



ЦЕНТР
ДОПОЛНИТЕЛЬНОГО
ОБРАЗОВАНИЯ
МГТУ им. Н.Э. Баумана

Прогнозирование конечных свойств новых материалов (композиционных материалов)

Петров Александр Евгеньевич



Постановка задачи

1. изучить предметную область
2. провести разведочный анализ данных
3. разделить данные на тренировочную и тестовую выборки
4. разработать приложение
5. выполнить препроцессинг (предобработку)
6. выбрать базовую модель и модели для подбора
7. подобрать гиперпараметры с помощью с помощью поиска по сетке с перекрестной проверкой
8. сравнить модели после подбора гиперпараметров и выбрать лучшую
9. сравнить качество лучшей и базовой моделей на тестовой выборке
10. сравнить качество лучшей модели на тренировочной и тестовой выборке



Разведочный анализ данных

```
<class 'pandas.core.frame.DataFrame'>
```

```
Index: 1023 entries, 0 to 1022
```

```
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)
```

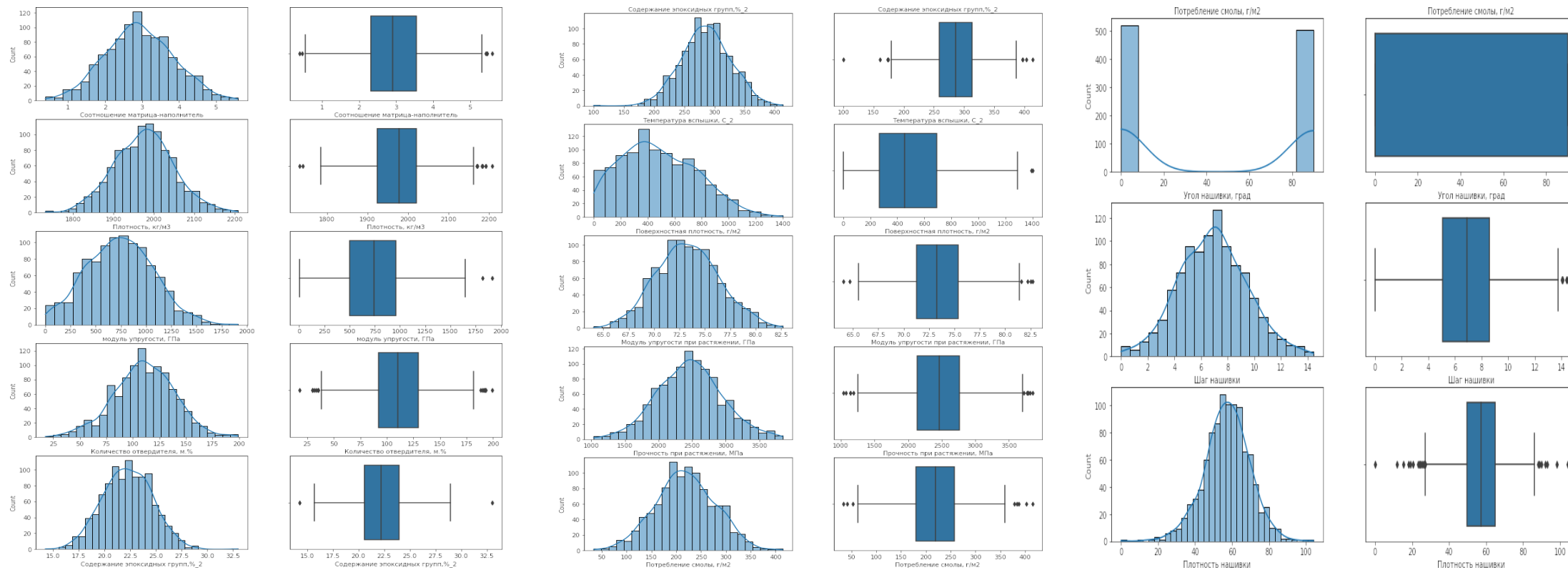
```
memory usage: 111.9 KB
```

	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

Описательная статистика признаков датасета



Аналитическая часть



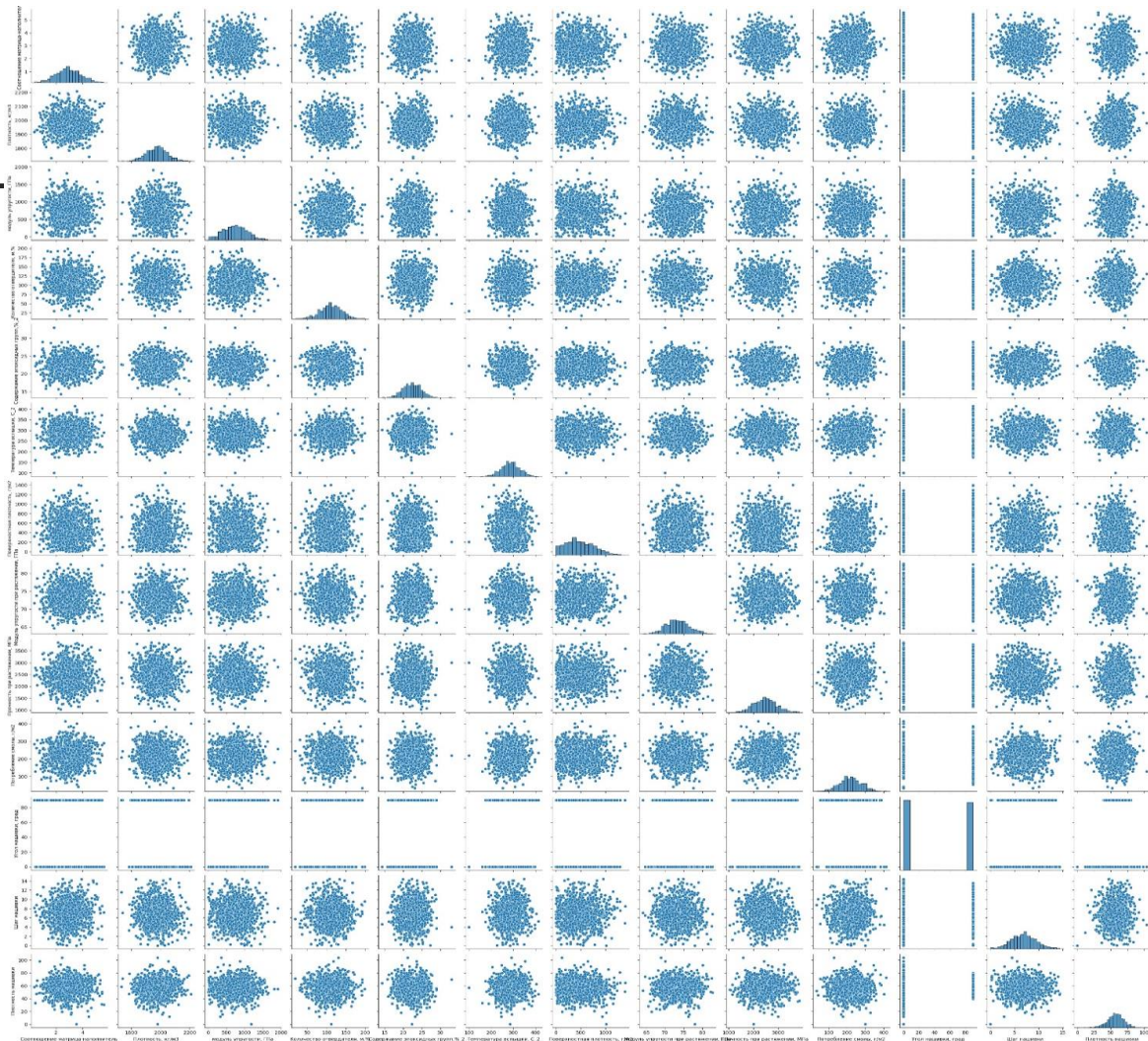
Гистограммы распределения и диаграммы “ящик с усами”



