

Образовательный центр МГТУ им. Н.Э. Баумана

# **Выпускная квалификационная работа по курсу "Data Science"**

**Тема: Прогнозирование конечных свойств  
новых материалов (композиционных материалов)**

**Слушатель: Урушадзе Тамара**

# Постановка задачи

- изучить предметную область
- провести разведочный анализ данных
- разделить данные на тренировочную и тестовую выборки
- выполнить препроцессинг (предобработку)
- выбрать базовую модель и модели для подбора
- сравнить модели с гиперпараметрами по умолчанию
- подобрать гиперпараметры с помощью поиска по сетке с перекрестной проверкой
- сравнить модели после подбора гиперпараметров и выбрать лучшую
- сравнить качество лучшей и базовой моделей на тестовой выборке
- сравнить качество лучшей модели на тренировочной и тестовой выборке
- разработать приложение

# Разведочный анализ данных

X\_bp (матрица из базальтопластика):

признаков: 10 и индекс строк: 1023

X\_nip (наполнитель из углепластика):

признаков: 3 и индекс строк: 1040

Объединение с типом INNER по индексу, получилось:

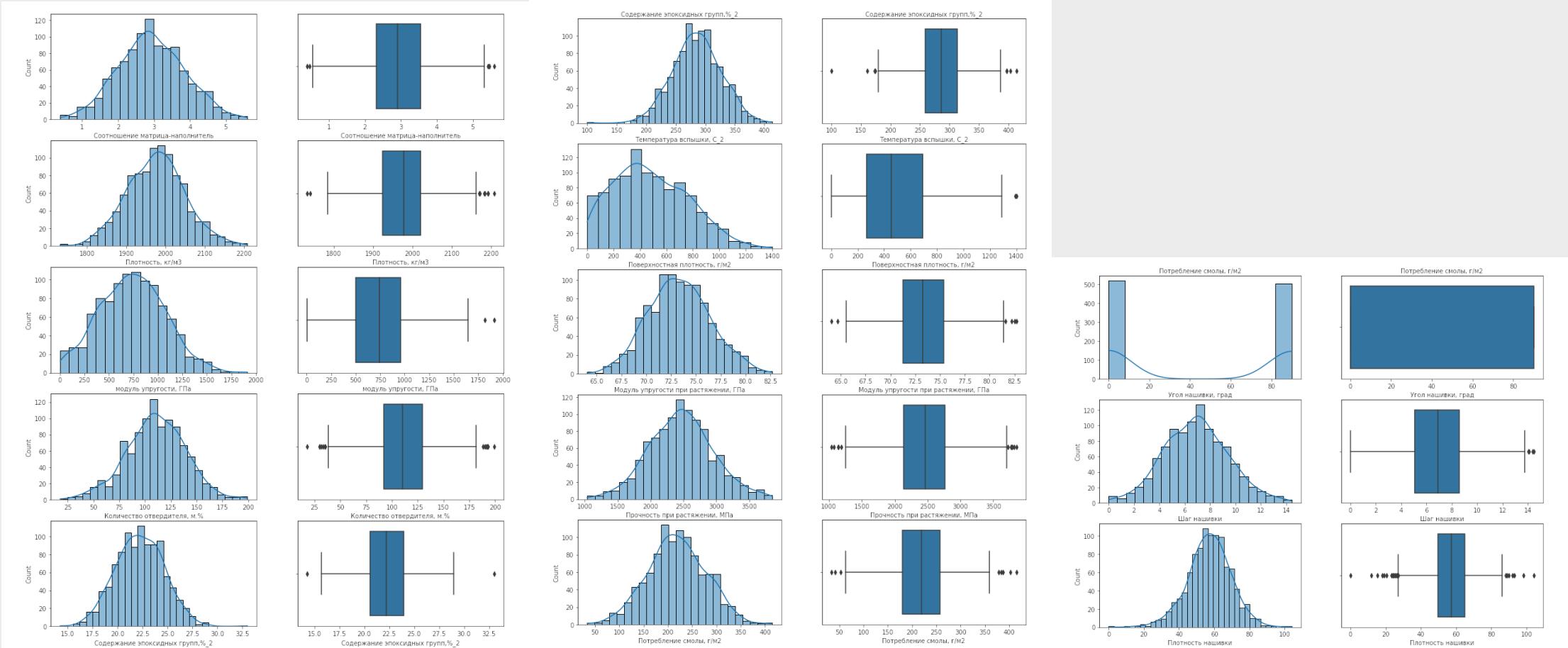
признаков: 13 строк: 1023

# Разведочный анализ данных

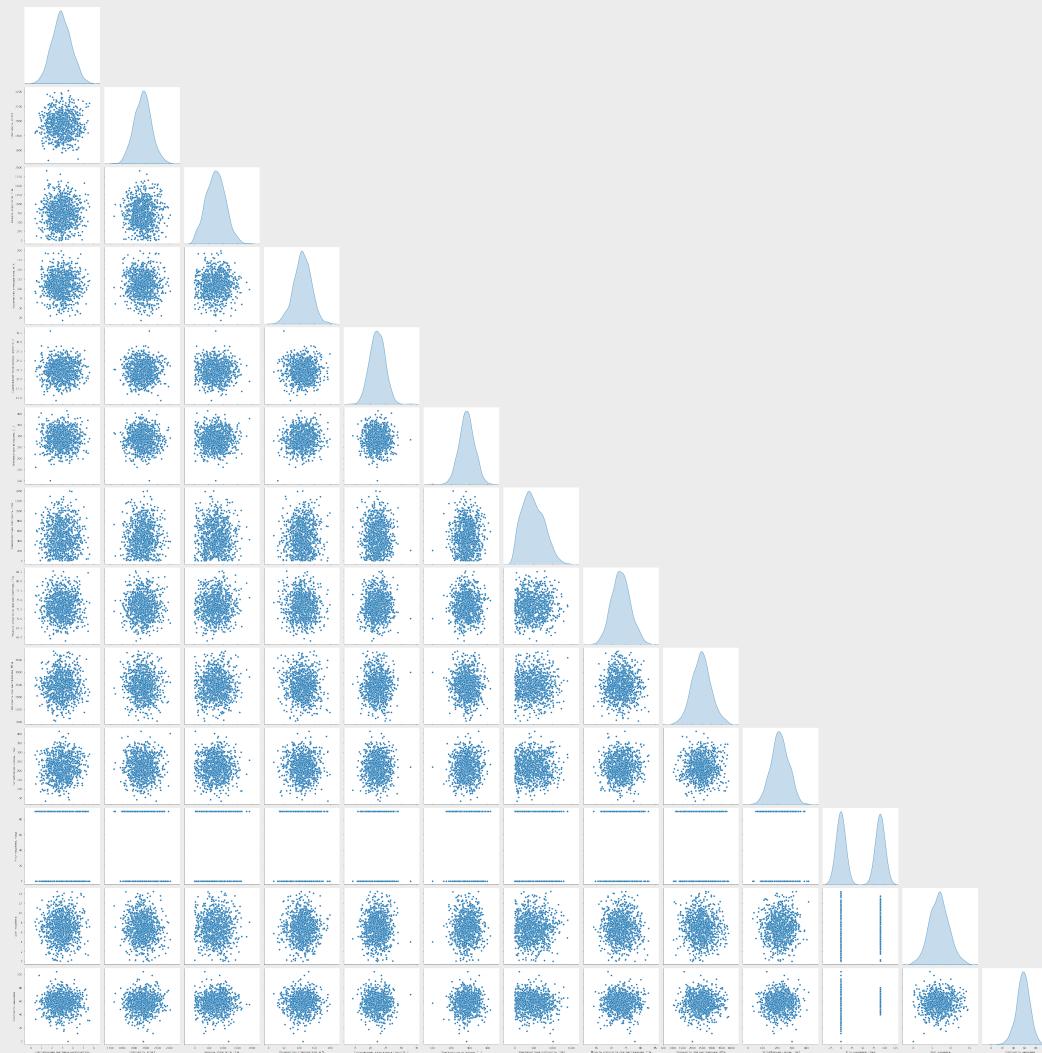
Название	Файл	Тип данных	Непустых значений	Уникальных значений
Соотношение матрица-наполнитель	X_bp	float64	1023	1014
Плотность, кг/м3	X_bp	float64	1023	1013
модуль упругости, ГПа	X_bp	float64	1023	1020
Количество отвердителя, м.%	X_bp	float64	1023	1005
Содержание эпоксидных групп,%_2	X_bp	float64	1023	1004
Температура вспышки, С_2	X_bp	float64	1023	1003
Поверхностная плотность, г/м2	X_bp	float64	1023	1004
Модуль упругости при растяжении, ГПа	X_bp	float64	1023	1004
Прочность при растяжении, МПа	X_bp	float64	1023	1004
Потребление смолы, г/м2	X_bp	float64	1023	1003
Угол нашивки, град	X_nip	float64	1023	2
Шаг нашивки	X_nip	float64	1023	989
Плотность нашивки	X_nip	float64	1023	988

