



ЦЕНТР
ДОПОЛНИТЕЛЬНОГО
ОБРАЗОВАНИЯ
МГТУ им. Н.Э. Баумана

Прогнозирование свойств КОМПОЗИТНЫХ материалов

Шипов Михаил Всеволодович



Постановка задачи

Построить модели для прогноза конечных свойств получаемых композиционных материалов:

1 Модуль упругости при растяжении

2 Прочность при растяжении

3 Соотношение матрица-наполнитель
Нейронная сеть

Композиционные материалы обладают следующими характеристиками

- экологическая безопасность;
- прочность на изгиб, сжатие, растяжение;
- стойкость к механическим и вибрационным нагрузкам;
- упругость, жесткость, удельный вес;
- износостойкость, электропроводимость;
- тепло- и звукоизоляционные характеристики;
- химическая, биологическая, коррозионная устойчивость



Характеристика датасета

Данные представлены двумя xlsx файлами

Размерность: X_br (1023, 10), X_upr (1040, 3)

Признаки:

- 12 – количественных (float64);
- 1 – категориальный «Угол нашивки» (int64)

Пустые значения: отсутствуют

Вероятная ошибка в данных:

- X_upr – первые 40 строк – повторяющиеся значения;
- X_br – первые 23 строки – протянутые значения;
- Без них общее кол-во строк обоих файлов 1000

Строки удалены из датасета со сбросом индексов.
Итоговая размерность общего датасета (1000, 13).
Количественные признаки – уникальные значения

X_upr.xlsx

	Угол нашивки, гр	Шаг нашивки	Плотность нашивки	
0	0	4	57	
1	0	4	60	
2	0	4	70	
3	0	5	47	
4	0	5	57	
5	0	5	60	
6	0	5	70	
7	0	7	47	
8	0	7	57	
9	0	7	60	
10	0	7	70	
11	0	9	47	
12	0	9	57	
13	0	9	60	
14	0	9	70	
15	0	10	47	

Угол нашивки, гр	Шаг нашивки	Плотность нашивки	
0	0	57	
1	0	4	60
2	0	4	70
3	0	5	47
4	0	5	57
5	0	5	60
6	0	5	70
7	0	7	47
8	0	7	57
9	0	7	60
10	0	7	70
11	0	9	47
12	0	9	57
13	0	9	60
14	0	9	70
15	0	10	47
16	0	10	57
17	0	10	60
18	0	10	70
19	0	0	0
20	90	4	47
21	90	4	57
22	90	4	60
23	90	4	70
24	90	5	47
25	90	5	57
26	90	5	60
27	90	5	70
28	90	7	47
29	90	7	57
30	90	7	60
31	90	7	70
32	90	9	47
33	90	9	57
34	90	9	60
35	90	9	70
36	90	10	47
37	90	10	57
38	90	10	60
39	90	10	70
40	0	7,856166547	64,30196385
41	0	7,401542567	19,25053314
42	0	6,675780339	76,62329934