Аналитическая часть

Попарные графики рассеяния точек

- Выбросы есть
- Зависимостей нет





Аналитическая часть

Выбросы

	доля выбросов
Соотношение матрица-наполнитель	0.59%
Плотность, кг/м3	0.88%
модуль упругости, ГПа	0.20%
Количество отвердителя, м.%	1.37%
Содержание эпоксидных групп, %_2	0.20%
Температура вспышки, С_2	0.78%
Поверхностная плотность, г/м2	0.20%
Модуль упругости при растяжении, ГПа	0.59%
Прочность при растяжении, МПа	1.08%
Потребление смолы, г/м2	0.78%
Угол нашивки, град	0.00%
Шаг нашивки	0.39%
Плотность нашивки	2.05%

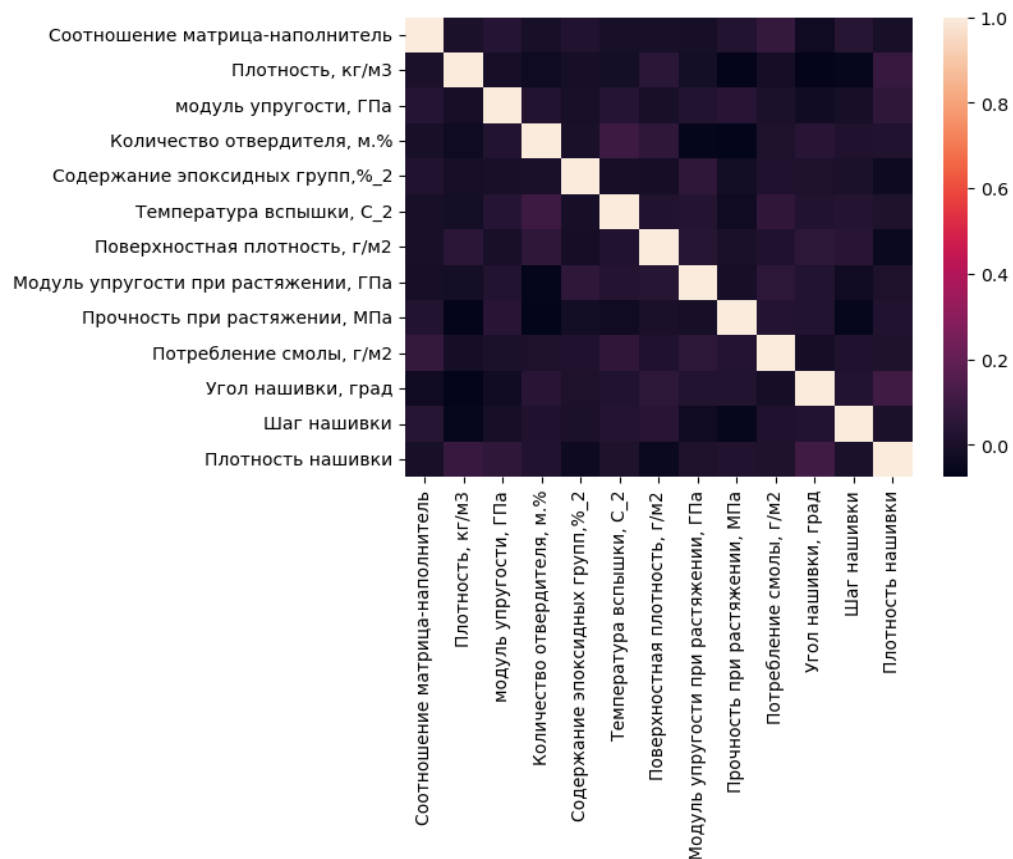
Межквартильный диапазон, часто сокращенно IQR, представляет собой разницу между 25-м процентилем (Q1) и 75-м процентилем (Q3) в наборе данных. Он измеряет разброс средних 50% значений.

Один из популярных методов состоит в том, чтобы объявить наблюдение выбросом, если его значение в 1,5 раза больше, чем IQR, или в 1,5 раза меньше, чем IQR.