	Среднее	Стандартное отклонение	Минимум	Максимум	Медиана
Соотношение матрица-наполнитель	2.9304	0.9132	0.3894	5.5917	2.9069
Плотность, кг/м3	1975.7349	73.7292	1731.7646	2207.7735	1977.6217
модуль упругости, ГПа	739.9232	330.2316	2.4369	1911.5365	739.6643
Количество отвердителя, м.%	110.5708	28.2959	17.7403	198.9532	110.5648
Содержание эпоксидных групп, %_2	22.2444	2.4063	14.2550	33.0000	22.2307
Температура вспышки, С_2	285.8822	40.9433	100.0000	413.2734	285.8968
Поверхностная плотность, г/м2	482.7318	281.3147	0.6037	1399.5424	451.8644
Модуль упругости при растяжении, ГПа	73.3286	3.1190	64.0541	82.6821	73.2688
Прочность при растяжении, МПа	2466.9228	485.6280	1036.8566	3848.4367	2459.5245
Потребление смолы, г/м2	218.4231	59.7359	33.8030	414.5906	219.1989
Угол нашивки, град	44.2522	45.0158	0.0000	90.0000	0.0000
Шаг нашивки	6.8992	2.5635	0.0000	14.4405	6.9161
Плотность нашивки	57.1539	12.3510	0.0000	103.9889	57.3419

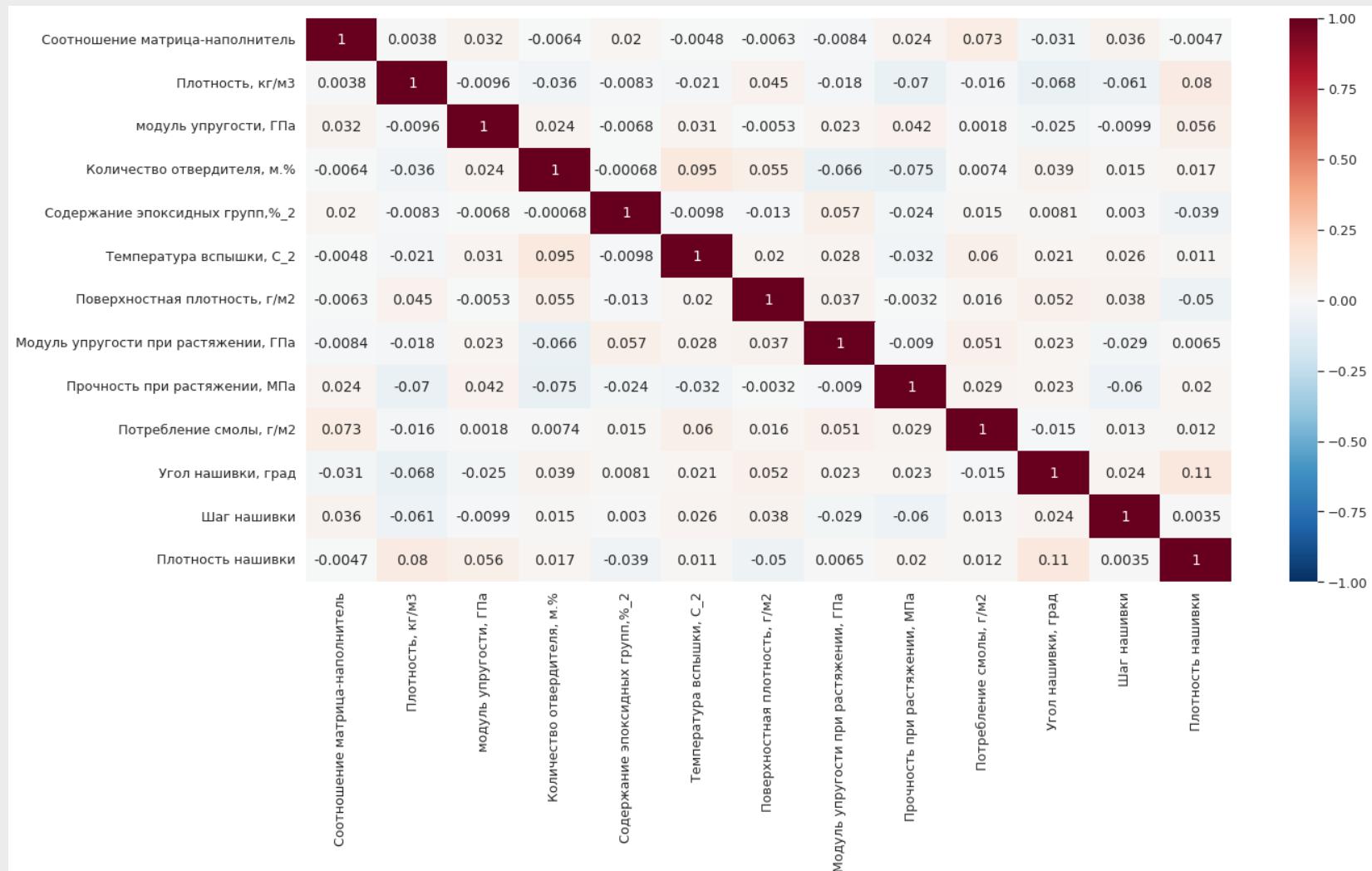
# Гистограммы распределения и диаграммы “ящик с усами”



