	111.86	22.267857	284.615385	210.0	70.0	3000.0	220.0	0	5.0	70.0
7	2.561475	1900.0	535.000000	111.86	22.267857	284.615385	380.0	75.0	1800.0	120.0	0	7.0	47.0
8	3.557018	1930.0	889.000000	129.00	21.250000	300.000000	380.0	75.0	1800.0	120.0	0	7.0	57.0
9	3.532338	2100.0	1421.000000	129.00	21.250000	300.000000	1010.0	78.0	2000.0	300.0	0	7.0	60.0

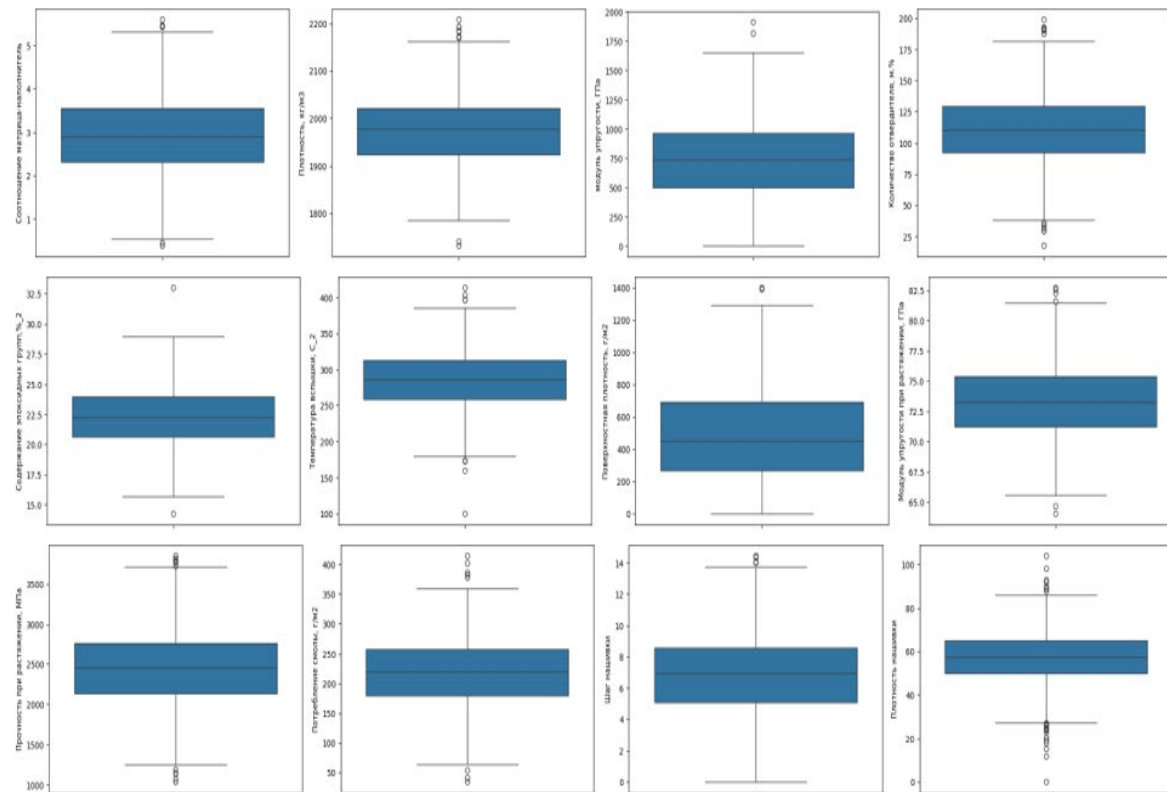
	Соотношение матрица-наполнитель	Плотность, кг/м3	модуль упругости, ГПа	Количество отвердителя, м.%	Содержание эпоксидных групп,%_2	Температура вспышки, С_2	Поверхностная плотность, г/м2	Модуль упругости при растяжении, ГПа	Прочность при растяжении, МПа	Потребление смолы, г/м2	Угол нашивки, град	Шаг нашивки	Плотность нашивки
count	1023.000000	1023.000000	1023.000000	1023.000000	1023.000000	1023.000000	1023.000000	1023.000000	1023.000000	1023.000000	1023.000000	1023.000000	1023.000000
mean	2.930366	1975.734888	739.923233	110.570769	22.244390	285.882151	482.731833	73.328571	2466.922843	218.423144	44.252199	6.899222	57.153929
std	0.913222	73.729231	330.231581	28.295911	2.406301	40.943260	281.314690	3.118983	485.628006	59.735931	45.015793	2.563487	12.350969
min	0.389403	1731.764635	2.436909	17.740275	14.254985	100.000000	0.603740	64.054061	1036.856605	33.803026	0.000000	0.000000	0.000000
25%	2.317887	1924.155467	500.047452	92.443497	20.608034	259.066528	266.816645	71.245018	2135.850448	179.627520	0.000000	5.080033	49.799212
50%	2.906878	1977.621657	739.664328	110.564840	22.230744	285.896812	451.864365	73.268805	2459.524526	219.198882	0.000000	6.916144	57.341920
75%	3.552660	2021.374375	961.812526	129.730366	23.961934	313.002106	693.225017	75.356612	2767.193119	257.481724	90.000000	8.586293	64.944961
max	5.591742	2207.773481	1911.536477	198.953207	33.000000	413.273418	1399.542362	82.682051	3848.436732	414.590628	90.000000	14.440522	103.988901



