



ЦЕНТР  
ДОПОЛНИТЕЛЬНОГО  
ОБРАЗОВАНИЯ  
МГТУ им. Н.Э. Баумана

# Выпускная квалификационная работа

по курсу «Data Science»

Тема: Прогнозирование конечных свойств новых  
материалов (композиционных материалов)

Слушатель: Степанова Василиса Валерьевна



# Разведочный анализ данных

1

Объединила датасеты и изучила описательную статистику с добавлением медианы.

2

Проверила типы данных признаков

	count	mean	std	min	25%	50%	75%	max	median
Соотношение матрица-наполнитель	1023.0000	2.9304	0.9132	0.3894	2.3179	2.9069	3.5527	5.5917	2.9069
Плотность, кг/м3	1023.0000	1975.7349	73.7292	1731.7646	1924.1555	1977.6217	2021.3744	2207.7735	1977.6217
модуль упругости, ГПа	1023.0000	739.9232	330.2316	2.4369	500.0475	739.6643	961.8125	1911.5365	739.6643
Количество отвердителя, м.%	1023.0000	110.5708	28.2959	17.7403	92.4435	110.5648	129.7304	198.9532	110.5648
Содержание эпоксидных групп,%_2	1023.0000	22.2444	2.4063	14.2550	20.6080	22.2307	23.9619	33.0000	22.2307
Температура вспышки, С_2	1023.0000	285.8822	40.9433	100.0000	259.0665	285.8968	313.0021	413.2734	285.8968
Поверхностная плотность, г/м2	1023.0000	482.7318	281.3147	0.6037	266.8166	451.8644	693.2250	1399.5424	451.8644
Модуль упругости при растяжении, ГПа	1023.0000	73.3286	3.1190	64.0541	71.2450	73.2688	75.3566	82.6821	73.2688
Прочность при растяжении, МПа	1023.0000	2466.9228	485.6280	1036.8566	2135.8504	2459.5245	2767.1931	3848.4367	2459.5245
Потребление смолы, г/м2	1023.0000	218.4231	59.7359	33.8030	179.6275	219.1989	257.4817	414.5906	219.1989
Угол нашивки, град	1023.0000	44.2522	45.0158	0.0000	0.0000	0.0000	90.0000	90.0000	0.0000
Шаг нашивки	1023.0000	6.8992	2.5635	0.0000	5.0800	6.9161	8.5863	14.4405	6.9161
Плотность нашивки	1023.0000	57.1539	12.3510	0.0000	49.7992	57.3419	64.9450	103.9889	57.3419

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, С_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)

3

Проверила датасет на наличие:

- Пропусков - 0
- Дубликатов - 0
- Уникальных значений

Соотношение матрица-наполнитель	1014
Плотность, кг/м3	1013
модуль упругости, ГПа	1020
Количество отвердителя, м.%	1005
Содержание эпоксидных групп,%_2	1004
Температура вспышки, С_2	1003
Поверхностная плотность, г/м2	1004
Модуль упругости при растяжении, ГПа	1004
Прочность при растяжении, МПа	1004
Потребление смолы, г/м2	1003
Угол нашивки, град	2
Шаг нашивки	989
Плотность нашивки	988

