

Выпускная квалификационная работа

по курсу

«Data Science»

по теме:

Прогнозирование конечных свойств новых материалов
(композиционных материалов)

Слушатель: Тулупов Роман Иванович

Этапы работы

- Загрузка и объединение датасетов.
- Разведочный анализ данных.
- Обработка дубликатов и выбросов.
- Гистограммы распределения каждой из переменной, диаграммы ящик с усами, попарные графики рассеяния точек.
- Предобработка данных.
- Обучение нескольких моделей для прогноза модуля упругости при растяжении и прочности при растяжении.
- Написание нейронной сети, которая будет рекомендовать соотношение матрица-наполнитель.
- Разработка приложения Flask.

Загрузка и объединение датасетов

```
X_bp = pd.read_excel(  
    'https://github.com/ritulupov/BKP/blob/main/data/X_bp.xlsx?raw=true',  
    index_col=0, engine='openpyxl')  
  
X_nup = pd.read_excel(  
    'https://github.com/ritulupov/BKP/blob/main/data/X_nup.xlsx?raw=true',  
    index_col=0, engine='openpyxl')
```

Файлы были заранее размещены в моем профиле на GitHub.

Имеется два файла X_bp.xlsx - с информацией о входных и выходных параметрах, состоящий из 1023 строк и 11 колонок и X_nup.xlsx - с информацией о нашивке, состоящий из 1040 строк и 4 колонок. Оба файла имеют колонки с индексами.

```
full_df = X_bp.join(X_nup)
```

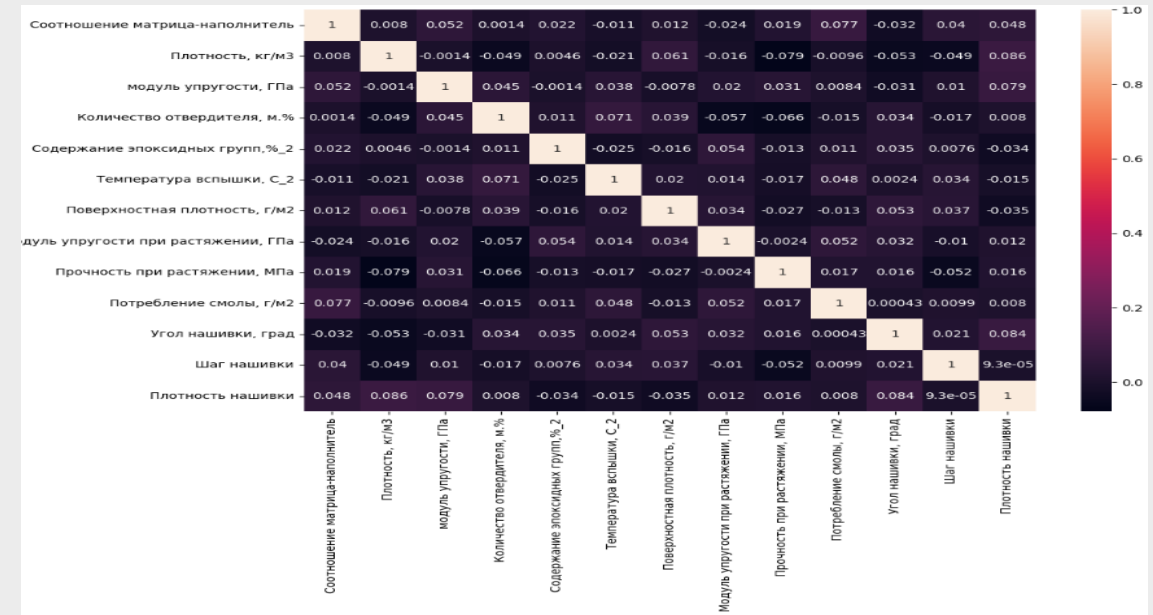
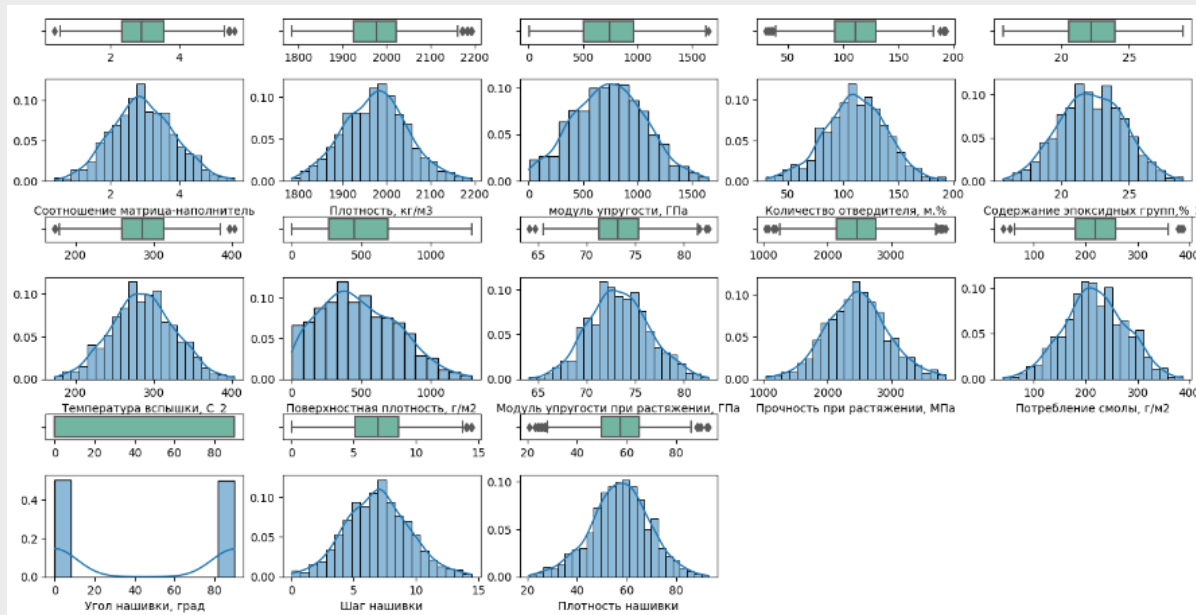
Файлы объединяются по индексу в датасет «full_df» тип объединения INNER. Объем полученной выборки составляет 1023 строки и 13 признаков, из них 8 являются входными переменными и 5 выходными переменными.