Анализ распределения переменных и выбросов



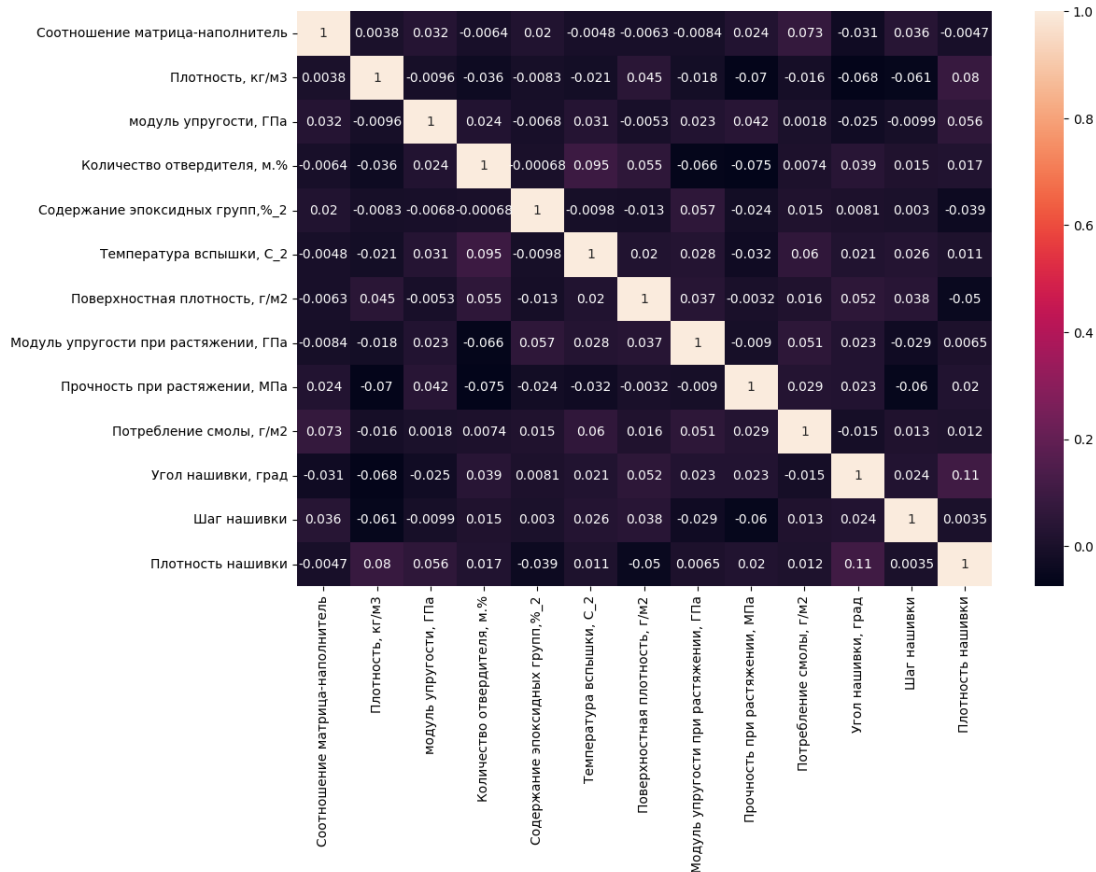
Гистограммы распределения
переменных



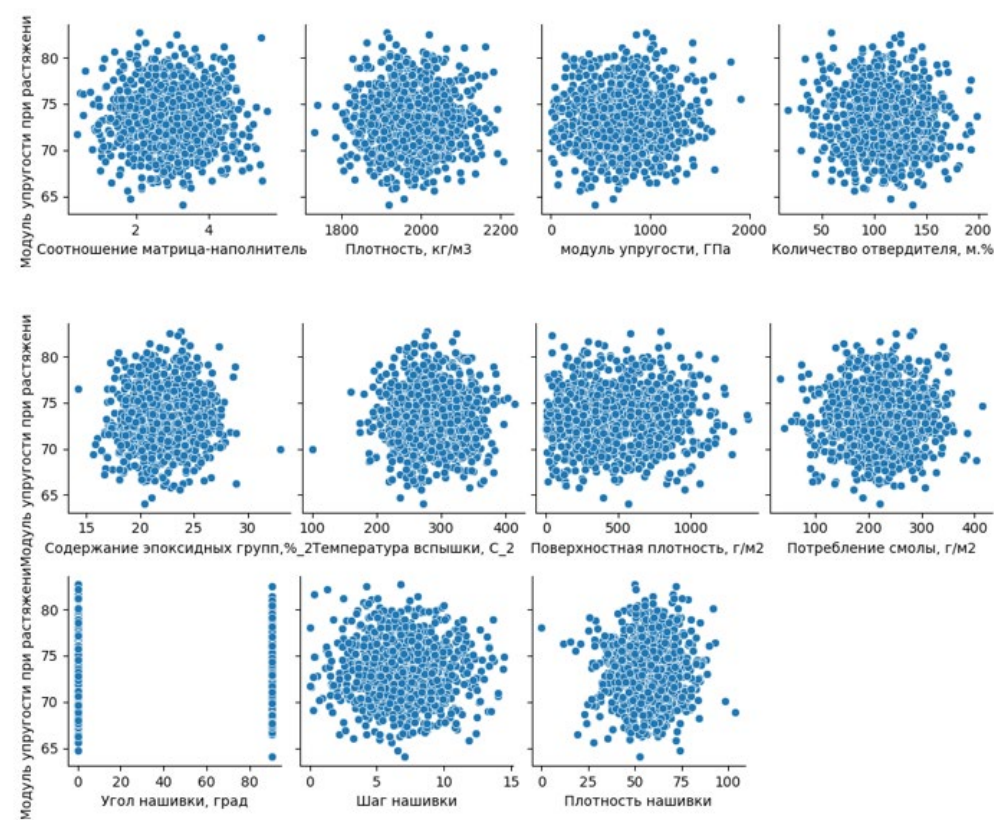
Диаграммы «ящик с усами»



Выявление зависимостей между целевыми переменными и признаками



Тепловая карта корреляции



Попарные графики рассеяния для
Модуля упругости при растяжении



Линейная регрессия

	MAE	MAPE	MSE	RMSE	R ²
Модуль упругости при растяжении					
Исходный датасет, полученный путем удаления всех одномерных выбросов	2,268	0,031	8,2	2,86	-0,023
Нормализованный датасет, полученный путем удаления всех одномерных выбросов	0,14	0,53	0,032	0,18	-0,023
Исходный датасет с признаками на основе тепловой карты корреляции	2,235	0,03	8,00	2,83	0,001
Исходный датасет с признаками на основе коэффициентов	2,267	0,031	8,23	2,87	-0,028
Прочность при растяжении					
Исходный датасет, полученный путем удаления всех одномерных выбросов	391,23	0,17	238112,7	487,97	-0,024
Нормализованный датасет, полученный путем удаления всех одномерных выбросов	0,159	0,58	0,039	0,198	-0,024



KNN-регрессия

	MAE	MAPE	MSE	RMSE	R2
Модуль упругости при растяжении					
Исходный датасет, полученный путем удаления всех одномерных выбросов	2,583	0,035	10,72	3,27	-0,339
Нормализованный датасет, полученный путем удаления всех одномерных выбросов	0,17	1,0	0,043	0,21	-0,132
Исходный датасет с признаками на основе тепловой карты корреляции	2,453	0,033	9,82	3,13	-0,226
Прочность при растяжении					
Исходный датасет, полученный путем удаления всех одномерных выбросов	414,21	0,18	265670,5	515,43	-0,14