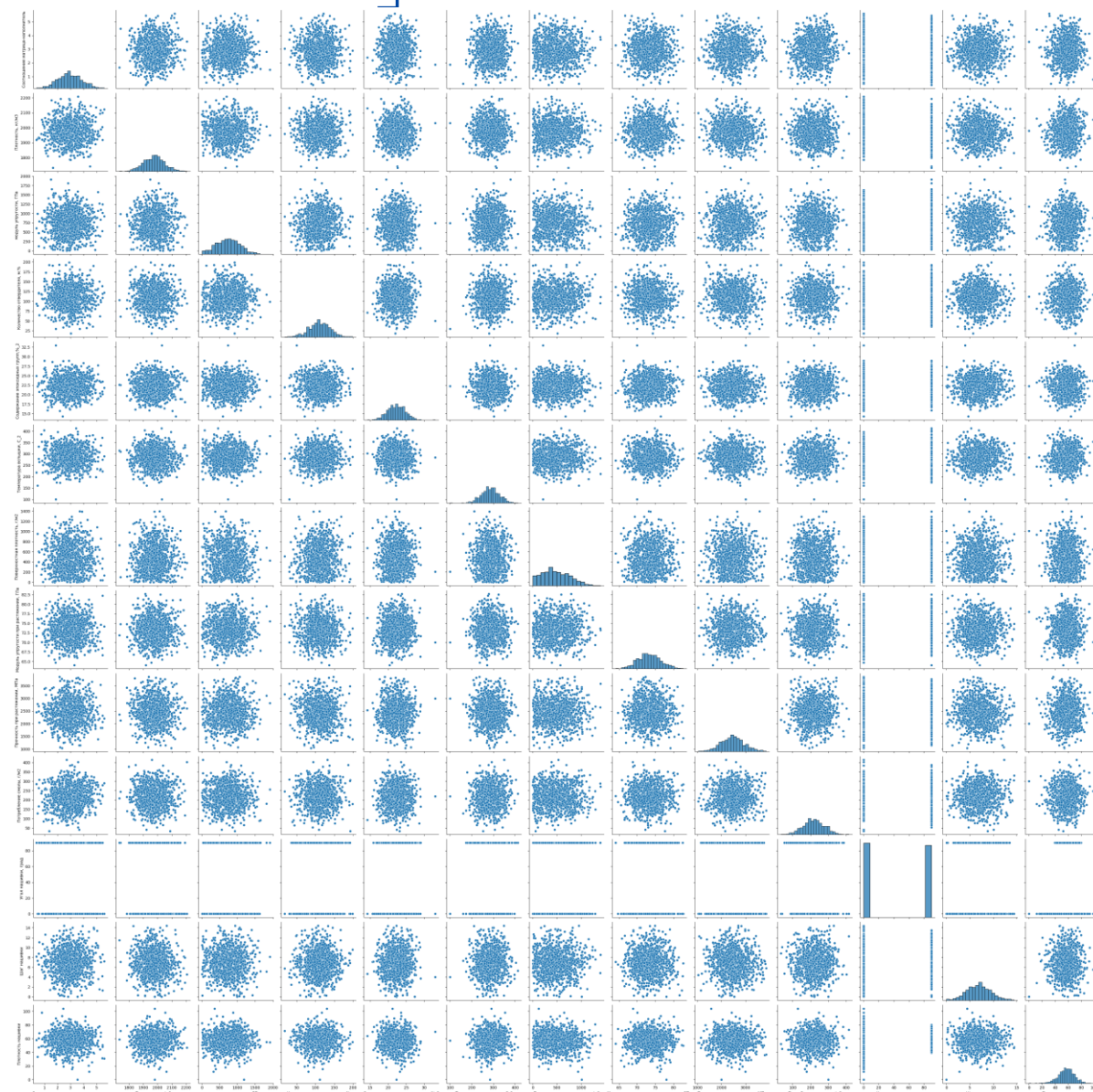
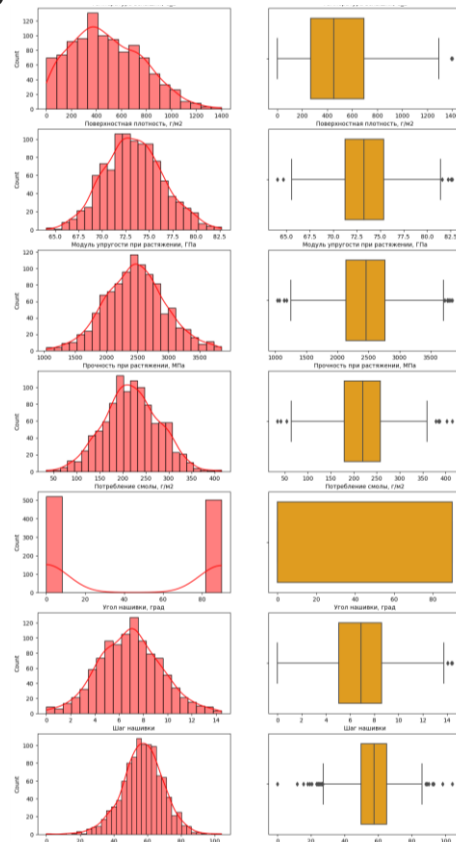
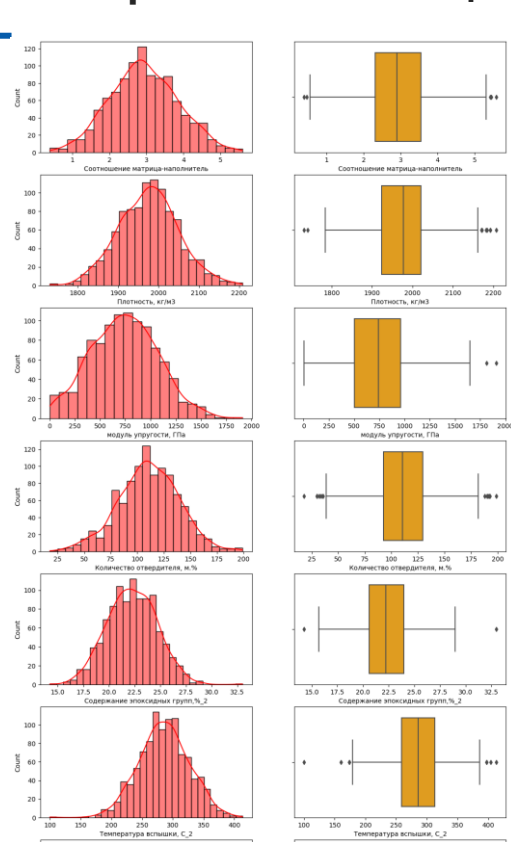


ЦЕНТР  
ДОПОЛНИТЕЛЬНОГО  
ОБРАЗОВАНИЯ  
МГУ им. Н.Э. Баумана

# Разведочный анализ данных

4 Построила попарные графики рассеяния

5 Построила графики распределения переменных и "ящики с усами"