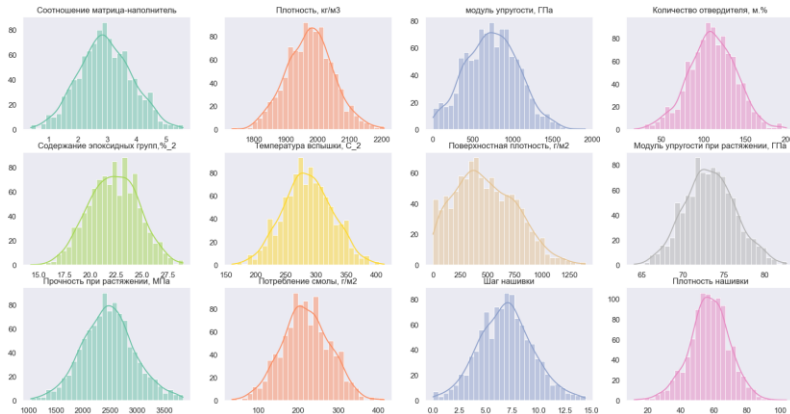
X_br.xlsx

	Соотношение м	Плотность, кг/м ²	модуль упругости	Количество откл	Содержание элс	Температура всл	Поверхностная г	Модуль упругости	Прочность при р	Потребление смолы, г/м ²	
0	1,857142857	2030	738,7368421	30	22,26785714	100	210	70	3000	220	
1	1,857142857	2030	738,7368421	50	23,75	284,6153846	210	70	3000	220	
2	1,857142857	2030	738,7368421	49,9	33	284,6153846	210	70	3000	220	
3	1,857142857	2030	738,7368421	129	21,25	300	210	70	3000	220	
4	2,771331058	2030	753	111,86	22,26785714	284,6153846	210	70	3000	220	
5	2,767918089	2000	748	111,86	22,26785714	284,6153846	210	70	3000	220	
6	2,569620253	1910	807	111,86	22,26785714	284,6153846	210	70	3000	220	
7	2,56147541	1900	535	111,86	22,26785714	284,6153846	380	75	1800	120	
8	3,557017541	1930	889	129	21,25	300	380	75	1800	120	
9	3,532338308	2100	1421	129	21,25	300	1010	78	2000	300	
10	2,919677836	2160	933	129	21,25	300	1010	78	2000	300	
11	2,877358491	1990	1628	129	21,25	300	1010	78	2000	300	
12	1,598173516	1950	827	129	21,25	300	470	73,33333333	2455,555556	220	
13	2,919677836	1980	568	129	21,25	300	470	73,33333333	2455,555556	220	
14	4,029126214	1910	800	129	21,25	300	470	73,33333333	2455,555556	220	
15	2,934782609	2030	302	129	21,25	300	210	70	3000	220	
16	3,557017541	1880	313	129	21,25	300	210	70	3000	220	
17	4,193548387	1950	506	129	21,25	300	380	75	1800	120	
18	4,897959184	1890	540	129	21,25	300	380	75	1800	120	
19	3,532338308	1980	1183	111,86	22,26785714	284,6153846	1010	78	2000	300	
20	2,877358491	2000	205	111,86	22,26785714	284,6153846	1010	78	2000	300	
21	1,598173516	1920	456	111,86	22,26785714	284,6153846	470	73,33333333	2455,555556	220	
22	4,029126214	1880	622	111,86	22,26785714	284,6153846	470	73,33333333	2455,555556	220	

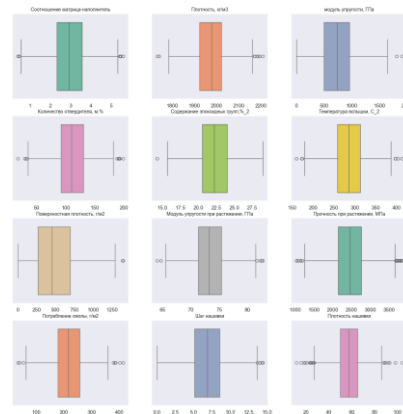


Разведочный анализ и предобработка данных

Анализ распределения данных и кодирование категориального признака



Распределение количественных признаков



Ящик с усами

	count	mean	std	min	25%	50%	75%	max
Соотношение матрица-наполнитель	1000.00	2.93	0.91	0.39	2.32	2.91	3.55	5.59
Плотность, кг/м3	1000.00	1975.67	73.80	1731.76	1924.20	1977.57	2021.16	2207.77
модуль упругости, ГПа	1000.00	739.95	330.33	2.44	498.44	741.15	962.85	1911.54
Количество отвердителя, м.%	1000.00	110.54	28.30	17.74	92.17	110.16	130.31	198.95
Содержание эпоксидных групп,%_2	1000.00	22.24	2.41	14.25	20.56	22.23	23.98	28.96
Температура вспышки, C_2	1000.00	285.91	40.96	160.26	258.54	285.85	313.58	413.27
Поверхностная плотность, г/м2	1000.00	483.02	280.81	0.60	268.06	452.97	694.21	1399.54
Модуль упругости при растяжении, ГПа	1000.00	73.33	3.12	64.05	71.30	73.25	75.38	82.68
Прочность при растяжении, МПа	1000.00	2467.18	485.62	1036.86	2143.83	2461.25	2760.16	3848.44
Потребление смолы, г/м2	1000.00	218.39	59.82	33.80	179.19	217.28	257.50	414.59
Угол нашивки, град	1000.00	45.00	45.02	0.00	0.00	45.00	90.00	90.00
Шаг нашивки	1000.00	6.91	2.56	0.04	5.14	6.91	8.57	14.44
Плотность нашивки	1000.00	57.25	12.34	11.74	49.97	57.41	65.11	103.99