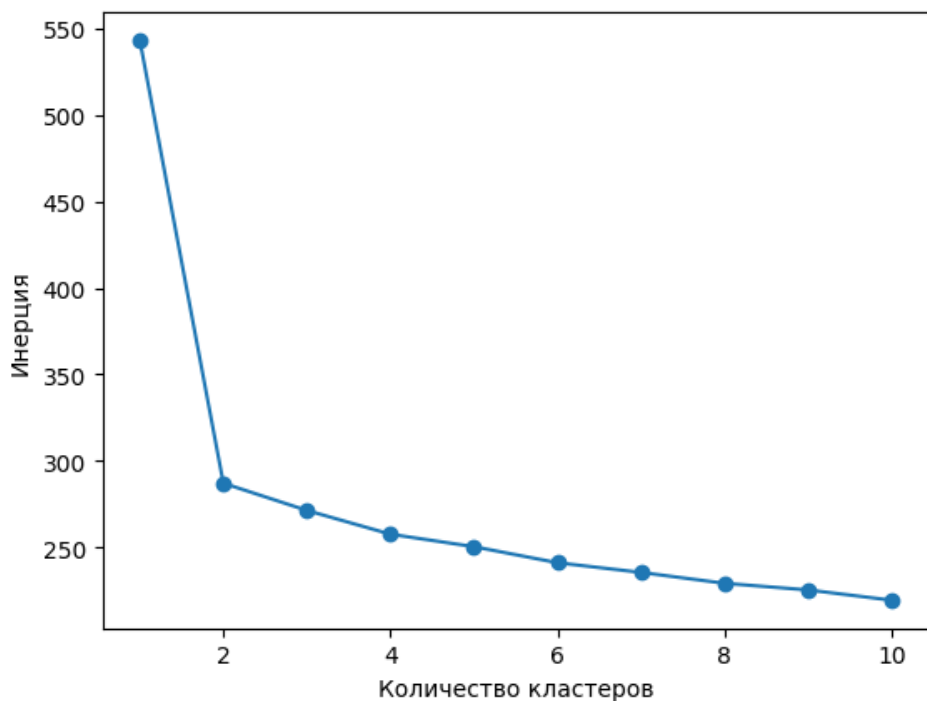


Дерево решений, Градиентный бустинг, Случайный лес

	MAE	MAPE	MSE	RMSE	R ²
Модуль упругости при растяжении					
Дерево решений Исходный датасет, полученный путем удаления всех одномерных выбросов	2,52	0,034	9,83	3,13	-0,227
Градиентный бустинг Исходный датасет, полученный путем удаления всех одномерных выбросов	2,39	0,03	9,01	3,01	-0,13
Градиентный бустинг Нормализованный датасет, полученный путем удаления всех одномерных выбросов	0,15	0,56	0,036	0,19	-0,125
Случайный лес Исходный датасет с признаками на основе тепловой карты корреляции	2,27	0,031	8,38	2,9	-0,05
Прочность при растяжении					
Дерево решений Исходный датасет, полученный путем удаления всех одномерных выбросов	420,26	0,18	281805,9	530,85	-0,212



Кластеризация



	MAE	MAPE	MSE	RMSE	R ²
Модуль упругости при растяжении					
Линейная регрессия	0,14	0,42	0,03	0,18	-0,053
Дерево решений	0,16	0,44	0,04	0,21	-0,043
Случайный лес	0,14	0,42	0,033	0,18	-0,114

Метод «локтя» для определения количества кластеров



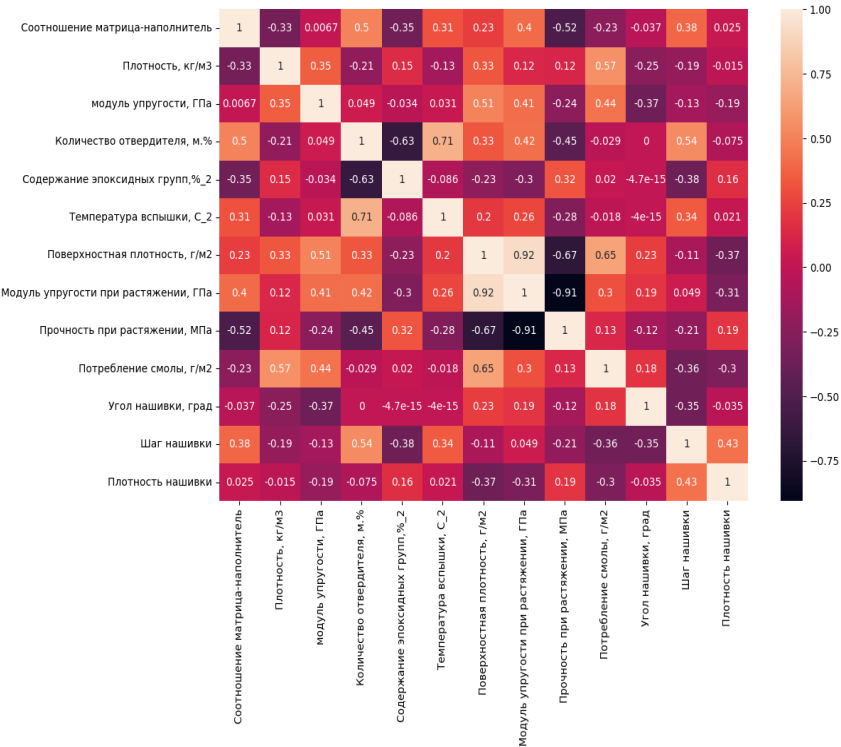
Уменьшение размерности

	0	1	2	3
0	-4.897077	0.823177	-0.323954	-1.116187
1	-2.324463	0.545291	0.434537	-0.239440
2	-2.290845	-0.013652	0.524039	-1.098307
3	-0.780114	0.120092	0.044025	0.915351
4	-1.074763	0.178294	0.751700	0.243356
...
1018	0.098297	0.532227	-1.165392	-0.831011
1019	0.210419	0.801068	-1.101005	0.059975
1020	-0.016972	0.612981	-0.399278	-0.021863
1021	0.606579	1.061286	0.294269	1.002596
1022	1.454526	0.096202	-0.523548	-0.824007