# Разведочный анализ данных

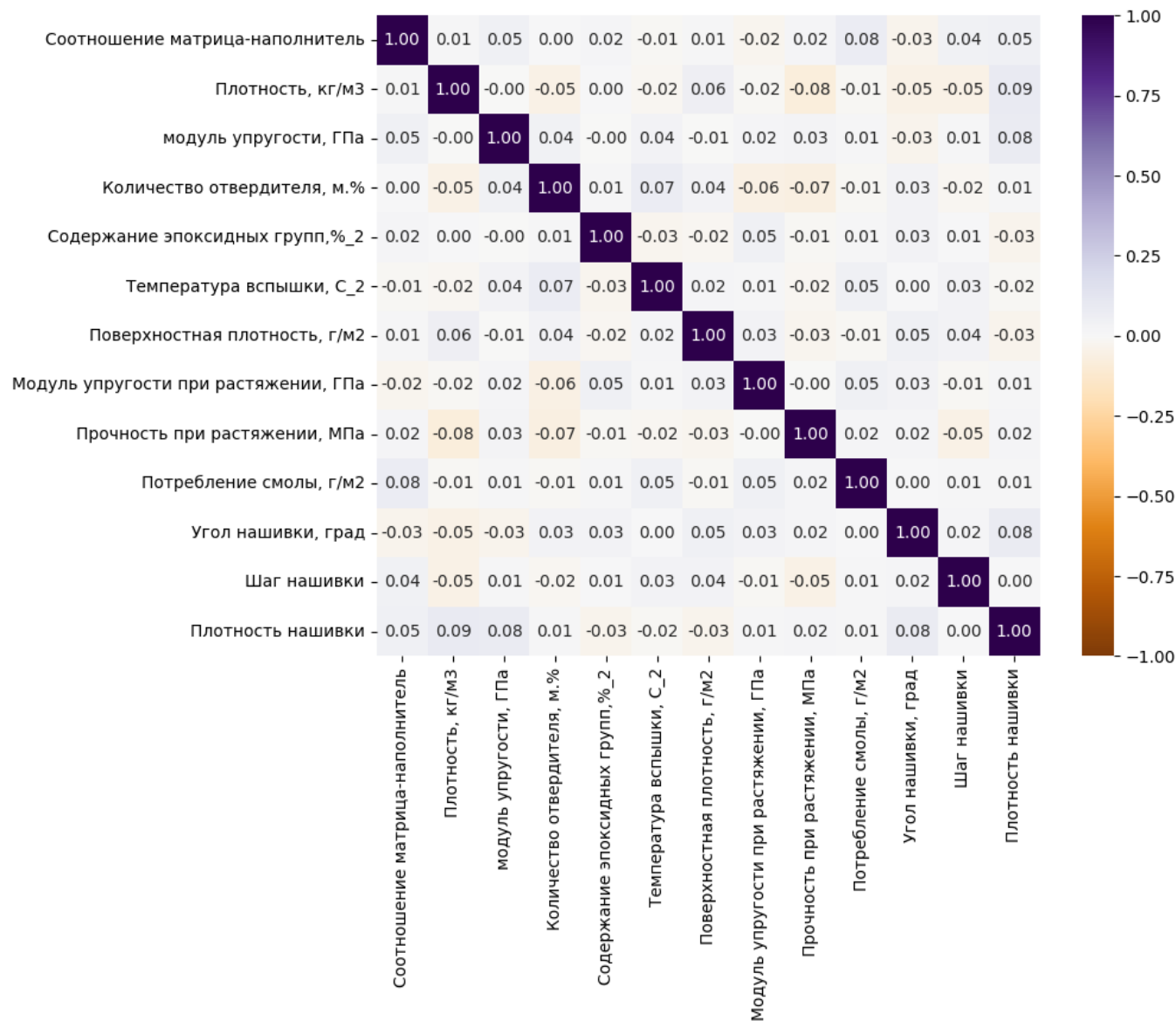
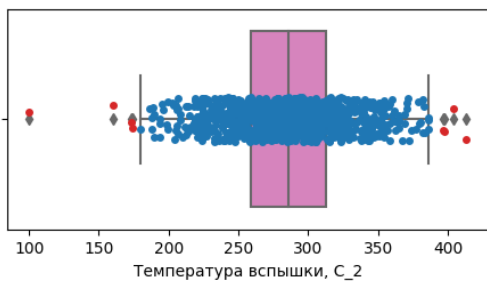
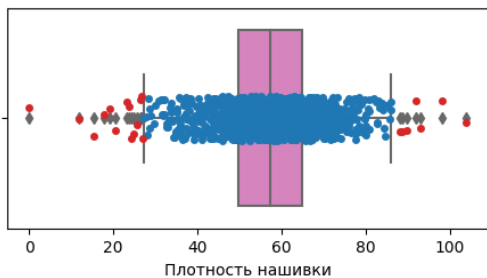
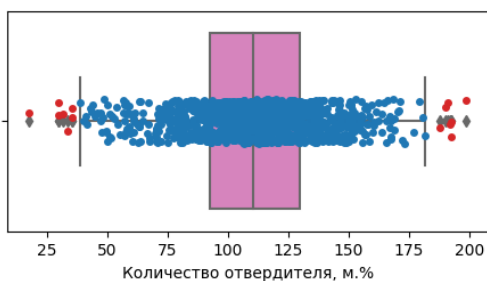
6

## Рассмотрела матрицу корреляции

7

## Определила выбросы

## Примеры:





## Построенные модели

- Линейная регрессия
- Лассо (LASSO) регрессия
- Гребневая (Ridge) регрессия
- Метод опорных векторов для регрессии
- Метод k-ближайших соседей
- Деревья решений
- Случайный лес
- Градиентный бустинг
- Многослойный персептрон
- Нейронная сеть