В данном задании колонки «Соотношение матрица-наполнитель», «Модуль упругости при растяжении, ГПа», «Прочность при растяжении, МПа» являются целевыми.

Разведочный анализ данных

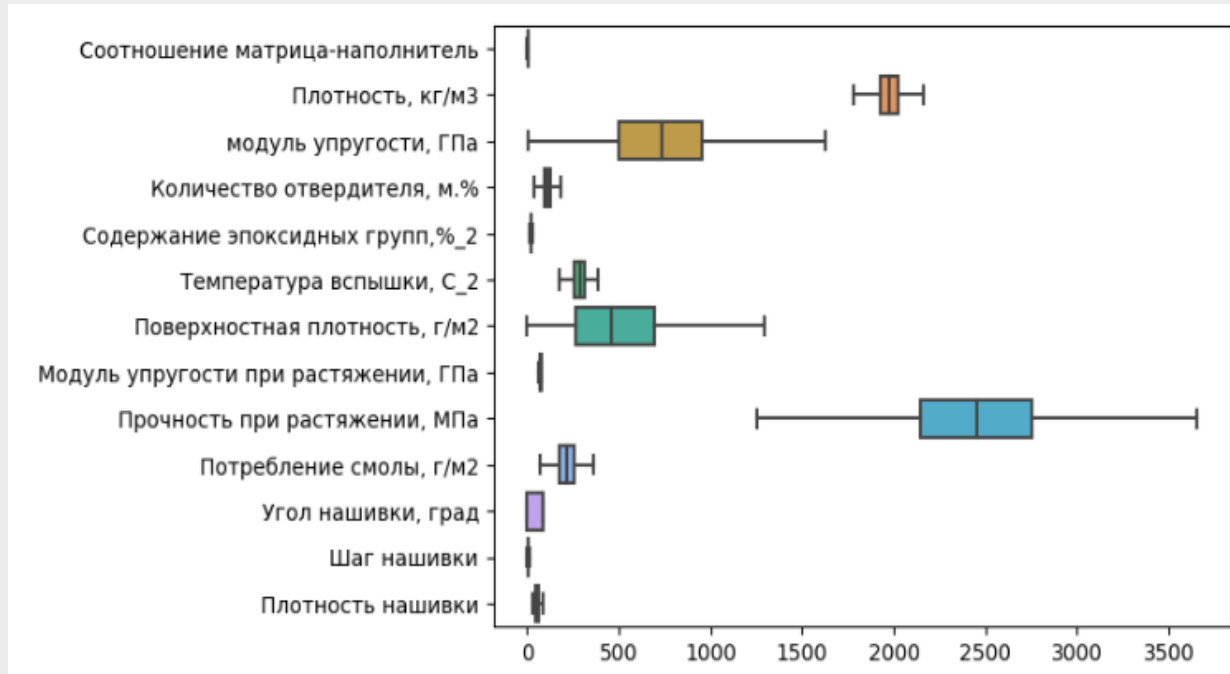
Выясняется, что в данных нет дубликатов, нет пропусков.

