



ЦЕНТР
ДОПОЛНИТЕЛЬНОГО
ОБРАЗОВАНИЯ
МГТУ им. Н.Э. Баумана

Прогнозирование конечных свойств новых материалов (композиционных материалов)

Островерх Владимир Степанович



ЦЕНТР
ДОПОЛНИТЕЛЬНОГО
ОБРАЗОВАНИЯ
МГТУ им. Н.Э. Баумана

В рамках ВКР было выполнено

1

Анализ данных

2

Подготовка данных

3

Подбор и обучение моделей машинного обучения

4

Создание и обучение нейронной сети

5

Создание приложения Flask





Обзор данных

Изучены характеристики датасета:

1. Типы данных;
2. Количество пропусков;
3. Количество уникальных значений;
4. Статистическая сводка;
5. Матрица корреляций;
6. Таблица рассеивания значений;
7. Выбросы.

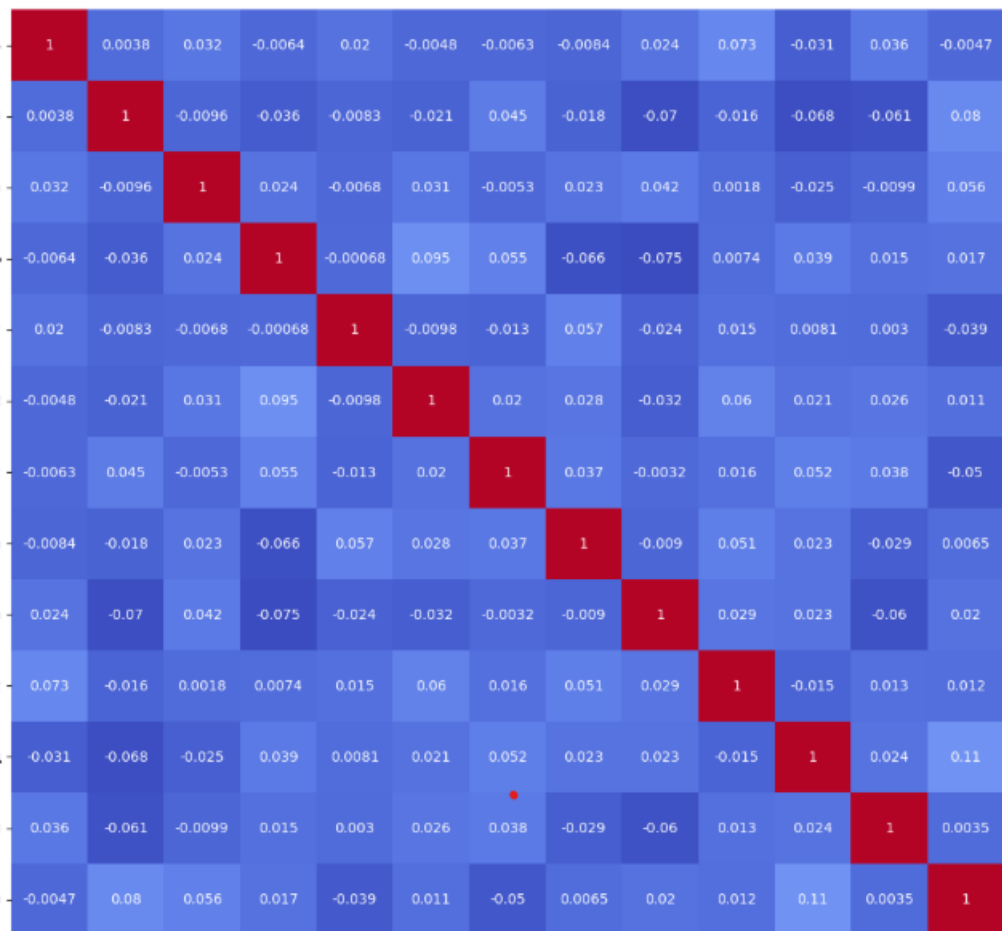
Сводная информация

	Тип данных	Количество_пропусков	Количество_уникальных_значений	Процент_пропусков
id_x	int64	0	1023	0.0
Соотношение матрица-наполнитель	float64	0	1014	0.0
Плотность, кг/м3	float64	0	1013	0.0
модуль упругости, ГПа	float64	0	1020	0.0
Количество отвердителя, м.%	float64	0	1005	0.0
Содержание эпоксидных групп,%_2	float64	0	1004	0.0
Температура вспышки, C_2	float64	0	1003	0.0
Поверхностная плотность, г/м2	float64	0	1004	0.0
Модуль упругости при растяжении, ГПа	float64	0	1004	0.0
Прочность при растяжении, МПа	float64	0	1004	0.0
Потребление смолы, г/м2	float64	0	1003	0.0
id_y	int64	0	1023	0.0
Угол нашивки, град	int64	0	2	0.0
Шаг нашивки	float64	0	989	0.0
Плотность нашивки	float64	0	988	0.0