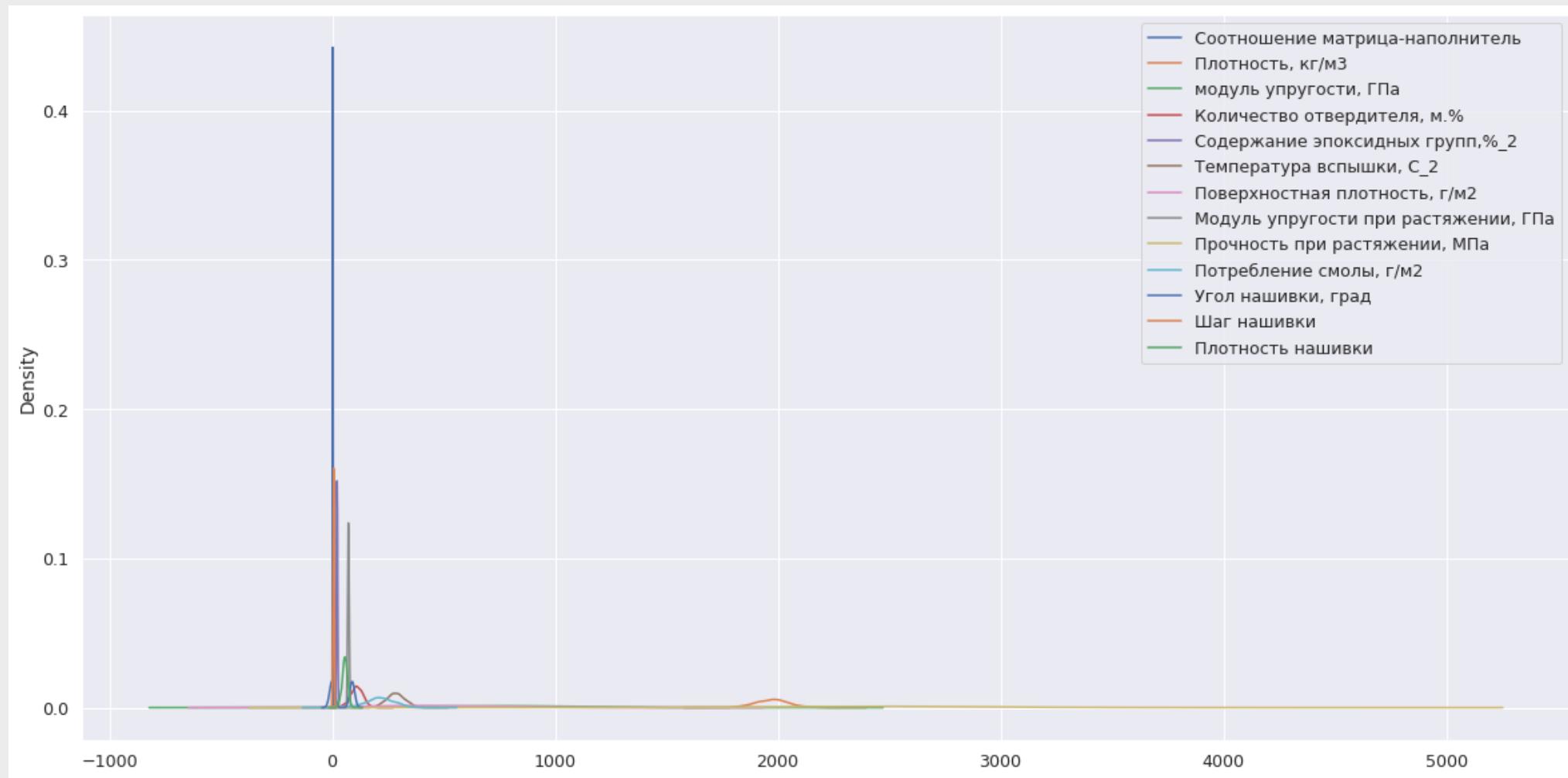
# Попарные графики рассеяния точек



# Матрица корреляции



# Оценка плотности ядра



# Метрики качества

- R<sup>2</sup> или коэффициент детерминации
- RMSE (Root Mean Squared Error) или корень из средней квадратичной ошибки
- MAE (Mean Absolute Error) или средняя абсолютная ошибка
- MAPE (Mean Absolute Percentage Error) или средняя абсолютная процентная ошибка
- max error или максимальная ошибка данной модели

# Модели

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

# Модель для модуля упругости при растяжении

		R2	RMSE	MAE	MAPE	max_error
	<b>dummy_regressor</b>	-0.004036	3.052892	2.463833	0.033723	9.159206
	<b>linear_regression</b>	-0.027923	3.088995	2.498601	0.034196	8.733386
	<b>ridge</b>	-0.013398	3.067093	2.478443	0.033920	8.865129
	<b>lasso</b>	-0.014245	3.068374	2.477739	0.033912	8.899312
	<b>svr</b>	-0.004368	3.053397	2.466900	0.033778	8.981321
	<b>knr</b>	-0.043621	3.112491	2.511645	0.034383	8.619261
	<b>dtr</b>	-0.004052	3.052917	2.463877	0.033724	9.158813
	<b>rfr</b>	-0.009565	3.061287	2.466319	0.033751	9.112006