# Модель для модуля упругости при растяжении

## Сравнение моделей с параметрами по умолчанию

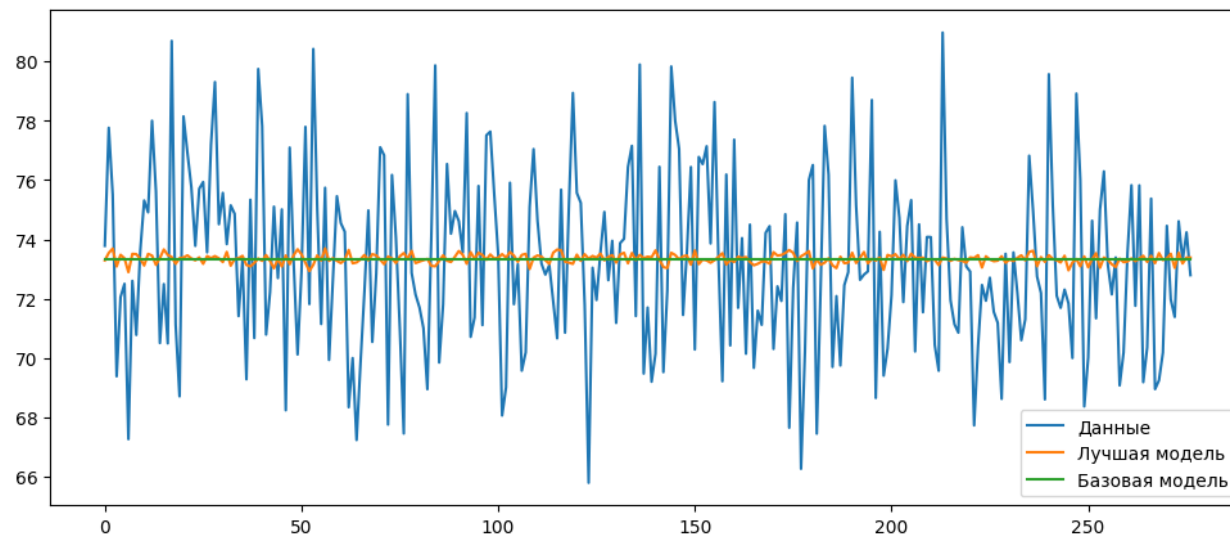
	R2	RMSE	MAE	MAPE	max_error
DummyRegressor	-0.011479	-3.018468	-2.434764	-0.033263	-7.216742
LinearRegression	-0.022599	-3.034619	-2.453669	-0.033520	-7.222509
Ridge	-0.021050	-3.032391	-2.451847	-0.033495	-7.212702
Lasso	-0.011479	-3.018468	-2.434764	-0.033263	-7.216742
SVR	-0.050196	-3.073227	-2.486431	-0.033930	-7.562206
KNeighborsRegressor	-0.252309	-3.350693	-2.673319	-0.036522	-8.485262
DecisionTreeRegressor	-1.317465	-4.505732	-3.660577	-0.050026	-11.406496
RandomForestRegressor	-0.081437	-3.117793	-2.525298	-0.034499	-7.342254

## Сравнение моделей с подобранными параметрами

	R2	RMSE	MAE	MAPE	max_error
Ridge(alpha=60, positive=True, solver='lbfgs')	-0.008194	-3.013781	-2.433792	-0.033249	-7.173646
Lasso(alpha=0.15)	-0.011479	-3.018468	-2.434764	-0.033263	-7.216742
SVR(C=0.03)	-0.009769	-3.016089	-2.434354	-0.033249	-7.220076
KNeighborsRegressor(n_neighbors=29)	-0.033495	-3.050098	-2.470094	-0.033784	-7.219134
DecisionTreeRegressor(max_depth=1, max_features=1, random_state=3128, splitter='random')	-0.011451	-3.018141	-2.426078	-0.033145	-7.182651
RandomForestRegressor(bootstrap=False, criterion='absolute_error', max_depth=2, max_features=1, n_estimators=50, random_state=3128)	-0.011578	-3.018652	-2.438736	-0.033310	-7.189810



# Модель для модуля упругости при растяжении



	R2	RMSE	MAE	MAPE	max_error
Базовая модель	-0.001840	-3.023023	-2.480618	-0.033942	-7.628576
Лучшая модель (модель Ridge)	-0.007314	-3.031272	-2.486880	-0.034029	-7.581602

	R2	RMSE	MAE	MAPE	max_error
Модуль упругости, тренировочный	0.010635	-3.008967	-2.422298	-0.033091	-7.839390
Модуль упругости, тестовый	-0.007314	-3.031272	-2.486880	-0.034029	-7.581602



## Модель для прочности при растяжении

### Сравнение моделей с параметрами по умолчанию

	R2	RMSE	MAE	MAPE	max_error
DummyRegressor	-0.014754	-458.517408	-366.664655	-0.160630	-1095.981614
LinearRegression	-0.017429	-459.567129	-368.992333	-0.161357	-1121.299260
Ridge	-0.016246	-459.284035	-368.747797	-0.161270	-1119.398974
Lasso	-0.009446	-457.664727	-367.497218	-0.160752	-1112.987433
SVR	-0.012648	-458.069725	-366.556719	-0.160477	-1093.217827
DecisionTreeRegressor	-1.101336	-648.893181	-524.028712	-0.223497	-1648.797617
GradientBoostingRegressor	-0.121957	-482.055962	-389.345452	-0.169754	-1194.692003