Построение гистограмм распределения каждой из переменной, диаграмм ящик с усами, тепловая карта корреляций



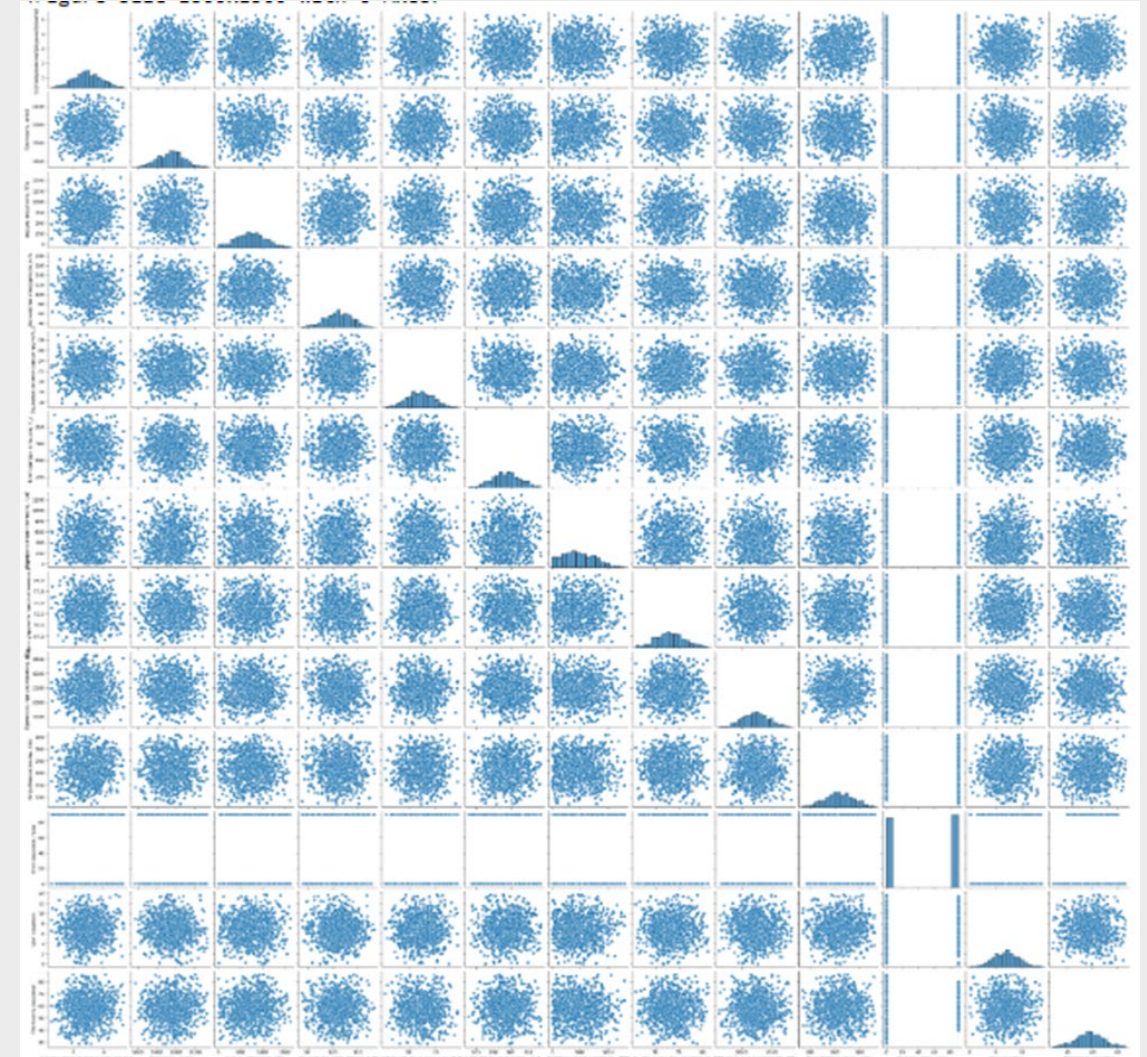
Максимальная корреляция между плотностью нашивки и плотностью составляет 0.086, корреляция между всеми параметрами очень близка к 0, значит нет корреляционной связи между переменными. Есть выбросы.