1023 rows × 4 columns

Датасет с 4 признаками

	MAE	MAPE	MSE	RMSE	R ²
Модуль упругости при растяжении					
Линейная регрессия	0,79	1,05	0,99	0,99	-0,008
Дерево решений	0,83	1,87	1,06	1,03	-0,1
Случайный лес	0,81	1,4	1,05	1,03	-0,09



Тепловая карта корреляции для

датасета из первых 23 строк

	MAE	MAPE	MSE	RMSE	R2
Модуль упругости при растяжении					
Линейная регрессия	0,0	0,0	0,0	0,0	1,0
KNN-регрессия	1,28	0,017	3,44	1,86	0,64
KNN-регрессия. Нормализация	0,17	-	0,04	0,21	0,7
KNN-регрессия. Гиперпараметры	0,05	-	0,019	0,14	0,87
Дерево решений	0,0	0,0	0,0	0,0	1,0
Градиентный бустинг	0,25	0,003	0,47	0,69	0,95
Случайный лес	0,33	0,004	0,29	0,53	0,97
Прочность при растяжении					
Линейная регрессия	0,0	0,0	0,0	0,0	1
KNN-регрессия	251,9	0,11	146229	382,4	0,38
KNN-регрессия. Нормализация	0,2	-	0,055	0,24	0,66
KNN-регрессия. Гиперпараметры	0,04	0,13	0,01	0,1	0,94
Дерево решений	0,0	0,0	0,0	0,0	1,0
Градиентный бустинг	15,16	0,007	1830,2	42,78	0,99
Случайный лес	103,3	0,05	19172	138,47	0,92

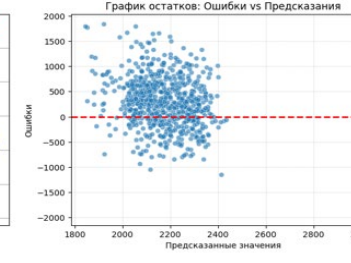
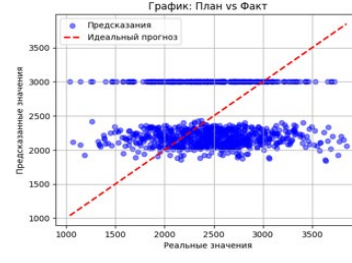
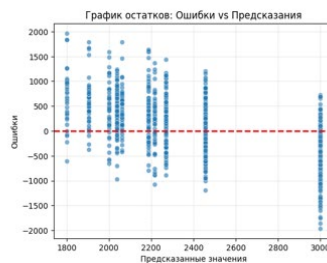
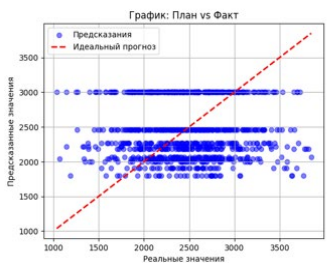
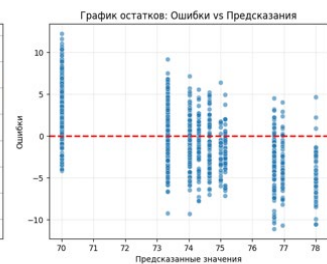
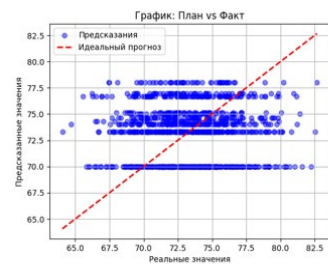
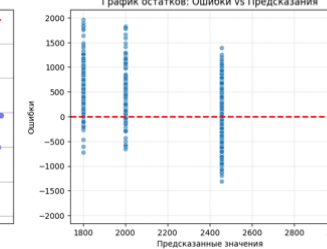
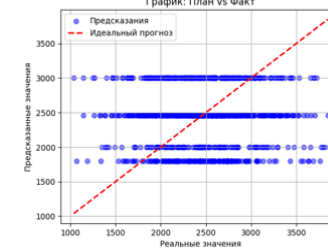
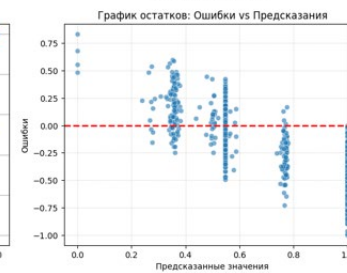
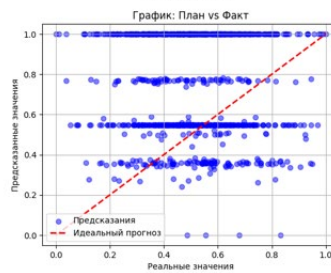
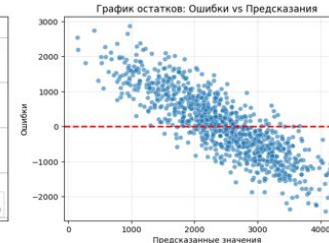
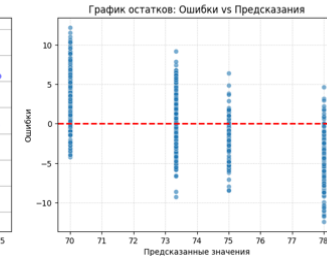
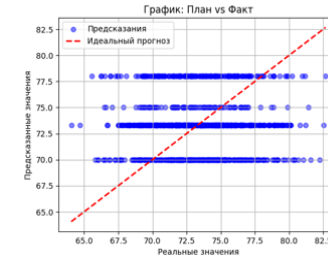
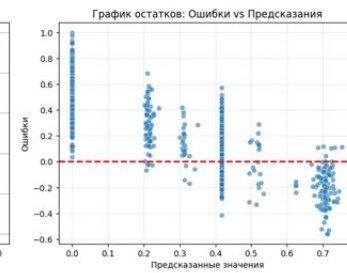
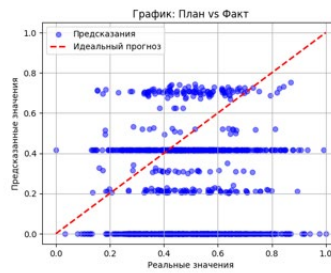
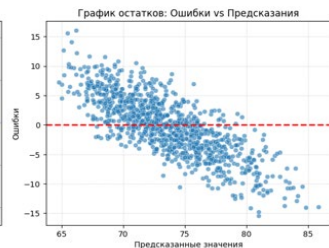
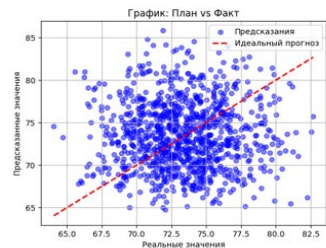