Анализ данных

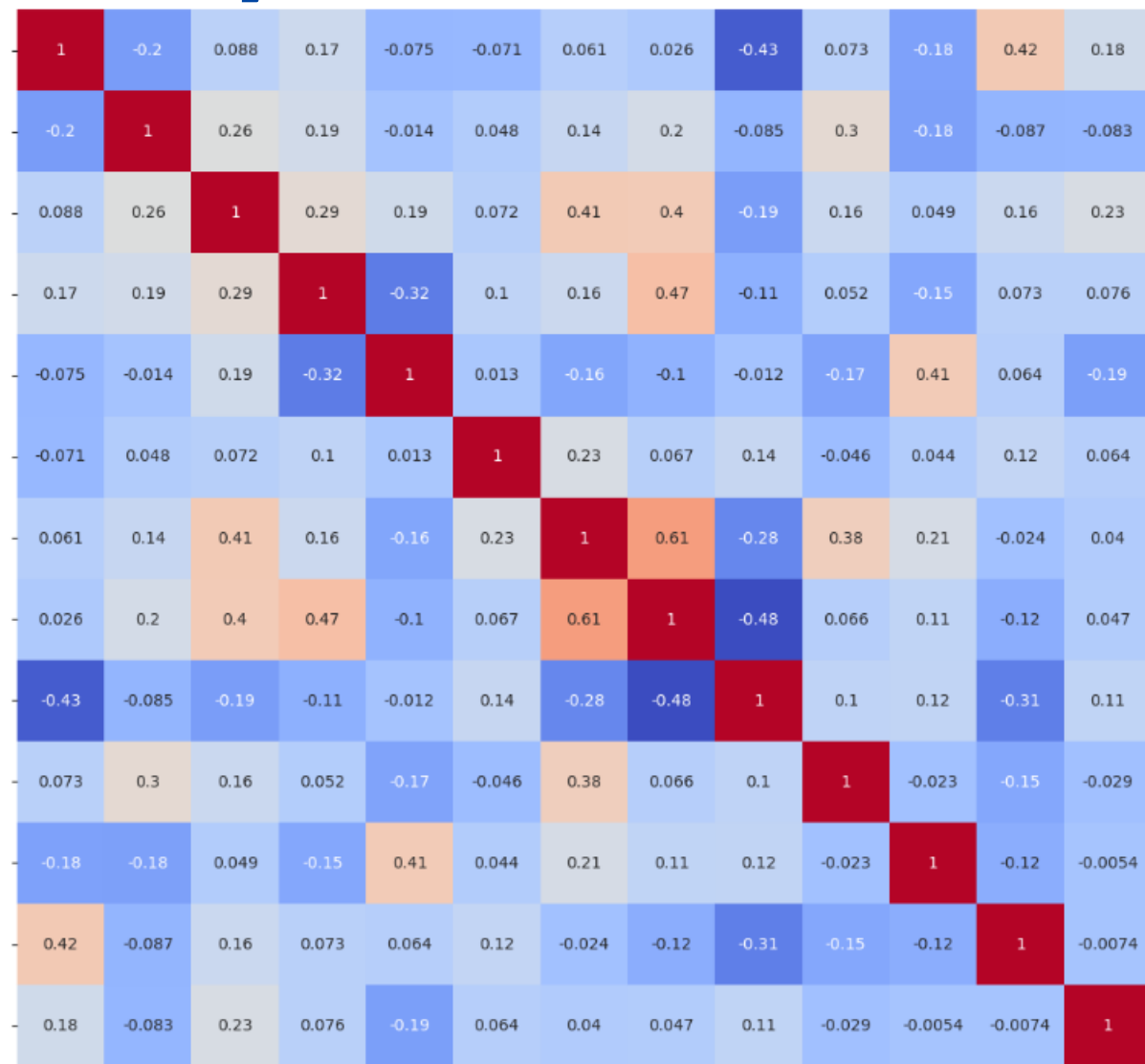
Визуализация корреляции и рассеивания данных





Анализ данных

Было взято за основу, синтетические данные – это нецелочисленные значения. Были отобраны только записи, для которых в столбце «Плотность нашивки» целое число. По полученным данным была составлена карта корреляции. Видно, что на таких данных зависимости уже более выражены. На полученном датасете далее будем проводить обучение.





Поиск гиперпараметров по сетке с перекрестной проверкой, y_1

1. LinearRegression
2. GammaRegressor
3. ARDRegression
4. AdaBoostRegressor
5. ExtraTreesRegressor
6. GradientBoostingRegressor
7. BaggingRegressor
8. CatBoostRegressor

	model	best_score	best_params
0	LinearRegression	-17.919547	{}
1	GammaRegressor	-14.156777	{}
2	ARDRegression	-6.943522	{'tol': 0.001}
4	BaggingRegressor	-5.173590	{'max_features': 0.5, 'max_samples': 1.0, 'n_e...
3	AdaBoostRegressor	-4.036066	{'loss': 'linear', 'n_estimators': 5}
5	ExtraTreesRegressor	-3.861599	{'max_depth': 10, 'min_samples_split': 5, 'n_e...
6	GradientBoostingRegressor	-3.782408	{'learning_rate': 0.5, 'loss': 'huber', 'max_d...
7	CatBoostRegressor	-3.059801	{'depth': 5, 'iterations': 500, 'learning_rate...