Доля выбросов



Матрица корреляции признаков



Линейной зависимости нет



Описательная статистика выходного признака

Прочность при растяжении, МПа

count	936.000000
mean	2467.488822
std	463.838911
min	1250.392802
25%	2146.936034
50%	2457.959767
75%	2755.169485
max	3705.672523

Модуль упругости при растяжении, ГПа

count	936.000000
mean	73.305127
std	3.037381
min	65.553336
25%	71.248823
50%	73.259230
75%	75.310788
max	81.417126

Соотношение матрица-наполнитель

count	936.000000
mean	2.925683
std	0.893712
min	0.547391
25%	2.321931
50%	2.904731
75%	3.546650
max	5.314144

Для каждого признака — отдельная модель

- модуль упругости при растяжении
- прочность при растяжении
- соотношение матрица-наполнитель



Препроцессинг

	count	mean	std	min	25%	50%	75%	max
Соотношение матрица-наполнитель	936.0	-3.985416e-16	1.000535	-2.662564	-0.675917	-0.023457	0.695189	2.673947
Плотность, кг/м3	936.0	-1.040004e-15	1.000535	-2.678494	-0.714937	0.045471	0.651668	2.649773
модуль упругости, ГПа	936.0	1.480297e-16	1.000535	-2.239686	-0.729517	0.001489	0.670165	2.773444
Количество отвердителя, м.%	936.0	-4.934325e-17	1.000535	-2.673520	-0.680609	0.007288	0.706247	2.624101
Содержание эпоксидных групп, %_2	936.0	-8.264994e-16	1.000535	-2.721073	-0.684125	-0.010160	0.732284	2.818386
Температура вспышки, С_2	936.0	2.827748e-16	1.000535	-2.708660	-0.681975	-0.000414	0.684390	2.540085
Поверхностная плотность, г/м2	936.0	1.537232e-16	1.000535	-1.722572	-0.774851	-0.090207	0.760272	2.886532
Модуль упругости при растяжении, ГПа	936.0	-4.531987e-15	1.000535	-2.553495	-0.677361	-0.015119	0.660679	2.672150
Прочность при растяжении, МПа	936.0	7.591269e-16	1.000535	-2.625366	-0.691456	-0.020555	0.620548	2.670853
Потребление смолы, г/м2	936.0	-1.622634e-16	1.000535	-2.663276	-0.659631	0.013415	0.671035	2.447193
Угол нашивки, град	936.0	1.214603e-16	1.000535	-1.023787	-1.023787	0.976766	0.976766	0.976766
Шаг нашивки	936.0	-2.068621e-16	1.000535	-2.742040	-0.709873	0.011064	0.668120	2.717671
Плотность нашивки	936.0	-1.821904e-16	1.000535	-2.686557	-0.644710	0.011780	0.653975	2.542482

Поиск гиперпараметров по сетке

Поиск гиперпараметров по сетке реализует класс `GridSearchCV` из `sklearn`. Он получает модель и набор гиперпараметров, поочередно передает их в модель, выполняет обучение и определяет лучшие комбинации гиперпараметры. Перекрестная проверка уже встроена в этот класс.

Описательная статистика входных признаков до и после предобработки



Модели машинного обучения

- 1 **DummyRegressor** — базовая модель
- 2 **LinearRegression** — линейная регрессия
- 3 **Ridge** — гребневая регрессия
- 4 **RandomForestRegressor** — случайный лес
- 5 **KneighborsRegressor** — метод ближайших соседей
- 6 **Нейронная сеть**