Гистограммы распределения каждой переменной, диаграммы ящик с усами, попарные графики рассеяния точек



Выбросы удалены. Данные имеют разный масштаб.

На попарных графиках распределения не видно корреляции между признаками.

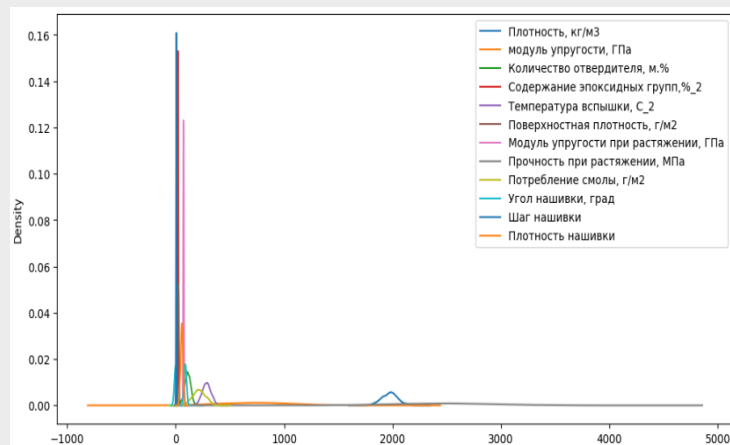


Предобработка данных

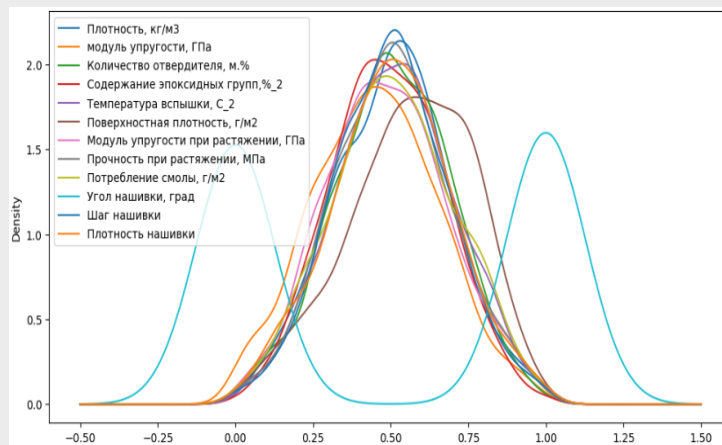
	mean	50%
Соотношение матрица-наполнитель	2.927964	2.907832
Плотность, кг/м3	1974.118744	1977.321002
модуль упругости, ГПа	736.119982	736.178435
Количество отвердителя, м.%	111.136066	111.162090
Содержание эпоксидных групп,%_2	22.200570	22.177681
Температура вспышки, C_2	286.181128	286.220763
Поверхностная плотность, г/м2	482.429070	457.732246
Модуль упругости при растяжении, ГПа	73.303464	73.247594
Прочность при растяжении, МПа	2461.491315	2455.974462
Потребление смолы, г/м2	218.048059	218.697660
Угол нашивки, град	45.976139	90.000000
Шаг нашивки	6.931939	6.972862
Плотность нашивки	57.562887	57.584225

Среднее и медианное значения по всем колонкам приблизительно равны за исключением поверхностной плотности. Здесь отклонение составляет 5.2%. Ранее на гистограмме можно было видеть на этом признаке смещение распределения влево. Для исправления можно извлечь корень.