Результат обучения

«Модуль упругости при растяжении, ГПа»

Модели	Без нормализации			С нормализацией		
	MAE	MSE	r2	MAE	MSE	r2
1	2	3	4	5	6	7
1. LinearRegression	2,08	5,75	0,4	2,08	5,75	0,4
2. GammaRegressor	1,84	4,29	0,55	2,62	9	0,06
3. ARDRegression	2,6	8,3	0,13	1,9	4,28	0,55
4. AdaBoostRegressor	1,38	3,20	0,67	1,37	3,23	0,66
5. ExtraTreesRegressor	1,57	4,45	0,54	1,7	4,88	0,49
6. GradientBoostingRegressor	1,34	3,62	0,62	1,44	3,85	0,6
7. BaggingRegressor	1,74	4,93	0,48	1,63	5,92	0,38
8. CatBoostRegressor	2,01	6,98	0,27	2,01	6,98	0,27



Поиск гиперпараметров по сетке с перекрестной проверкой, y_2

1. LinearRegression
2. GammaRegressor
3. ARDRegression
4. AdaBoostRegressor
5. ExtraTreesRegressor
6. GradientBoostingRegressor
7. BaggingRegressor
8. CatBoostRegressor

	model	best_score	best_params
0	LinearRegression	-10.358839	{}
2	ARDRegression	-5.373563	{'tol': 0.01}
6	GradientBoostingRegressor	-3.616197	{'learning_rate': 0.01, 'loss': 'huber', 'max_...
1	GammaRegressor	-2.569186	{}
7	CatBoostRegressor	-2.321341	{'depth': 7, 'iterations': 100, 'learning_rate...
3	AdaBoostRegressor	-2.243952	{'loss': 'exponential', 'n_estimators': 30}
5	ExtraTreesRegressor	-1.972986	{'max_depth': 10, 'min_samples_split': 2, 'n_e...
4	BaggingRegressor	-1.668895	{'max_features': 0.5, 'max_samples': 1.0, 'n_e...



Результат обучения

«Прочность при растяжении, МПа»

Модели	Без нормализации			С нормализацией		
	MSE	MAE	r2	MSE	MAE	r2
1	2	3	4	5	6	7
1. LinearRegression	623,77	473638,04	-1,05	623,77	473638,04	-1,05
2. GammaRegressor	456,87	262844,88	-0,14	450,56	257126,75	-0,11
3. ARDRegression	449,03	276128,64	-0,19	460,21	278366,25	-0,2
4. AdaBoostRegressor	296,21	152585,1	0,34	330,68	159189,23	0,31
5. ExtraTreesRegressor	337,34	173246,8	0,25	376,45	184493,12	0,2
6. GradientBoostingRegressor	431,24	243393,79	-0,05	430,28	242669,69	-0,05
7. BaggingRegressor	345,91	180277,16	0,22	343,56	194889,73	0,18
8. CatBoostRegressor	346,87	173018,99	0,25	346,87	173018,99	0,25



Нейронная сеть

«Соотношение матрица-наполнитель»

Результаты:

R2: -0,48.

MSE: 0.53;

MAE: 0.57.

```
def baseline_model():
    model = Sequential()
    normalizer
    model.add(Dense(20, input_dim=10, activation='tanh', bias_initializer='he_normal', kernel_initializer='he_normal'))
    model.add(Dense(150, input_dim=20, activation='tanh', bias_initializer='he_normal', kernel_initializer='he_normal'))
    model.add(Dense(150, input_dim=150, activation='tanh', bias_initializer='he_normal', kernel_initializer='he_normal'))
    model.add(Dense(100, input_dim=150, activation='tanh', bias_initializer='he_normal', kernel_initializer='he_normal'))
    model.add(Dense(20, input_dim=100, activation='tanh', bias_initializer='he_normal', kernel_initializer='he_normal'))
    model.add(Dense(1, input_dim=20, activation='linear', bias_initializer='he_normal', kernel_initializer='he_normal'))

    adam = Adam()
    model.compile(loss='mean_squared_error', optimizer='adam')
    return model

model = baseline_model()
```



Приложение Flask

Приложение доступно по пути:

<https://ostroverkh-vkr-10.onrender.com/>

Рассчитывает прогноз по
переменным:

- «Модуль упругости при
растяжении, ГПа»;
- «Прочность при растяжении,
МПа».

Введите данные для расчета:

Плотность, кг/м³:

Модуль упругости, ГПа:

Количество отвердителя, м. %:

Содержание эпоксидных групп, %:

Температура вспышки, °C:

Поверхностная плотность, г/м²:

Потребление смолы, г/м²:

Угол нашивки, градусы:

Шаг нашивки:

Плотность нашивки:



ЦЕНТР
ДОПОЛНИТЕЛЬНОГО
ОБРАЗОВАНИЯ
МГТУ им. Н.Э. Баумана



do.bmstu.ru