Метрики качества

1

R² или коэффициент детерминации

2

MSE (Mean Squared Error) или корень из средней квадратичная ошибка

3

MAE (Mean Absolute Error) или средняя абсолютная ошибка

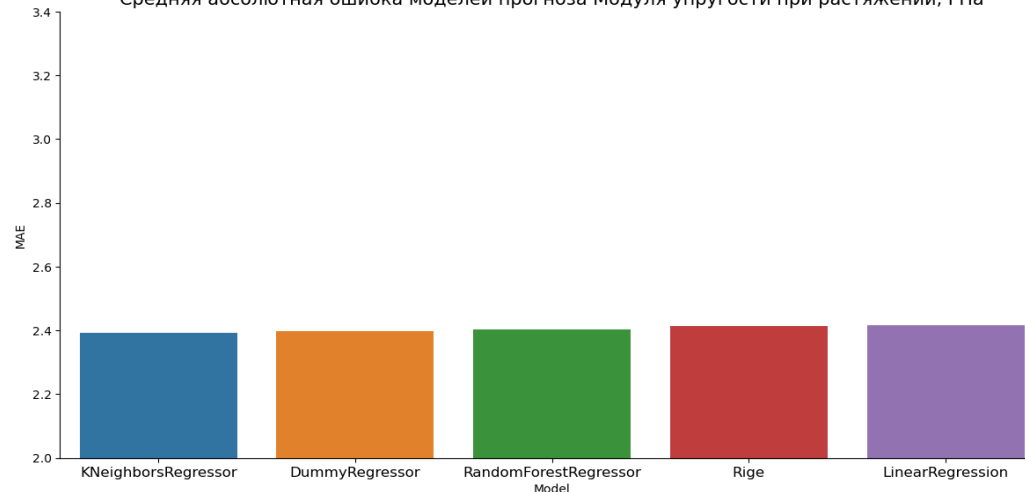


Практическая часть

Модель для модуля упругости при растяжении

	Model	MSE	MAE	R2 score
Модуль упругости при растяжение, ГПа	KNeighborsRegressor	8.678704	2.392482	0.00408
Модуль упругости при растяжении, ГПа	DummyRegressor	8.765902	2.398634	-0.00600
Модуль упругости при растяжение, ГПа	RandomForestRegressor	8.735981	2.404383	-0.00200
Модуль упругости при растяжение, ГПа	Ridge	8.732845	2.412596	-0.00200
Модуль упругости при растяжении, ГПа	LinearRegression	8.745602	2.416002	-0.00400

Средняя абсолютная ошибка моделей прогноза Модуля упругости при растяжении, ГПа



Результаты моделей после подбора гиперпараметров

	Model	MSE	MAE	R2 score
Модуль упругости при растяжении, ГПа	DummyRegressor	8.765902	2.398634	-0.00600
Модуль упругости при растяжение, ГПа	KNeighborsRegressor	8.678704	2.392482	0.00408

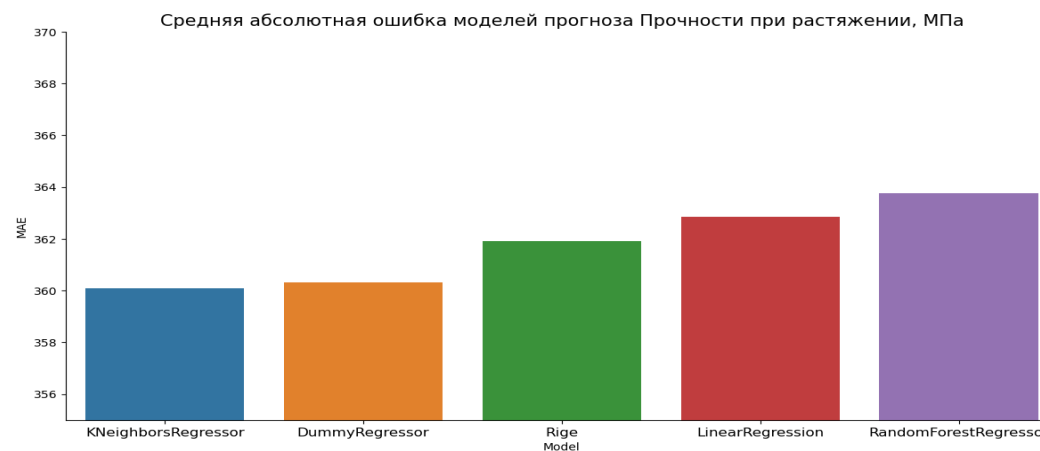
Метрики работы лучшей модели на тестовом множестве