Статистические характеристики

1. Распределение значений по всем количественным признакам является близким к нормальному;
2. Размах и масштаб – необходима нормализация данных;
3. Наличие выбросов в данных;
4. Угол нашивки: 500 значений 0 градусов, 500 – 90 градусов

```
labelencoder = LabelEncoder()
df['Угол нашивки, град'] = labelencoder.fit_transform(df['Угол нашивки, град'])
df['Угол нашивки, град'].value_counts()

✓ 0.0s

Угол нашивки, град
0    500
1    500
Name: count, dtype: int64
```



Разведочный анализ

Нахождение связей между признаками

Корреляция признаков композитных материалов около нулевая, изменение одной переменной не влияет на увеличение/уменьшение другой.

Для поиска нелинейной связи воспользуемся методом Phik, основанным на хи-квадрат тесте Пирсона.

```
phik_overview = df.phik_matrix()  
phik_overview['Модуль упругости при растяжении, ГПа'].sort_values(ascending=False)
```

В результате находим более значимые признаки, которые следует учесть при построении моделей

Модуль упругости при растяжении

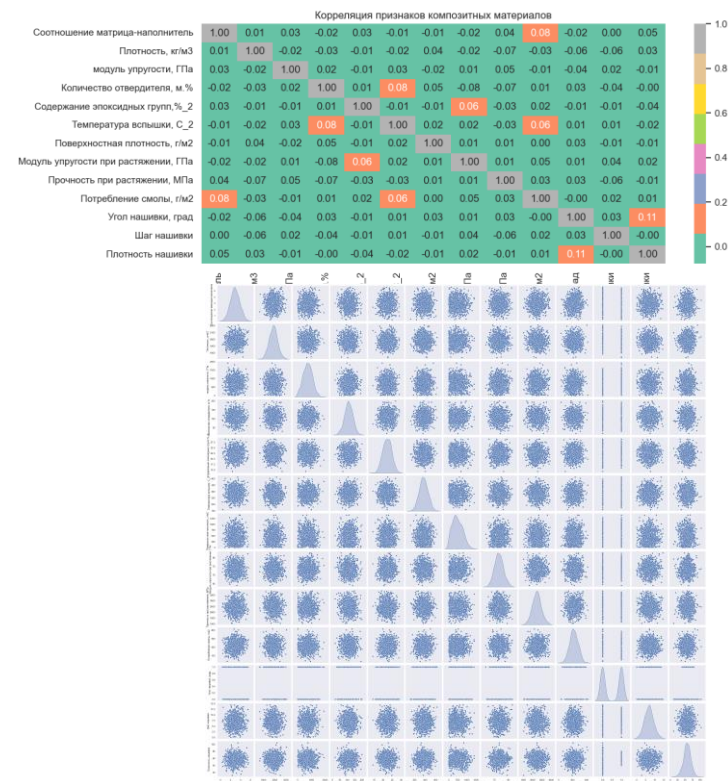
Модуль упругости при растяжении, ГПа	1.00
Шаг нашивки	0.26
модуль упругости, ГПа	0.17
Потребление смолы, г/м2	0.15
Прочность при растяжении, МПа	0.15
Угол нашивки, град	0.08

Прочность при растяжении

Прочность при растяжении, МПа	1.00
Количество отвердителя, м.%	0.22
Модуль упругости при растяжении, ГПа	0.15
Соотношение матрица-наполнитель	0.10
Потребление смолы, г/м2	0.10
Поверхностная плотность, г/м2	0.08
Плотность нашивки	0.08

Соотношение матрица-наполнитель

Соотношение матрица-наполнитель	1.00
Температура вспышки, С_2	0.13
Прочность при растяжении, МПа	0.10
Потребление смолы, г/м2	0.10
Плотность, кг/м3	0.06
Шаг нашивки	0.04