Тестирование моделей

	MAE	MAPE	MSE	RMSE	R2
Модуль упругости при растяжении					
Линейная регрессия	4,07	0,06	25,7	5,01	-1,64
KNN-регрессия. Нормализация	0,32	-	0,15	0,39	-4,37
Дерево решений	3,36	0,05	17,4	4,17	-0,79
Градиентный бустинг	3,26	0,04	16,13	4,01	-0,66
Случайный лес	3,2	0,04	15,4	3,92	-0,58
Прочность при растяжении					
Линейная регрессия	729,18	0,31	832440,2	912,38	-2,53
KNN-регрессия. Нормализация	0,32	-	0,15	0,39	-4,06
Дерево решений	525,92	0,23	425619,3	652,4	-0,81
Градиентный бустинг	509,9	0,22	399590,1	632,13	-0,7
Случайный лес	505,94	0,21	394156,2	627,8	-0,67

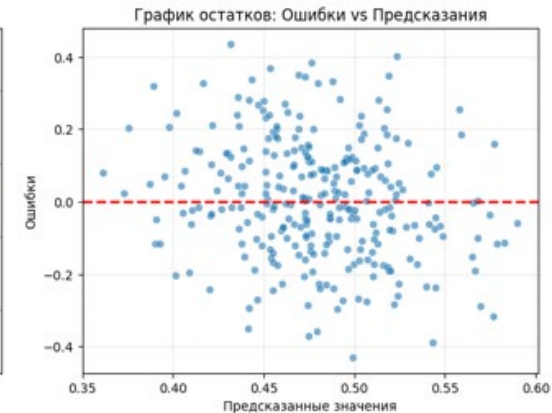
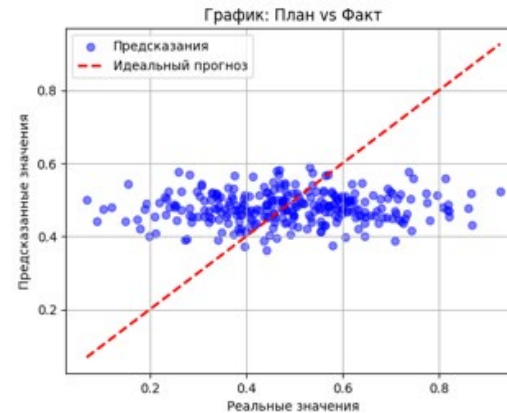
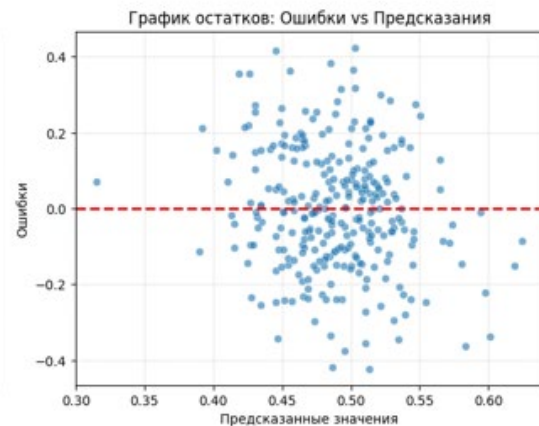
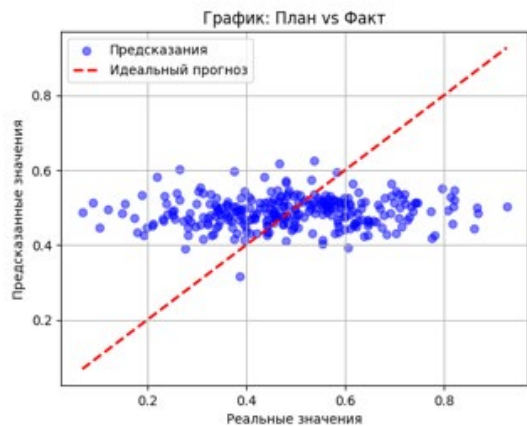
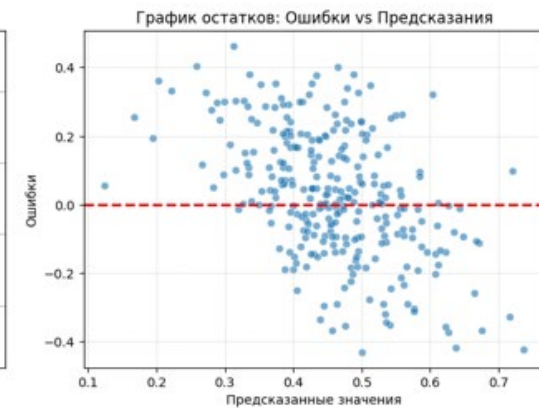
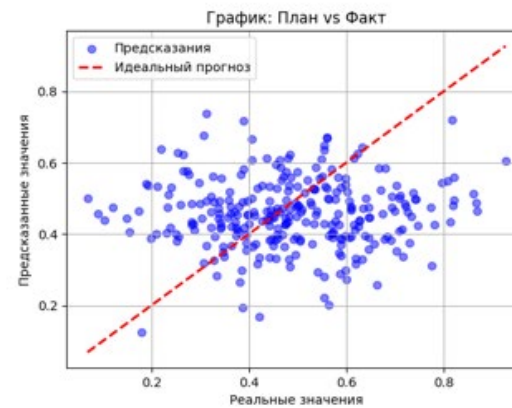
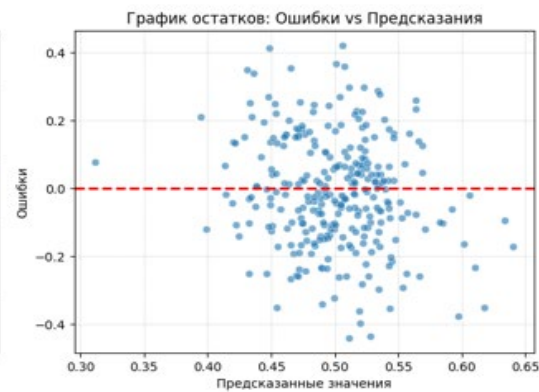
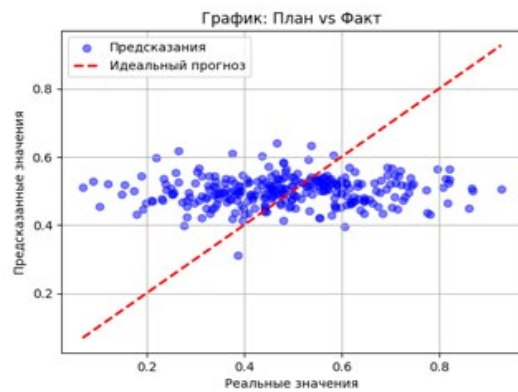