	<b>catboost</b>	-0.205467	3.345142	2.692858	0.036816	9.461555
	<b>lgbm</b>	-0.169878	3.295392	2.616058	0.035757	9.868853
	<b>gbr</b>	-0.010713	3.063027	2.478294	0.033933	8.852253

# Модель для модуля упругости при растяжении



```
{'dtr__max_depth': 1, 'dtr__max_features': 1, 'dtr__splitter': 'random'}  
-0.0037561810653206074
```

# Модель для прочности при растяжении

	R2	RMSE	MAE	MAPE	max_error
<b>dummy_regressor</b>	-0.000354	497.415096	385.693931	0.172664	1425.668631
<b>linear_regression</b>	-0.036405	506.298770	399.029498	0.176525	1386.349118
<b>ridge</b>	-0.020680	502.443149	394.279111	0.175026	1385.486336
<b>lasso</b>	-0.026818	503.951672	396.274695	0.175689	1384.647916
<b>svr</b>	-0.001087	497.597435	385.807793	0.172221	1418.711630
<b>knr</b>	-0.020638	502.432849	395.623664	0.176149	1451.614910
<b>dtr</b>	-0.018310	501.859389	387.371770	0.173459	1413.536032
<b>rfr</b>	-0.024930	503.488044	396.013842	0.176199	1426.954893
<b>catboost</b>	-0.206094	546.176072	432.798954	0.192129	1640.488442
<b>lgbm</b>	-0.135702	529.997996	412.111379	0.182783	1575.283923
<b>gbr</b>	0.001306	497.002154	387.459996	0.172412	1410.583328