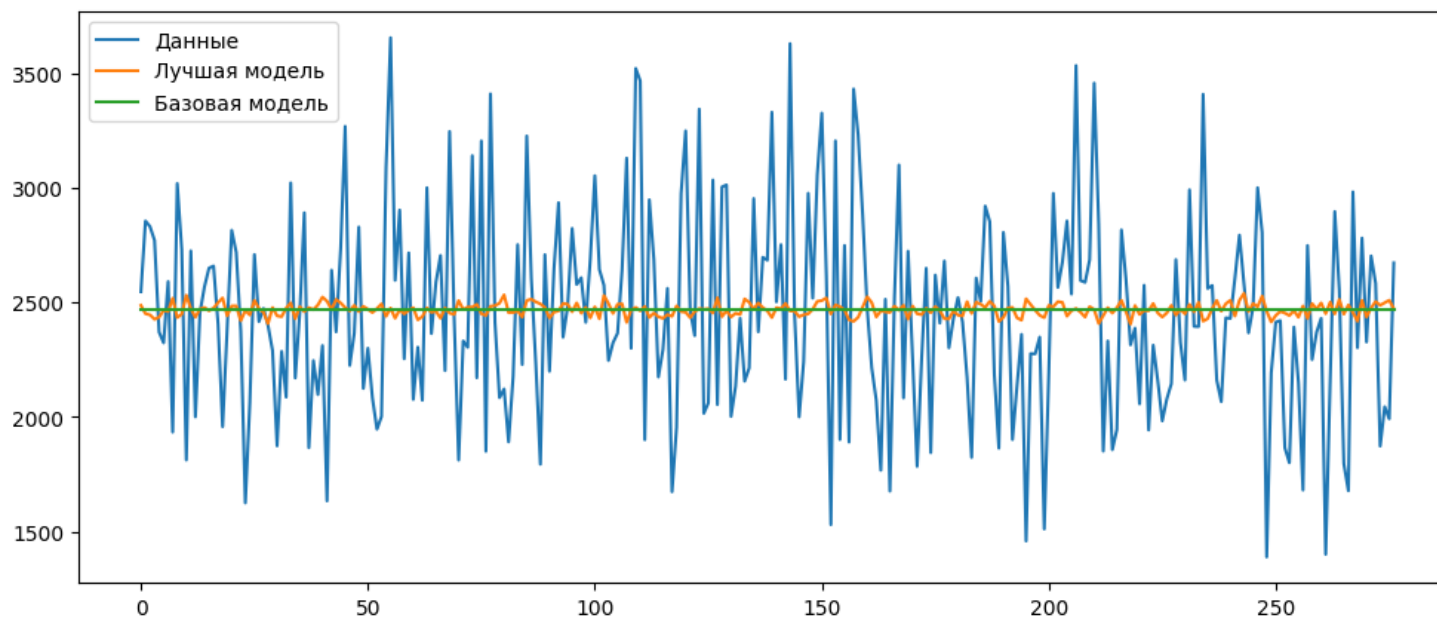
### Сравнение моделей с подобранными параметрами

	R2	RMSE	MAE	MAPE	max_error
Ridge(alpha=60, random_state=3128, solver='sag')	-0.012026	-458.047100	-366.642494	-0.160565	-1094.747442
Lasso(alpha=5)	-0.008467	-457.198647	-366.086794	-0.160291	-1090.943083
SVR(C=0.2)	-0.012996	-458.138246	-366.569324	-0.160528	-1093.961228
DecisionTreeRegressor(max_depth=1, max_features=3, random_state=3128)	-0.020410	-459.733570	-365.998753	-0.160581	-1107.058869
GradientBoostingRegressor(loss='absolute_error', max_depth=1, max_features=11, n_estimators=50, random_state=3128)	-0.022204	-460.206446	-368.817375	-0.161594	-1105.850304





# Модель для прочности при растяжении



	R2	RMSE	MAE	MAPE	max_error
Базовая модель	-0.001555	-439.676848	-350.354301	-0.151168	-1187.738138
Лучшая модель (модель Lasso)	-0.009342	-441.382723	-350.450799	-0.151207	-1179.461001

	R2	RMSE	MAE	MAPE	max_error
Прочность при растяжении, тренировочный	0.013067	-456.085754	-364.432618	-0.159590	-1266.944961
Прочность при растяжении, тестовый	-0.009342	-441.382723	-350.450799	-0.151207	-1179.461001