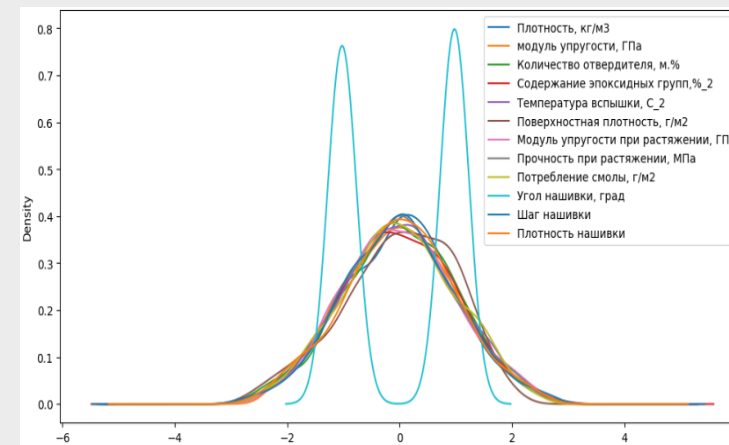
```
sqrt_df['Поверхностная плотность, г/м2'] = np.sqrt(sqrt_df['Поверхностная плотность, г/м2'])
```



None



MinMaxScaler



StandardScaler

Обучение нескольких моделей для прогноза модуля упругости при растяжении и прочности при растяжении

Для дальнейшей работы были выбраны пять моделей. Для каждой из двух целевых переменных были проделаны следующие шаги:

- Предварительно произведен подбор нормализатора и выборки (с выбросами и без);
- Результаты отсортированы и сведены в таблицу;
- Прежде чем проводить подбор параметров по сетке, был проведен рандомный поиск параметров и визуализация результатов для выявления тенденций;
- Результаты всех экспериментов были занесены в таблицу, после чего модель с лучшим результатом была сохранена на диске для дальнейшего использования в приложении.

В итоге на диске сохранены две обученные модели с лучшими характеристиками.

Написание нейронной сети, которая будет рекомендовать соотношение матрица-наполнитель

Для конструирования нейронной сети и оптимизации гиперпараметров, воспользовался инструментом *kerastuner*.

```
model.summary()
```

Model: "sequential"

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 576)	7488
dense_1 (Dense)	(None, 480)	276960
dense_2 (Dense)	(None, 1)	481
Total params: 284,929		
Trainable params: 284,929		
Non-trainable params: 0		

```
model.get_compile_config()
```

```
{'optimizer': 'adam',  
 'loss': 'mean_squared_error',  
 'metrics': ['mean_squared_error'],  
 'loss_weights': None,  
 'weighted_metrics': None,  
 'run_eagerly': None,  
 'steps_per_execution': None,  
 'jit_compile': None}
```

Итогом работы *kerastuner* стал выбор модели с характеристиками отображенными на картинке слева.

Модель сохранена на диске и может, в дальнейшем, быть использована в приложении или до обучена.

Разработка приложения Flask

Для прогнозирования модуля упругости при растяжении и прочности при растяжении разработано Flask приложение, запустить которое можно на сервере с поддержкой Flask или локальном компьютере, установив предварительно зависимости. После запуска интерфейс приложения доступен в браузере по адресу `hostname:5000/`

Чтобы получить прогноз, необходимо заполнить все поля формы и нажать кнопку «рассчитать».

Результат отобразится под кнопкой.

Прогноз

модуля упругости при растяжении, прочности при растяжении

Соотношение матрица-наполнитель	<input type="text" value="2.334566013"/>
Плотность, кг/м3	<input type="text" value="2020.97"/>
модуль упругости, ГПа	<input type="text" value="1013.514"/>
Количество отвердителя, м.%	<input type="text" value="109.5051"/>
Содержание эпоксидных групп, %_2	<input type="text" value="24.3309"/>
Температура вспышки, С_2	<input type="text" value="266.2586"/>
Поверхностная плотность, г/м2	<input type="text" value="719.87"/>
Потребление смолы, г/м2	<input type="text" value="264.8356"/>
Угол нашивки, град	<input type="text" value="0"/>
Шаг нашивки	<input type="text" value="7.3636"/>
Плотность нашивки	<input type="text" value="69.17234"/>

Ожидаемый модуль упругости при растяжении, ГПа: 73.938

Ожидаемая прочность при растяжении, МПа: 2472.436