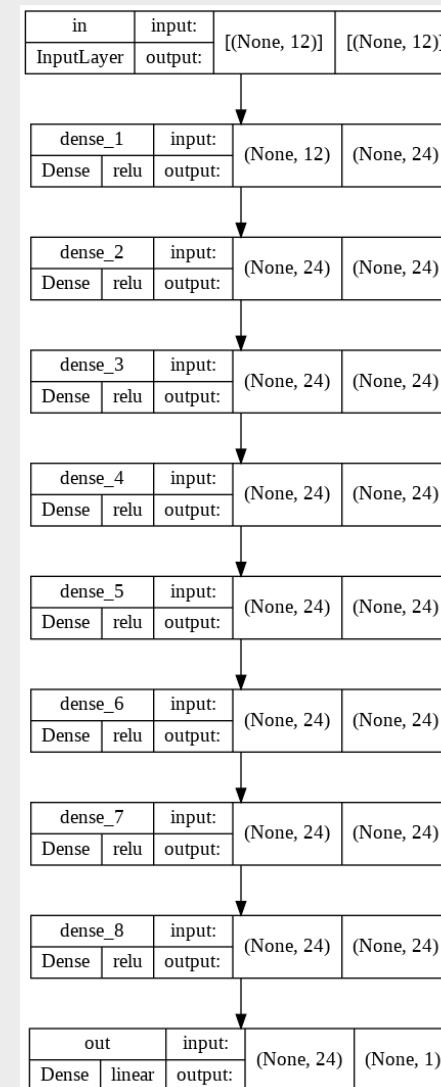
# Модель для прочности при растяжении



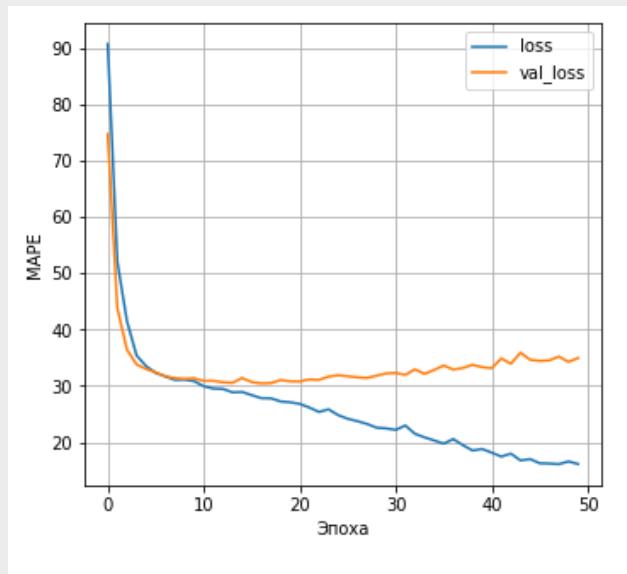
```
{'gbr__loss': 'squared_error',
 'gbr__max_depth': 2,
 'gbr__max_features': 1,
 'gbr__n_estimators': 50}
0.013064598263368551
```

# Модель для соотношения матрица-наполнитель

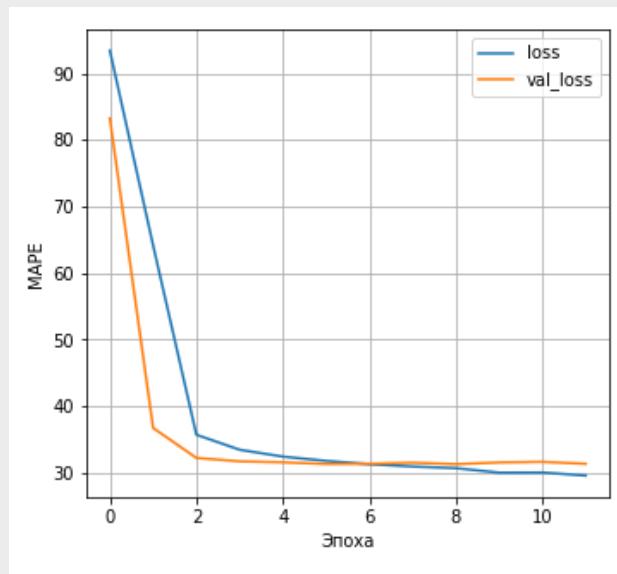
Layer (type)	Output Shape	Param #
<hr/>		
dense_1 (Dense)	(None, 24)	312
dense_2 (Dense)	(None, 24)	600
dense_3 (Dense)	(None, 24)	600
dense_4 (Dense)	(None, 24)	600
dense_5 (Dense)	(None, 24)	600
dense_6 (Dense)	(None, 24)	600
dense_7 (Dense)	(None, 24)	600
dense_8 (Dense)	(None, 24)	600
out (Dense)	(None, 1)	25
<hr/>		
Total params: 4,537		
Trainable params: 4,537		
Non-trainable params: 0		



