

The slide features a white background with decorative blue triangles in the corners. A large blue triangle is in the top right corner. In the bottom left corner, there are two overlapping triangles: a darker blue one on top and a lighter blue one below it.

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

по курсу
«Data Science»

Слушатель: Долматова Анастасия Александровна

Постановка задачи:

Цель решения задачи: прогнозировать характеристики композиционного материала на основе имеющихся данных.

Входные данные:

Общее описание свойств композиционного материала

- два датасета, которые содержат данные о количественных характеристиках различных свойств и составляющих композитного материала. Всего 13 характеристик.

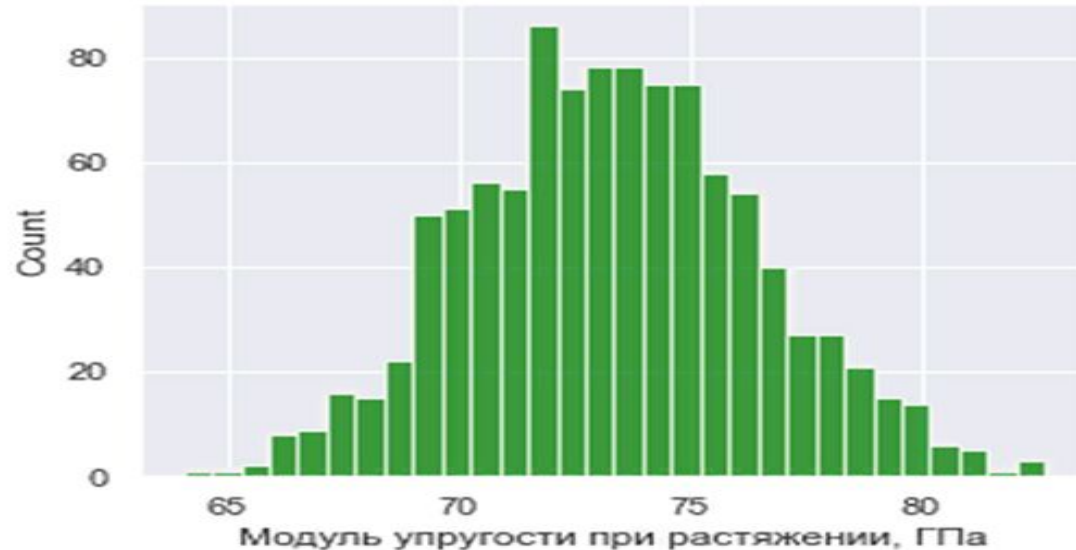
Постановка задач для решения с помощью методов машинного обучения:

- решение задачи регрессии для прогнозирования двух из 13 представленных характеристик
- создание нейронной сети для прогнозирования показателя «Соотношение матрица-наполнитель»

2 Этап. Разведочный анализ данных

Описательная статистика:

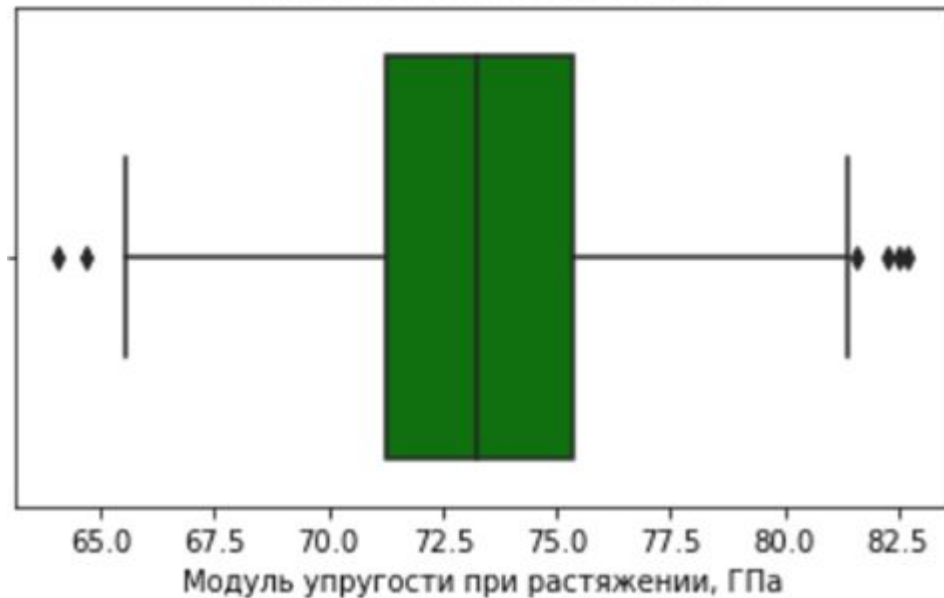
- Количество уникальных значений в каждом столбце. Выявлена одна дискретная величина
- Количество пропусков. Не обнаружено.
- графики распределения переменных.