Практическая часть

Модель для прочности при растяжении

	Model	MSE	MAE	R2 score
Прочность при растяжении, МПа	KNeighborsRegressor	209750.489526	360.097838	-0.028
Прочность при растяжении, МПа	DummyRegressor	210162.565630	360.306251	-0.030
Прочность при растяжении, МПа	Ridge	212349.035083	361.903011	-0.040
Прочность при растяжении, МПа	LinearRegression	213381.584134	362.854584	-0.045
Прочность при растяжении, МПа	RandomForestRegressor	212325.620988	363.774842	-0.040



Результаты моделей после подбора гиперпараметров

	Model	MSE	MAE	R2 score
Прочность при растяжении, МПа	DummyRegressor	210162.565630	360.306251	-0.030
Прочность при растяжении, МПа	KNeighborsRegressor	209750.489526	360.097838	-0.028

Метрики работы лучшей модели на тестовом множестве



Практическая часть

Модель для соотношения матрица наполнитель

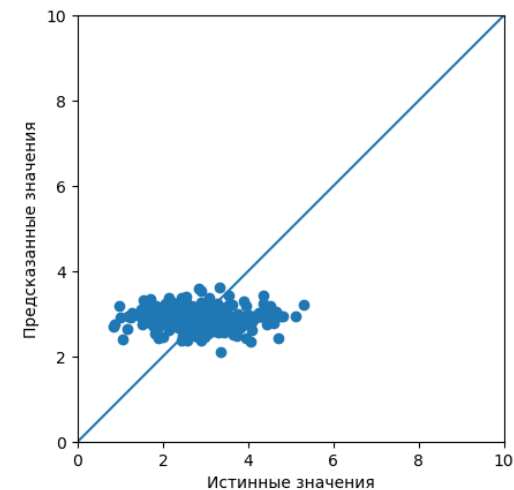
Layer (type)	Output Shape	Param #
dense_112 (Dense)	(None, 12)	156
batch_normalization_57 (BatchNormalization)	(None, 12)	48
dense_113 (Dense)	(None, 4)	52
batch_normalization_58 (BatchNormalization)	(None, 4)	16
dense_114 (Dense)	(None, 1)	5

Total params: 277 (1.08 KB)

Trainable params: 245 (980.00 B)

Non-trainable params: 32 (128.00 B)