Тестирование моделей





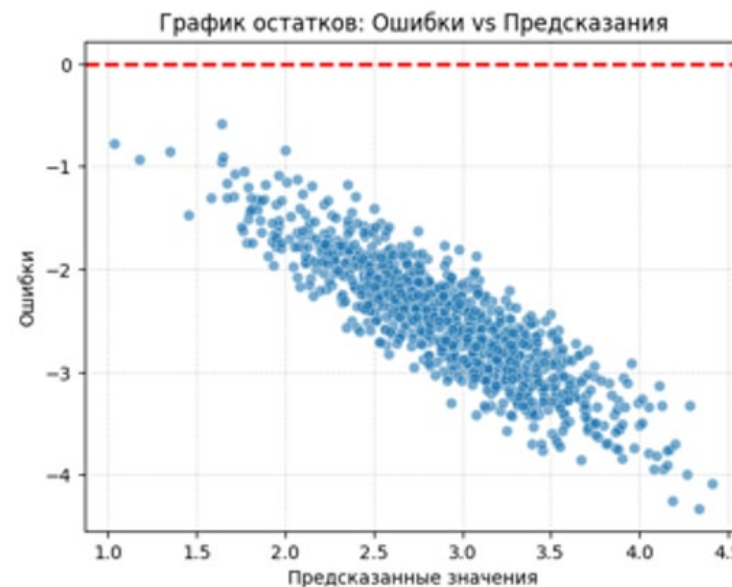
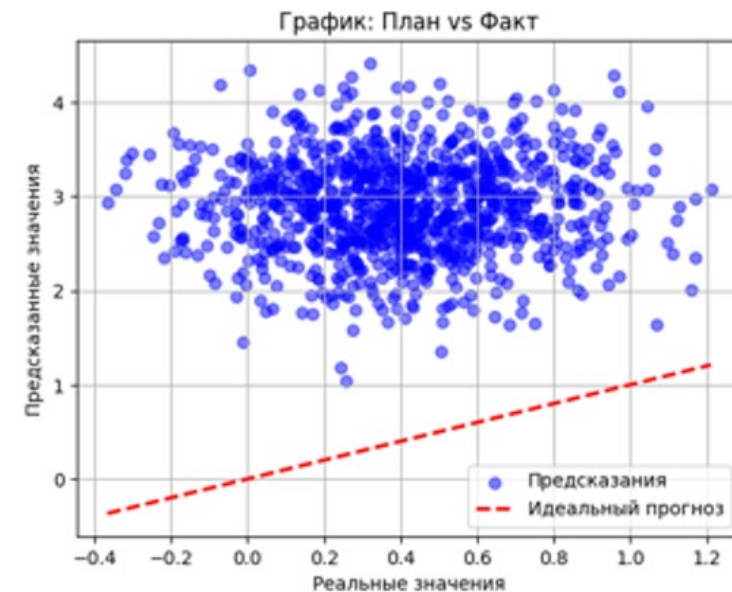
Нейронная сеть





Нейронная сеть

	MAE	MAPE	MSE	RMSE	R ²
1 слой, 32 нейрона, adam	0,13	0,37	0,02	0,16	-0,03
2 слоя, 64 и 32 нейрона, adam	0,14	0,39	0,03	0,19	-0,35
1 слой, 32 нейрона, SGD	0,13	0,36	0,026	0,16	-0,02
2 слоя, 32 и 16 нейронов, SGD	0,13	0,36	0,027	0,16	-0,04
100, 12 признаков	0,78	0,3	1,01	1,01	-0,3
100, 9 признаков	0,78	0,31	1,00	1,00	-0,29
100, 5 признаков	0,81	0,31	0,97	0,98	-0,24
train-23	2,5	28,6	6,57	2,56	-84,8





ЦЕНТР
ДОПОЛНИТЕЛЬНОГО
ОБРАЗОВАНИЯ
МГТУ им. Н.Э. Баумана

Разработка приложения

Simple Browser X

← → ↻ http://127.0.0.1:5000/

Заполните параметры

Соотношение матрица-наполнитель

Плотность, кг/м3

модуль упругости, ГПа

Количество отвердителя, м.%

Содержание эпоксидных групп, %_2

Температура вспышки, C_2

Поверхностная плотность, г/м2

Потребление смолы, г/м2

Угол нашивки, град, г/м2

Шаг нашивки

Плотность нашивки

Simple Browser X

← → ↻ http://127.0.0.1:5000/

Результат

Значение модуля упругости при растяжении

[73.74]

Интерфейс страницы приложения по расчету Модуля
упругости при растяжении



ЦЕНТР
ДОПОЛНИТЕЛЬНОГО
ОБРАЗОВАНИЯ
МГТУ им. Н.Э. Баумана



do.bmstu.ru