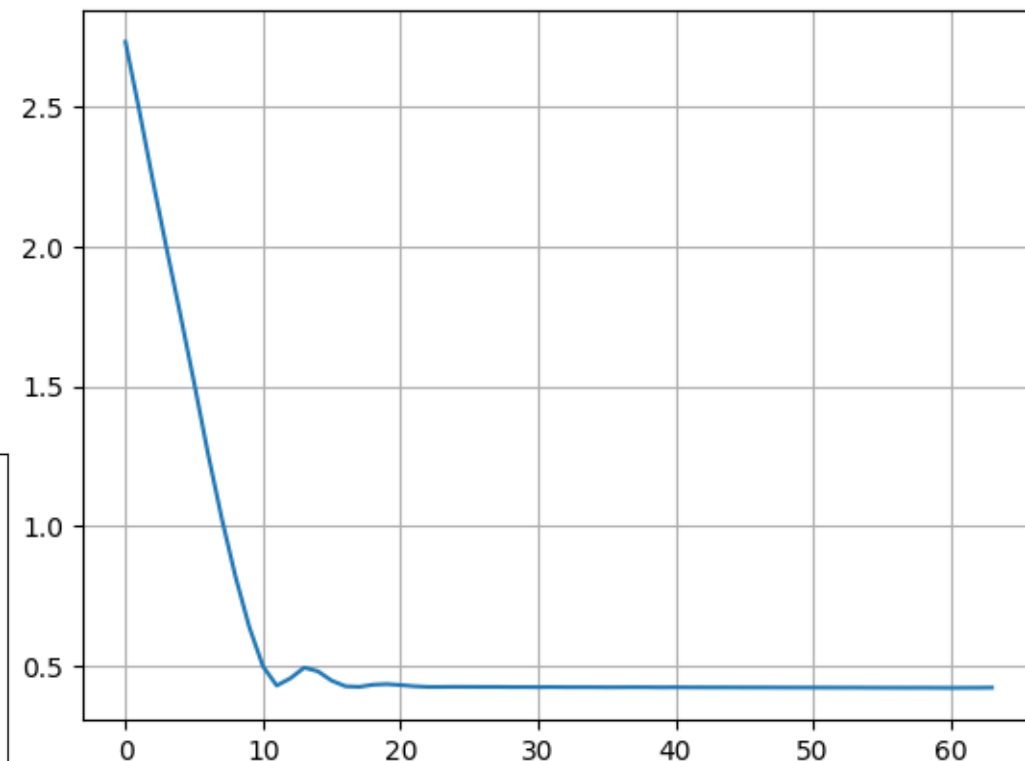
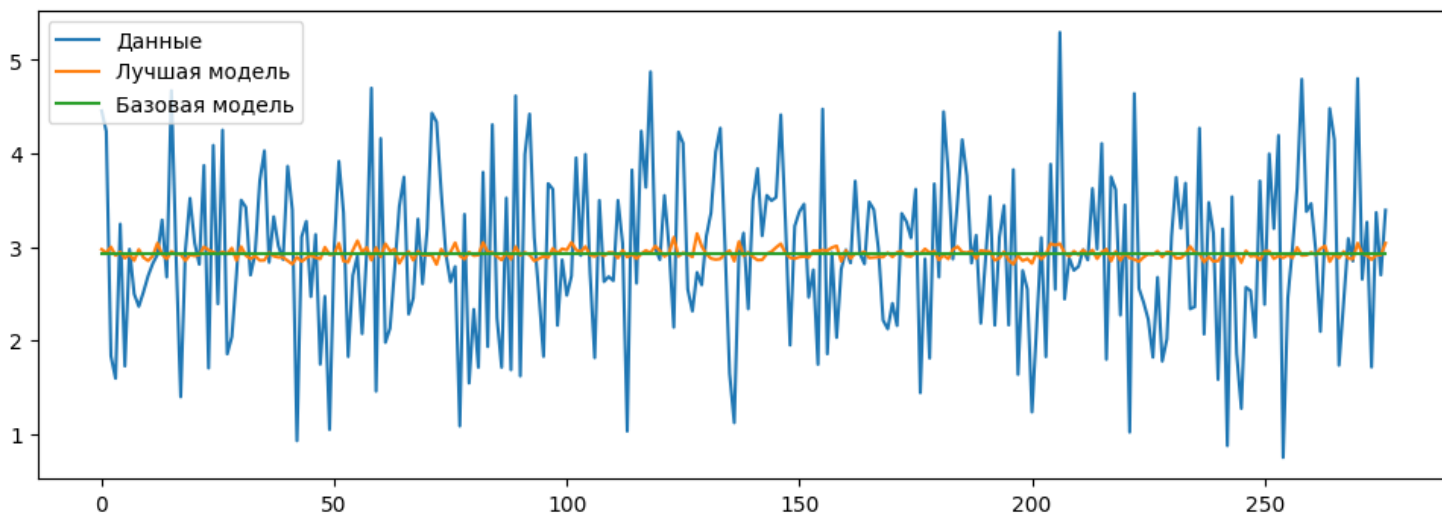


# Модель для соотношения матрица-наполнитель

## Модель MLPRegressor из библиотеки sklearn

### MLPRegressor

```
MLPRegressor(early_stopping=True,  
             hidden_layer_sizes=(24, 24, 24, 24, 24, 24, 24, 24), max_iter=5000,  
             random_state=3128, validation_fraction=0.3, verbose=True)
```



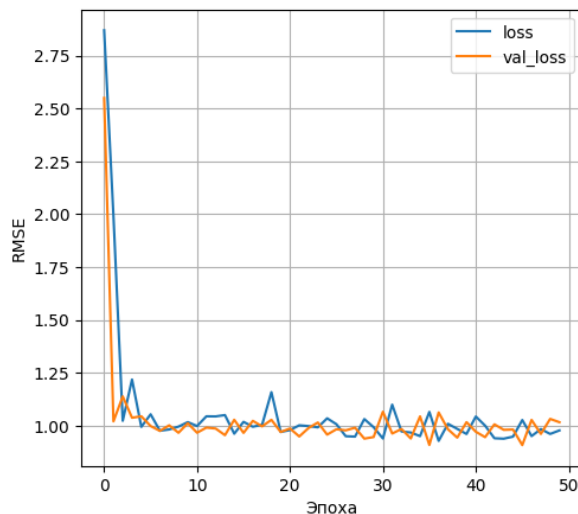
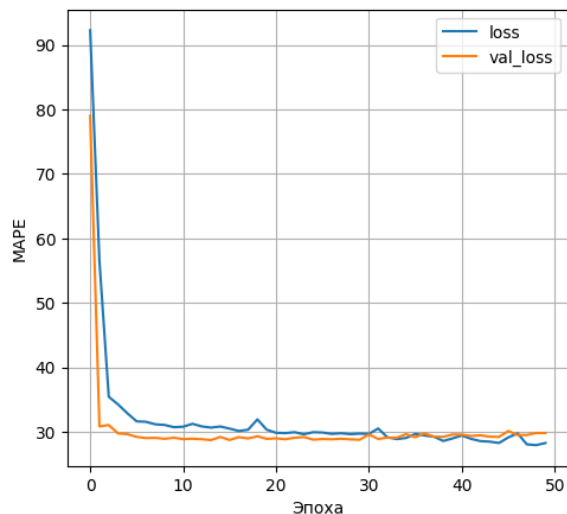
	R2	RMSE	MAE	MAPE	max_error
DummyRegressor	-0.000074	-0.868130	-0.691547	-0.296759	-2.370117
MLPRegressor	0.007322	-0.864914	-0.694770	-0.296923	-2.260243

	R2	RMSE	MAE	MAPE	max_error
Соотношение матрица-наполнитель, тренировочный	0.007198	-0.903012	-0.728448	-0.315134	-2.401036
Соотношение матрица-наполнитель, тестовый	0.007322	-0.864914	-0.694770	-0.296923	-2.260243