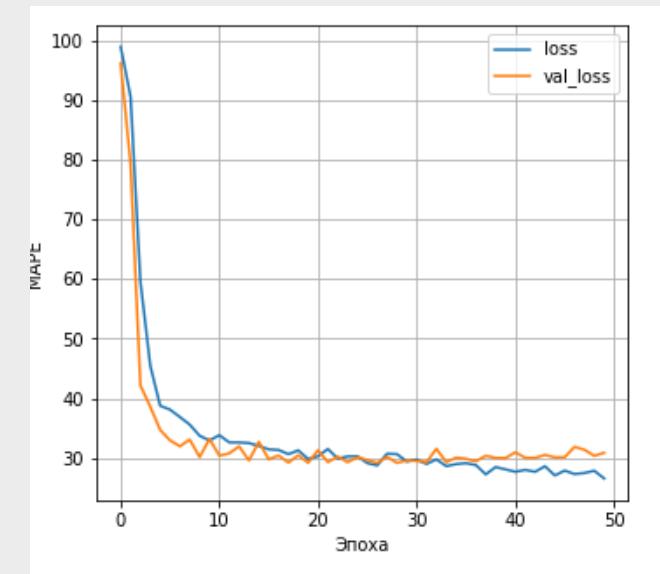
# Обучение нейросети

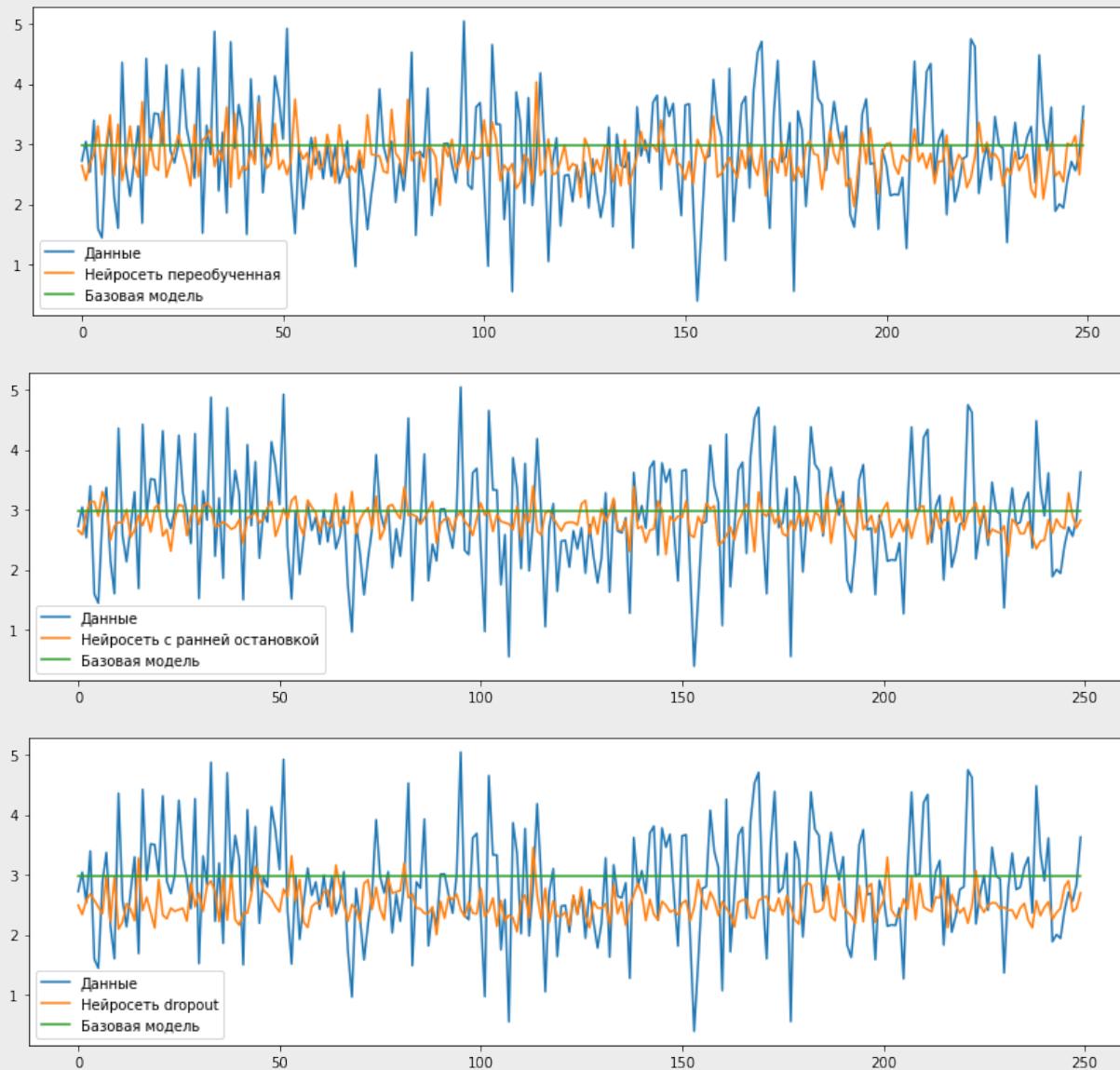


Борьба с  
переобучением:  
ранняя остановка



Борьба с  
переобучением:  
Dropout





	<b>R2</b>	<b>RMSE</b>	<b>MAE</b>	<b>MAPE</b>	<b>max_error</b>
<b>DummyRegressor</b>	-0.020151	0.882319	0.686249	0.341972	2.577898
<b>Нейросеть переобученная</b>	-0.257839	0.979728	0.775280	0.318282	2.679286
<b>Нейросеть с ранней остановкой</b>	-0.030224	0.886664	0.688314	0.323383	2.342318
<b>Нейросеть dropout</b>	-0.274153	0.986061	0.779692	0.324875	2.624417

# Разработка веб-приложения

## Выпускная квалификационная работа по курсу «Data Science»

Прогнозирование конечных свойств новых материалов (композиционных материалов)

Студентка Тамара Урушадзе

[Прогнозирование модуля упругости при растяжении и прочности при растяжении](#)

## Прогнозирование модуля упругости при растяжении и прочности при растяжении

Соотношение матрица-наполнитель  
 Плотность, кг/м<sup>3</sup>  
 Модуль упругости, ГПа  
 Количество отвердителя, м.-%  
 Содержание эпоксидных групп,%\_2  
 Температура вспышки, С\_2  
 Поверхностная плотность, г/м<sup>2</sup>  
 Потребление смолы, г/м<sup>2</sup>  
 Угол нашивки, град  
 Шаг нашивки  
 Плотность нашивки

# Разработка веб-приложения

## Прогнозирование модуля упругости при растяжении и прочности при растяжении

Содержание эпоксидных групп,%\_2 – не должно быть отрицательным числом  