Корреляция признаков и график попарного рассеяния точек



Предобработка данных

Подготовка датасетов Упругости и Прочности к моделированию

Разбивка на обучающую и тестовую выборки

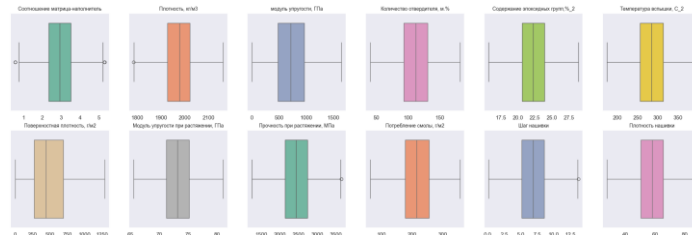
- Метод `train_test_split()`
- 30% - тестовая выборка

```
df_train, df_test = train_test_split(df, test_size=0.3, random_state=42)
df_train.shape, df_test.shape
✓ 0.0s
((700, 13), (300, 13))
```

Устранение выбросов

- Для обучающей выборки
- Метод межквартильного интервала

$$Q_1 - k * IQR, Q_3 + k * IQR$$



Ящик с усами

Нормирование данных

- Для каждого датасета
- Метод `MinMaxScaler()`

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

	count	mean	std	min	25%	50%	75%	max
Соотношение матрица-наполнитель	642.00	0.50	0.19	0.00	0.38	0.50	0.62	1.00
Плотность, кг/м3	642.00	0.51	0.18	0.00	0.38	0.51	0.63	1.00
модуль упругости, ГПа	642.00	0.44	0.20	0.00	0.29	0.43	0.58	1.00
Количество отвердителя, м.%	642.00	0.51	0.19	0.00	0.37	0.51	0.64	1.00
Содержание эпоксидных групп,%_2	642.00	0.50	0.18	0.00	0.37	0.50	0.62	1.00
Температура вспышки, С_2	642.00	0.50	0.19	0.00	0.37	0.49	0.63	1.00
Поверхностная плотность, г/м2	642.00	0.37	0.21	0.00	0.21	0.34	0.54	1.00
Прочность при растяжении, МПа	642.00	0.50	0.19	0.00	0.37	0.49	0.62	1.00
Потребление смолы, г/м2	642.00	0.52	0.20	0.00	0.39	0.52	0.66	1.00
Угол нашивки, град	642.00	0.52	0.50	0.00	0.00	1.00	1.00	1.00
Шаг нашивки	642.00	0.50	0.19	0.00	0.37	0.50	0.62	1.00
Плотность нашивки	642.00	0.50	0.19	0.00	0.37	0.50	0.63	1.00

Распределение количественных признаков



Разработка моделей

Формулировка гипотез

Количество признаков

Большее количество признаков позволяет лучше описать закономерности в условиях отсутствия корреляции между отдельными признаками

Влияние выбросов

Выбросы стоит удалять по минимуму, т.к. кол-во значений небольшое

Сложность модели

Ансамблевые модели работают лучше в силу более сложных алгоритмов

Применение гиперпараметров

Настройка гиперпараметров позволит улучшить качество модели



Разработка моделей

Результат проверки гипотез

Количество признаков

Решение дающее лучший результат

Необходимо сокращение признаков: Φ_{ik} для Модуля упругости при растяжении; PC для Прочности при растяжении

Влияние выбросов

Удаление «более грубым» методом межквартильного интервала

Сложность модели

Топ 3: Линейная регрессия, Случайный лес и Градиентный бустинг. Низкая ошибка достигается минимизацией отклонения от среднего

Применение гиперпараметров

После настройки гиперпараметров модели стали вести себя как линейная регрессия за исключением Catboost, который сохранил высокий R^2 на обучающей выборке и низкий на тестовой, что говорит о переобучении



Примеры предсказаний

Модуль упругости при растяжении

	R2_train	R2	MSE	RMSE	MAE	MAPE
lin_reg	0.01	-0.00	10.65	3.26	2.68	0.04
random_forest_reg	0.85	-0.08	11.46	3.38	2.75	0.04
knn_reg	0.23	-0.16	12.27	3.50	2.85	0.04
bag_reg	0.79	-0.13	11.95	3.46	2.83	0.04
gb_reg	0.42	-0.11	11.75	3.43	2.80	0.04
cat_reg	0.82	-0.23	13.00	3.61	2.93	0.04