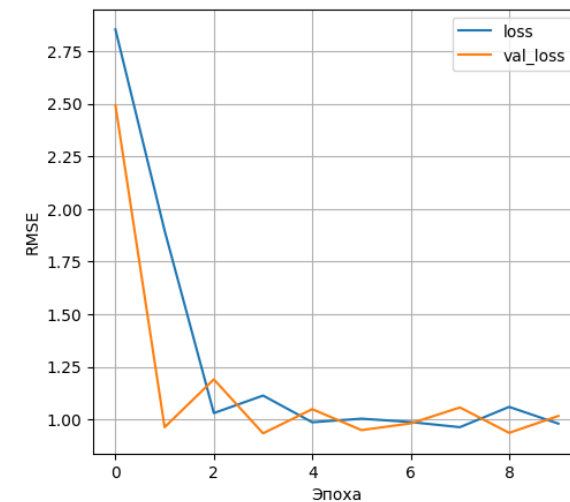
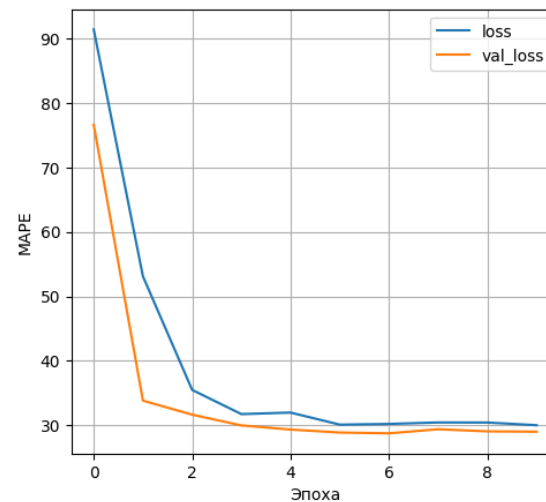


# Модель для соотношения матрица-наполнитель

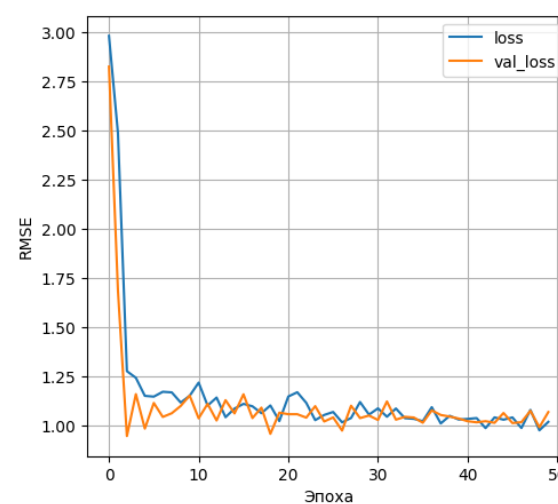
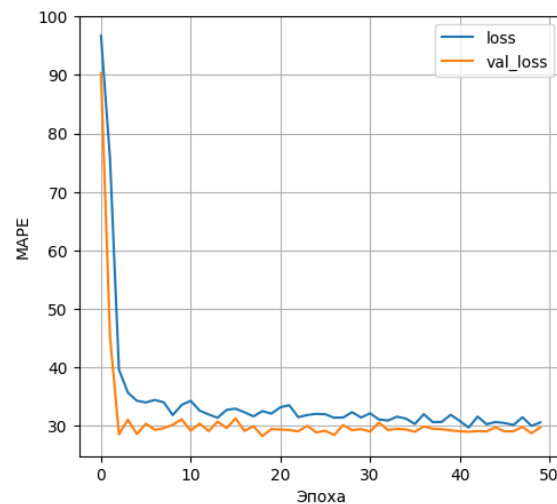
## ■ Обучение Нейросети