2 Этап. Разведочный анализ данных

Описательная статистика:

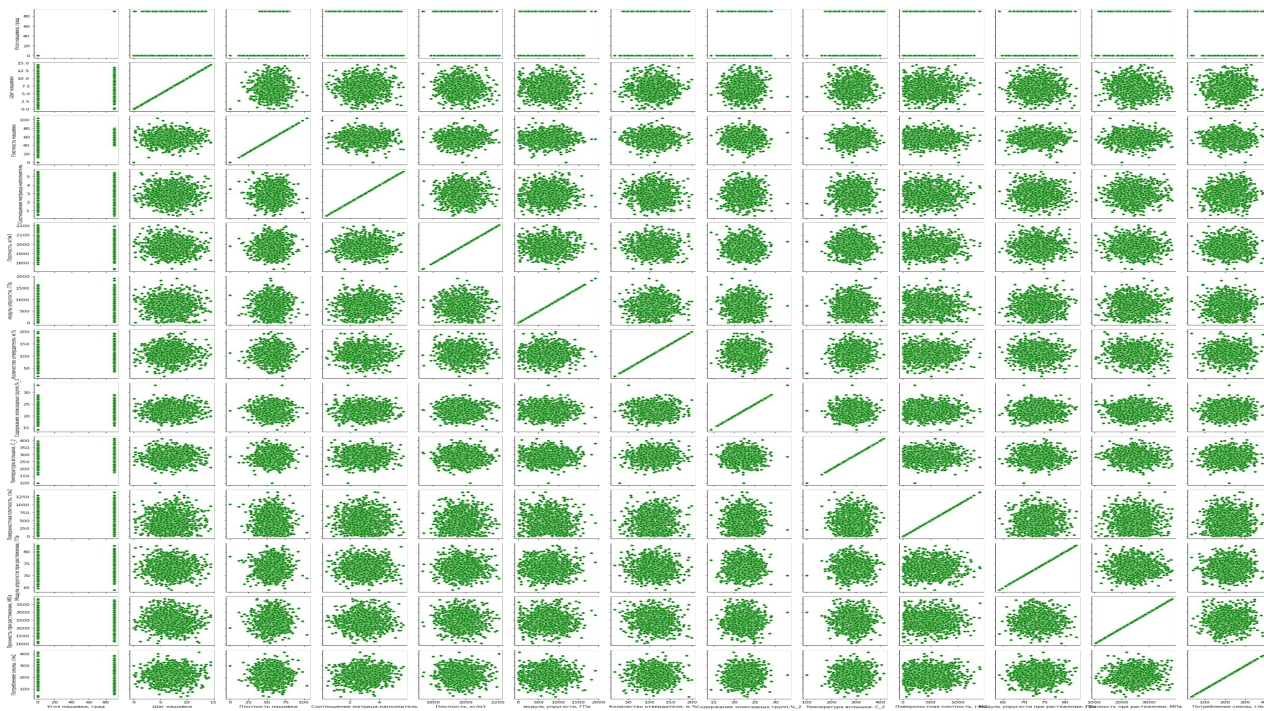
- Графики "ящики с усами"



2 Этап. Разведочный анализ данных

Описательная статистика:

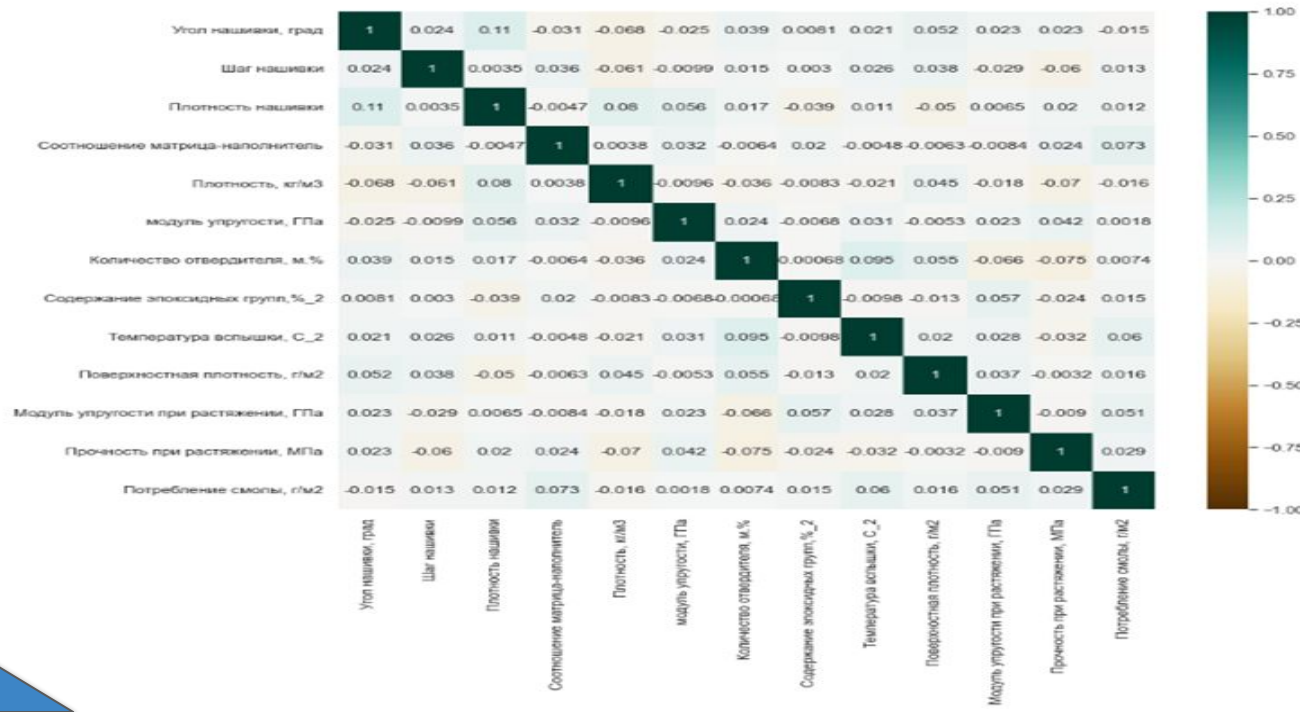
- Парные графики рассеивания



2 Этап. Разведочный анализ данных

Описательная статистика:

- Коэффициент корреляции
- Тепловая карта коэффициентов корреляции.



3. Этап. Предобработка данных

- Расчет количества выбросов и удаление выбросов.
- Для поиска выбросов использован метод 3-х сигм. Количество выбросов 24.
- Нормализация данных методом MinMaxScaler.
- Проверка данных при помощи показателей описательной статистики и метода describe().

	Угол нашивки, град	Шаг нашивки	Плотность нашивки	Соотношение матрица- наполнитель	Плотность, кг/м3	модуль упругости, ГПа
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	0.496000	0.477193	0.507027	0.489568	0.467648	0.447024
std	0.500234	0.177586	0.163634	0.174687	0.178696	0.198876
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.000000	0.351886	0.405037	0.370964	0.340831	0.302576
50%	0.000000	0.477999	0.509730	0.484284	0.472347	0.448525
75%	1.000000	0.593714	0.612766	0.608289	0.579727	0.582408
max	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000

4 Этап. Решение задачи регрессии

Для прогноза модуля упругости при растяжении и прочности при растяжении были использованы и настроены гиперпараметры для следующих моделей:

к – ближайших соседей

- Количество соседей:
`'n_neighbors': np.arange(1, 25)`

Случайный лес

- Количество деревьев:
`'n_estimators': [60, 80, 100]`
- Поиск функций:
`'max_features': ['auto', 'sqrt', 'log2']`
- Глубина дерева:
`'max_depth' : [3,4,5,6]`

4 Этап. Решение задачи регрессии

Для прогноза модуля упругости при растяжении и прочности при растяжении были использованы и настроены гиперпараметры для следующих моделей:

Линейная регрессия

Градиентный бустинг

- Количество деревьев:
'n_estimators': [60, 80, 100]
- Поиск функций:
'max_features': ['auto', 'sqrt', 'log2']
- Глубина дерева:
'max_depth' : [3,4,5,6]

5 Этап. Оценка качества моделей для задачи регрессии

наименование модели	mean_absolute_error	r2_score
Случайный лес_МУ	0.144007	-241.823329
К соседей_МУ	0.146184	-23.223745
Градиентный бустинг_МУ	0.145732	-25.175385
Линейная регрессия_МУ	0.145125	-43.538655
Случайный лес_ПР	0.145951	-176.576341
К соседей_ПР	0.146513	-26.181760
Градиентный бустинг_ПР	0.146387	-20.367530
Линейная регрессия_ПР	0.147455	-34.854171

Этап 6. Построение нейронной сети

keras.Sequential со следующими параметрами:

- входной слой нормализации 12 признаков;
- выходной слой для 1 признака;
- скрытых слоев: 2;
- нейронов в первом скрытом слое : 20;
- нейронов во втором скрытом слое : 10;
- активационная функция скрытых слоев: sigmoid;
- оптимизатор: RMSprop;
- loss-функция: MeanAbsoluteError.