Прочность при растяжении

	R2_train	R2	MSE	RMSE	MAE	MAPE
lin_reg	0.01	-0.01	226330.27	475.74	371.20	0.16
random_forest_reg	0.84	-0.06	235971.66	485.77	380.53	0.16
knn_reg	0.26	-0.21	269600.81	519.23	398.43	0.17
bag_reg	0.78	-0.12	249817.30	499.82	384.45	0.16
gb_reg	0.45	-0.07	239125.25	489.00	379.67	0.16
cat_reg	0.86	-0.15	255965.80	505.93	402.07	0.17



Phik

PC

Hyper + Phik



Выбор модели

Возможно выбрана не та модель?

	Adjusted R-Squared	R-Squared	RMSE	Time Taken
Model				
BayesianRidge	-0.05	-0.01	473.79	0.02
TweedieRegressor	-0.05	-0.01	473.90	0.02
ElasticNetCV	-0.05	-0.01	473.91	0.19
GammaRegressor	-0.05	-0.01	473.98	0.01
ElasticNet	-0.05	-0.01	474.42	0.02
DummyRegressor	-0.05	-0.01	474.61	0.02
LassoCV	-0.05	-0.01	474.73	0.15
LarsCV	-0.05	-0.01	474.74	0.04
LassoLarsCV	-0.05	-0.01	474.74	0.04
SVR	-0.05	-0.01	475.07	0.09
LassoLarsIC	-0.05	-0.01	475.30	0.02
NuSVR	-0.06	-0.01	475.65	0.06
SGDRegressor	-0.06	-0.02	476.43	0.02
RidgeCV	-0.06	-0.02	476.59	0.02
Lasso	-0.06	-0.02	476.60	0.02
LassoLars	-0.06	-0.02	476.60	0.02
Ridge	-0.06	-0.02	476.73	0.01
Lars	-0.06	-0.02	476.74	0.02
TransformedTargetRegressor	-0.06	-0.02	476.74	0.02
LinearRegression	-0.06	-0.02	476.74	0.01

Для повышения качества модели требуется:

- большой объем данных для обучения;
- перепроверка исходных данных на предмет ошибок заполнения

* Прогноз качества моделей получен при помощи LazyRegressor()



Построение нейронной сети

Подготовка данных и этапы разработки архитектуры нейронной сети для модели соотношения матрица-наполнитель

Подготовка датасета	Разбивка на обучающую (70%) и тестовую (30%) выборки. Встроенный слой нормализации	Необходимо сокращение количества признаков
Функция активации	Активации 'relu', 'tanh', 'sigmoid'. Активация 'linear' на выходе	'tanh' дает более стабильный результат. 'sigmoid' – dummy. На выходе не стоит задавать функцию активации
Сложность модели	Кол-во слоев активации 1-9. (i) Кол-во нейронов в слое 10-10*i	Увеличение числа слоев/нейронов приводит к «спрямлению» прогнозных результатов
Применение дополнительных слоев	Dropout, Batchnormalization	Качество модели повышается при использовании обоих слоев



Построение нейронной сети

Лучшие модели по итогам различных конфигураций нейронной сети

Модель 1

Model: "sequential_107"

Layer (type)	Output Shape	Param #
normalization_1 (Normalization)	(None, 5)	11
dense_496 (Dense)	(None, 10)	60
dropout_366 (Dropout)	(None, 10)	0
batch_normalization_348 (Batch Normalization)	(None, 10)	40
dense_497 (Dense)	(None, 1)	11

Total params: 122
Trainable params: 91
Non-trainable params: 31



	MSE	RMSE	MAE	R2	MAPE
HC tanh 1	0.85	0.92	0.74	0.02	0.34

- Normalization
- Dense (10, 'tanh')
- Dropout
- BatchNormalization
- Dense(1)

Модель 2

Model: "sequential_76"

Layer (type)	Output Shape	Param #
normalization_1 (Normalization)	(None, 5)	11
dense_343 (Dense)	(None, 10)	60
dense_344 (Dense)	(None, 10)	110
dense_345 (Dense)	(None, 1)	11