## ■ Нейросеть с ранней остановкой



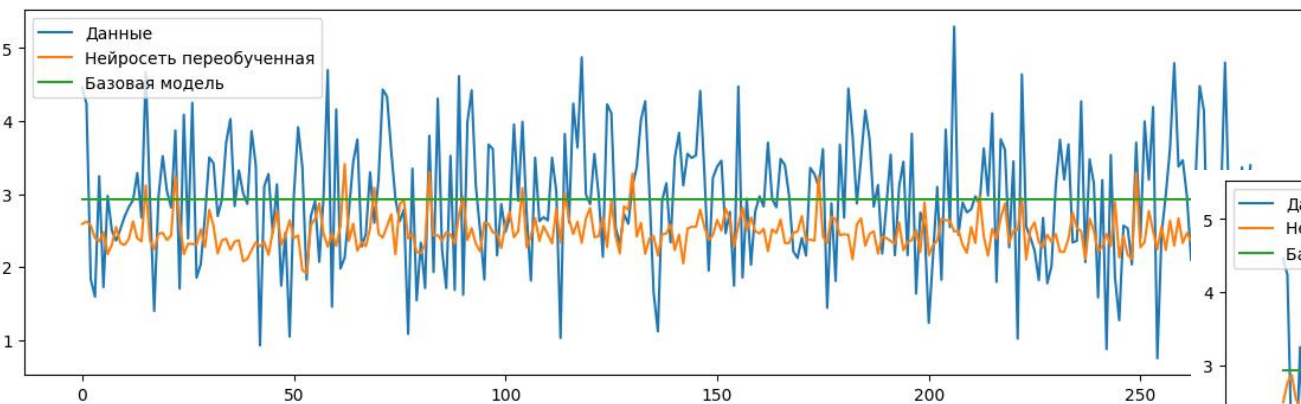
## ■ Нейросеть с Dropout-слоем



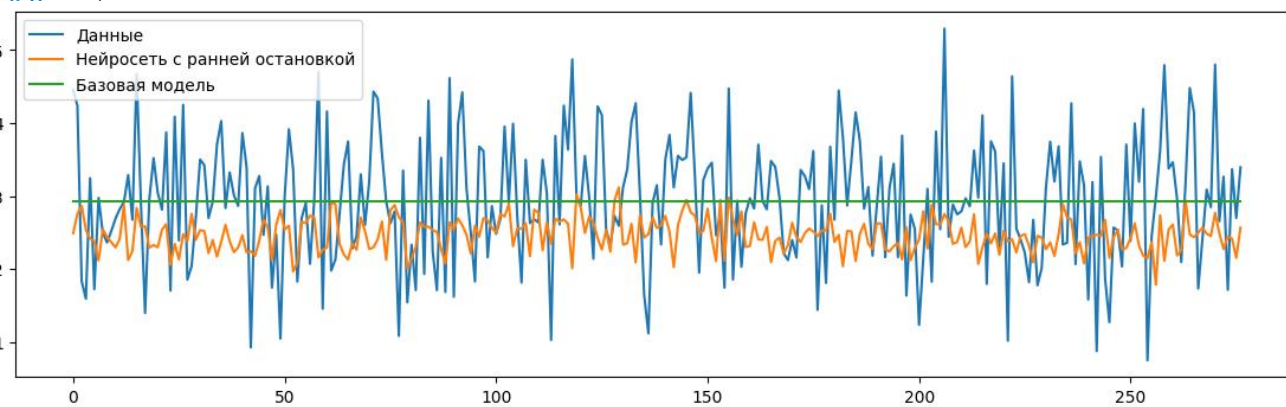


# Модель для соотношения матрица-наполнитель

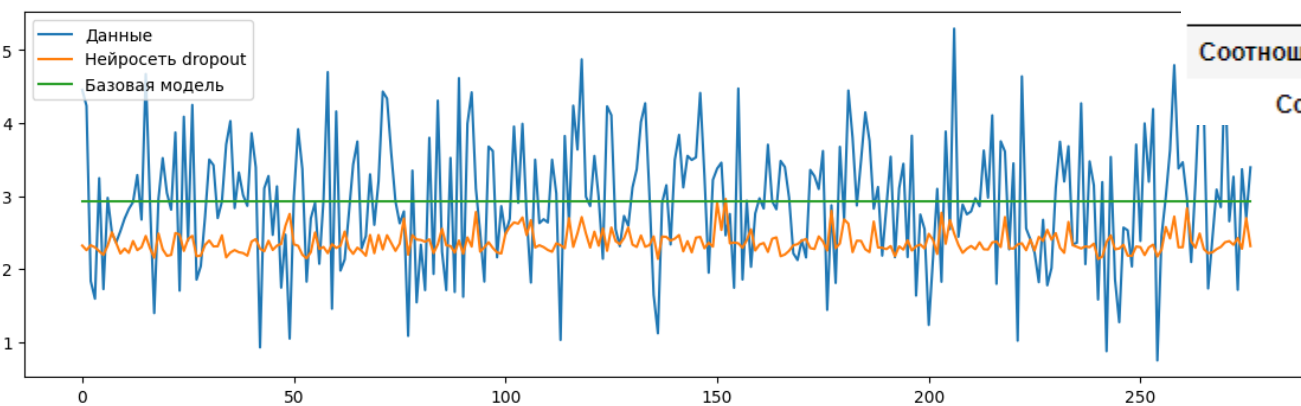
## ■ Обучение Нейросети



## ■ Нейросеть с ранней остановкой



## ■ Нейросеть с Dropout-слоем



	R2	RMSE	MAE	MAPE	max_error
Соотношение матрица-наполнитель, тренировочный	0.007198	-0.903012	-0.728448	-0.315134	-2.401036
Соотношение матрица-наполнитель, тестовый	0.007322	-0.864914	-0.694770	-0.296923	-2.260243

	R2	RMSE	MAE	MAPE	max_error
DummyRegressor	-0.000074	-0.868130	-0.691547	-0.296759	-2.370117
Нейросеть переобученная	-0.282342	-0.983039	-0.802209	-0.288258	-2.807613
Нейросеть с ранней остановкой	-0.337591	-1.003992	-0.816470	-0.293582	-2.862930
Нейросеть dropout	-0.432122	-1.038864	-0.847236	-0.294678	-2.807533





# Разработка приложения

- Разработала приложение для прогнозирования соотношения матрица-наполнитель

## Прогнозирование соотношения матрица-наполнитель

Введите значения, входящие в указанные диапазоны

Плотность, кг/м<sup>3</sup> (1700...2300)

Модуль упругости, ГПа (2...2000)

Количество отвердителя, м. % (17...200)

Содержание эпоксидных групп, %<sub>2</sub> (14...34)

Температура вспышки, С<sub>2</sub> (100...414)

Поверхностная плотность, г/м<sup>2</sup> (0.6...1400)

Модуль упругости при растяжении, ГПа (64...83)

Прочность при растяжении, МПа (1036...3849)

Потребление смолы, г/м<sup>2</sup> (33...414)

Угол нашивки, град(0 или 90)

Шаг нашивки(0...15)

Плотность нашивки (0...104)

Отправить



ЦЕНТР  
ДОПОЛНИТЕЛЬНОГО  
ОБРАЗОВАНИЯ  
МГТУ им. Н.Э. Баумана

# Спасибо за внимание!



ЦЕНТР  
ДОПОЛНИТЕЛЬНОГО  
ОБРАЗОВАНИЯ  
МГТУ им. Н.Э. Баумана



[do.bmstu.ru](https://do.bmstu.ru)