Потребление смолы, г/м2 – could not convert string to float: 'test'  
Угол нашивки, град – должно быть 0 или 90

Соотношение матрица-наполнитель  
 Плотность, кг/м3  
 Модуль упругости, ГПа  
 Количество отвердителя, м.%  
 Содержание эпоксидных групп,%\_2  
 Температура вспышки, С\_2  
 Поверхностная плотность, г/м2  
 Потребление смолы, г/м2  
 Угол нашивки, град  
 Шаг нашивки  
 Плотность нашивки

## Прогнозирование модуля упругости при растяжении и прочности при растяжении

Соотношение матрица-наполнитель  
 Плотность, кг/м3  
 Модуль упругости, ГПа  
 Количество отвердителя, м.%  
 Содержание эпоксидных групп,%\_2  
 Температура вспышки, С\_2  
 Поверхностная плотность, г/м2  
 Потребление смолы, г/м2  
 Угол нашивки, град  
 Шаг нашивки  
 Плотность нашивки

Входные данные:

	Соотношение матрица-наполнитель	Плотность, кг/м3	модуль упругости, ГПа	Количество отвердителя, м.%	Содержание эпоксидных групп,%_2	Температура вспышки, С_2	Поверхностная плотность, г/м2	Потребление смолы, г/м2	Угол нашивки, град	Шаг нашивки	Плотность нашивки
0	1.857143	2030.0	738.736842	50.0	23.75	284.615385	210.0	220.0	0.0	4.0	60.0

Результат модели:

Модуль упругости при растяжении Прочность при растяжении  
73.3696021503352 2510.8182638288736

# Спасибо за внимание!



edu.bmstu.ru

+7 495 182-83-85

[edu@bmstu.ru](mailto:edu@bmstu.ru)

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