Total params: 192
Trainable params: 181
Non-trainable params: 11



	MSE	RMSE	MAE	R2	MAPE
HC без dropout & batch 2	0.90	0.95	0.77	-0.04	0.36

Общие параметры

- Sequential()
- Скорость обучения 0.001
- loss = mse
- Кол-во эпох 1000
- Ранняя остановка

- Normalization
- Dense (10, 'relu')
- Dense(1)



Приложение на Flask

Прогнозирование соотношения матрица-наполнитель

Введите значения

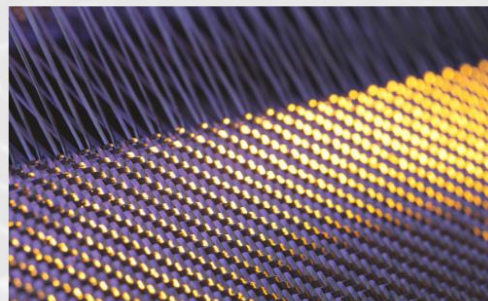
Плотность, кг/м^3 , (1731..2207)

Температура вспышки, $^{\circ}\text{C}_2$, (160..413)

Прочность при растяжении, МПа, (1036..3848)

Потребление смолы, г/м^2 , (34..414)

Шаг нашивки, (0..14)



Результат прогноза

Соотношение матрица-наполнитель для переданных параметров
2.9476655

Выпускная квалификационная работа по курсу "DataScience Pro"

Тема: Прогнозирование соотношения матрица-наполнитель

Выберите количество признаков для прогноза

[Основные 5
признаков](#)

[Все признаки](#)

Прогнозирование соотношения матрица-наполнитель

Введите значения

Плотность, кг/м^3 , (1731..2207)

Модуль упругости, ГПа, (2..1911)

Количество отвердителя, м.%, (18..198)

Содержание эпоксидных групп, %₂, (14..29)

Температура вспышки, $^{\circ}\text{C}_2$, (160..413)

Поверхностная плотность, г/м^2 , (1..1400)

Модуль упругости при растяжении, ГПа, (64..82)

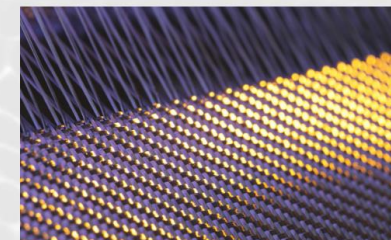
Прочность при растяжении, МПа, (1036..3848)

Потребление смолы, г/м^2 , (34..414)

Угол нашивки, (0, 90)

Шаг нашивки, (0..14)

Плотность нашивки, (12..104)



Результат прогноза

Соотношение матрица-наполнитель для переданных параметров
2.838562



ЦЕНТР
ДОПОЛНИТЕЛЬНОГО
ОБРАЗОВАНИЯ
МГТУ им. Н.Э. Баумана

Удаленный репозиторий

mshipov / BMSTU_VKR

[Code](#) [Issues](#) [Pull requests](#) [Actions](#) [Projects](#) [Wiki](#) [Security](#) [Insights](#) [Settings](#)

BMSTU_VKR Public Pin Unwatch 1

main 1 Branch 0 Tags

Add file Code

mshipov Update app.py 9aff050 · 25 minutes ago 13 Commits

Datasets	Add files via upload	27 minutes ago
models	Add files via upload	28 minutes ago
static	Add files via upload	28 minutes ago
templates	Add files via upload	28 minutes ago
README.md	Update README.md	40 minutes ago
VKR.ipynb	Add files via upload	44 minutes ago
app.py	Update app.py	25 minutes ago

README

BMSTU_VKR

Выпускная квалификационная работа по курсу "Data Science Pro" МГТУ им. Н. Э. Баумана Тема "Прогнозирование конечных свойств композиционных материалов" Цель работы - изучение способов прогнозирования конечных свойств новых композиционных материалов и разработка моделей для выполнения прогнозов.



ЦЕНТР
ДОПОЛНИТЕЛЬНОГО
ОБРАЗОВАНИЯ
МГТУ им. Н.Э. Баумана



do.bmstu.ru