Model: "sequential"

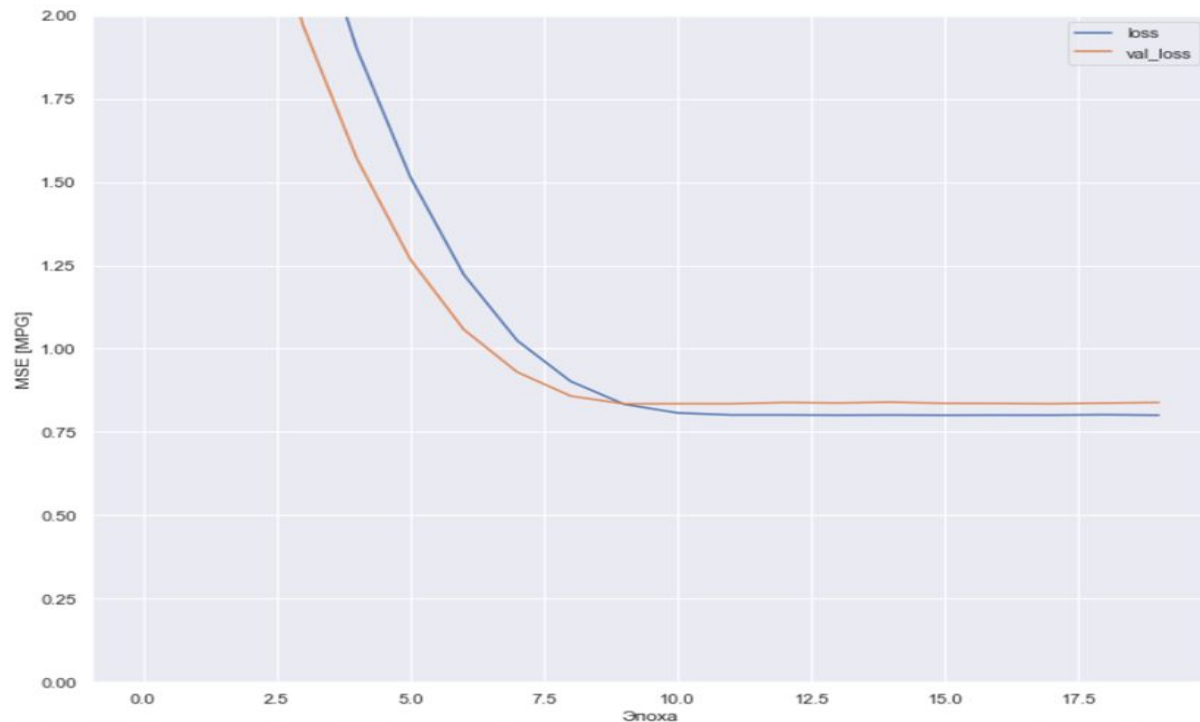
Layer (type)	Output Shape	Param #
=====		
normalization (Normalizatio n)	(None, 12)	25
dense (Dense)	(None, 200)	2600
dropout (Dropout)	(None, 200)	0
dense_1 (Dense)	(None, 20)	4020
dense_2 (Dense)	(None, 10)	210
dense_3 (Dense)	(None, 1)	11

=====

Total params: 6,866
Trainable params: 6,841
Non-trainable params: 25

Этап 6. Построение нейронной сети

Результаты обучения модели характеризуются изменением среднеквадратической ошибки



Этап 8. Разработка Flask приложения

- Разработано веб-приложение для модели «Соотношение матрица-наполнитель». Приложение разработано в среде разработки PyCharm.

Плотность, кг/м³

Модуль упругости, ГПа²

Количество отвердителя, м.%

Содержание эпоксидных групп, %₂

Температура вспышки, С₂

Поверхностная плотность, г/м²

Модуль упругости при растяжении, ГПа

Прочность при растяжении, МПа

Потребление смолы, г/м²

Угол нашивки, град

Шаг нашивки

Плотность нашивки



Спасибо за внимание!