Архитектура нейросети из библиотеки keras



Предсказание на тестовой выборке



Практическая часть

График потерь модели

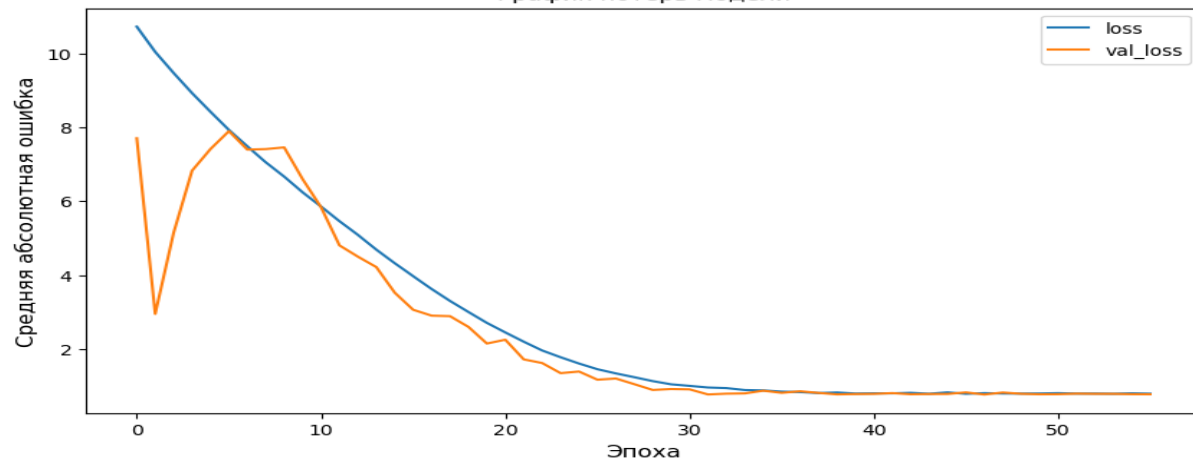
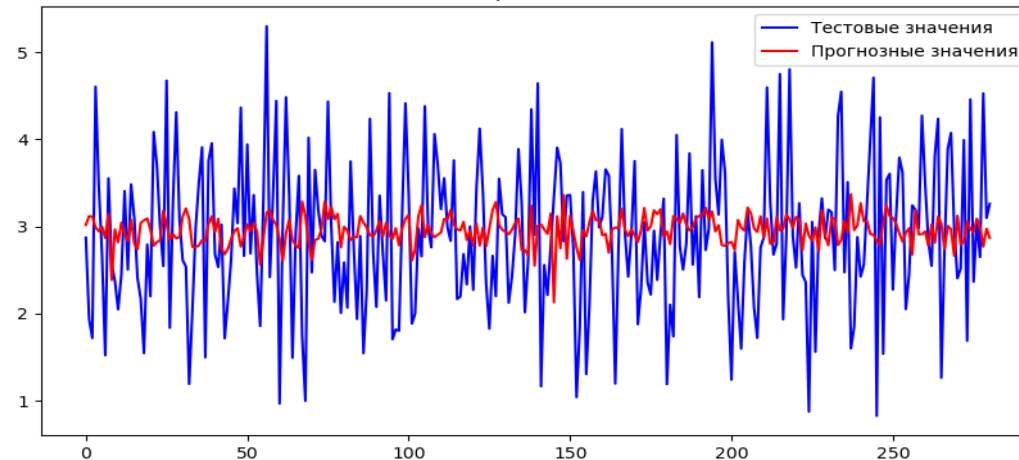


График потерь модели на тренировочной и тестовой выборках

Тестовые и прогнозные значения



Результат работы нейросети



Практическая часть

	Model	MSE	MAE	R2 score
Матрица-наполнитель	DummyRegressor	0.762885	0.694175	-0.002
Матрица-наполнитель	Нейронная сеть	0.781559	0.714372	-0.026

Метрики работы нейросети и базовой модели

	Model	MSE	MAE	R2 score
Матрица-наполнитель trein	Нейронная сеть	0.862030	0.755862	-0.06
Матрица-наполнитель test	Нейронная сеть	0.876235	0.765448	-0.15

Сравнение ошибок модели для соотношения матрица-наполнитель на тренировочном и тестовом датасете.



Разработка веб-приложения

Входные параметры

Соотношение матрица-наполнитель

3,000000

-

+

Плотность, кг/м3

2000,000000

-

+

модуль упругости, ГПа

2000,000000

-

+

Количество отвердителя, м.%

2000,000000

-

+

Содержание эпоксидных групп, %_2

2000,000000

-

+

Температура вспышки, C_2

2000,000000

-

+

Поверхностная плотность, г/м2

2000,000000

-

+

Потребление смолы, г/м2

2000,000000

-

+

Угол нашивки, град

2000,000000

-

+

Шаг нашивки

2000,000000

-

+

Плотность нашивки

2000,000000

-

+

Предсказание свойств материала

Предсказание для прочности при растяжении: 2421.153 МПа

Предсказание для модуля упругости при растяжении: 73.215 ГПа



Результаты

Задача не решена

1. Дальнейшие поиски решения могли бы включать:
2. проконсультироваться у экспертов
3. уточнить постановку задачи
4. исследовать сырые данные
5. провести отбор признаков и уменьшение размерности
6. углубиться в нейросети

Спасибо за внимание!



ЦЕНТР
ДОПОЛНИТЕЛЬНОГО
ОБРАЗОВАНИЯ
МГТУ им. Н.Э. Баумана



do.bmstu.ru