

# ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

ПО КУРСУ

«DATA SCIENCE»

ПО ТЕМЕ:

Прогнозирование конечных свойств новых материалов  
(композиционных материалов)

Слушатель: Бережанская И.Ю.

# Характеристики анализируемого набора данных

- Два файла X\_br и X\_nir представлены в таблицах Excel и объединены в один
- Объем и характеристики датасета: 1023 строки и 13 колонок

	Соотношение матрица-наполнитель	Плотность, кг/м3	модуль упругости, ГПа	Количество отвердителя, м.%	Содержание эпоксидных групп,%_2	Температура вспышки, С_2	Поверхностная плотность, г/м2	Модуль упругости при растяжении, ГПа	Прочность при растяжении, МПа	Потребление смолы, г/м2	Угол нашивки, град	Шаг нашивки	Плотность нашивки
0	1.857143	2030.000000	738.736842	30.000000	22.267857	100.000000	210.000000	70.000000	3000.000000	220.000000	0.0	4.000000	57.000000
1	1.857143	2030.000000	738.736842	50.000000	23.750000	284.615385	210.000000	70.000000	3000.000000	220.000000	0.0	4.000000	60.000000
2	1.857143	2030.000000	738.736842	49.900000	33.000000	284.615385	210.000000	70.000000	3000.000000	220.000000	0.0	4.000000	70.000000
3	1.857143	2030.000000	738.736842	129.000000	21.250000	300.000000	210.000000	70.000000	3000.000000	220.000000	0.0	5.000000	47.000000
4	2.771331	2030.000000	753.000000	111.860000	22.267857	284.615385	210.000000	70.000000	3000.000000	220.000000	0.0	5.000000	57.000000
...	...	...	...	...	...	...	...	...	...	...	...	...	...
1018	2.271346	1952.087902	912.855545	86.992183	20.123249	324.774576	209.198700	73.090961	2387.292495	125.007669	90.0	9.076380	47.019770
1019	3.444022	2050.089171	444.732634	145.981978	19.599769	254.215401	350.660830	72.920827	2360.392784	117.730099	90.0	10.565614	53.750790
1020	3.280604	1972.372865	416.836524	110.533477	23.957502	248.423047	740.142791	74.734344	2662.906040	236.606764	90.0	4.161154	67.629684
1021	3.705351	2066.799773	741.475517	141.397963	19.246945	275.779840	641.468152	74.042708	2071.715856	197.126067	90.0	6.313201	58.261074
1022	3.808020	1890.413468	417.316232	129.183416	27.474763	300.952708	758.747882	74.309704	2856.328932	194.754342	90.0	6.078902	77.434468

1023 rows × 13 columns



# Этапы обработки данных

- Анализ признаков с целью выявления общих зависимостей, а также проведена визуализация данных
- Удаление дубликатов и выбросов
- Разработка признаков и трансформация существующих
- Нахождение коррелирующих признаков и отбор
- Для прогнозирования «Модуль упругости при растяжении, ГПА» и «Прочность при растяжении» построены регрессионные модели
- Для рекомендации «Соотношение матрица-наполнитель» нейронная сеть



# Этапы обработки данных

Гистограммы распределения

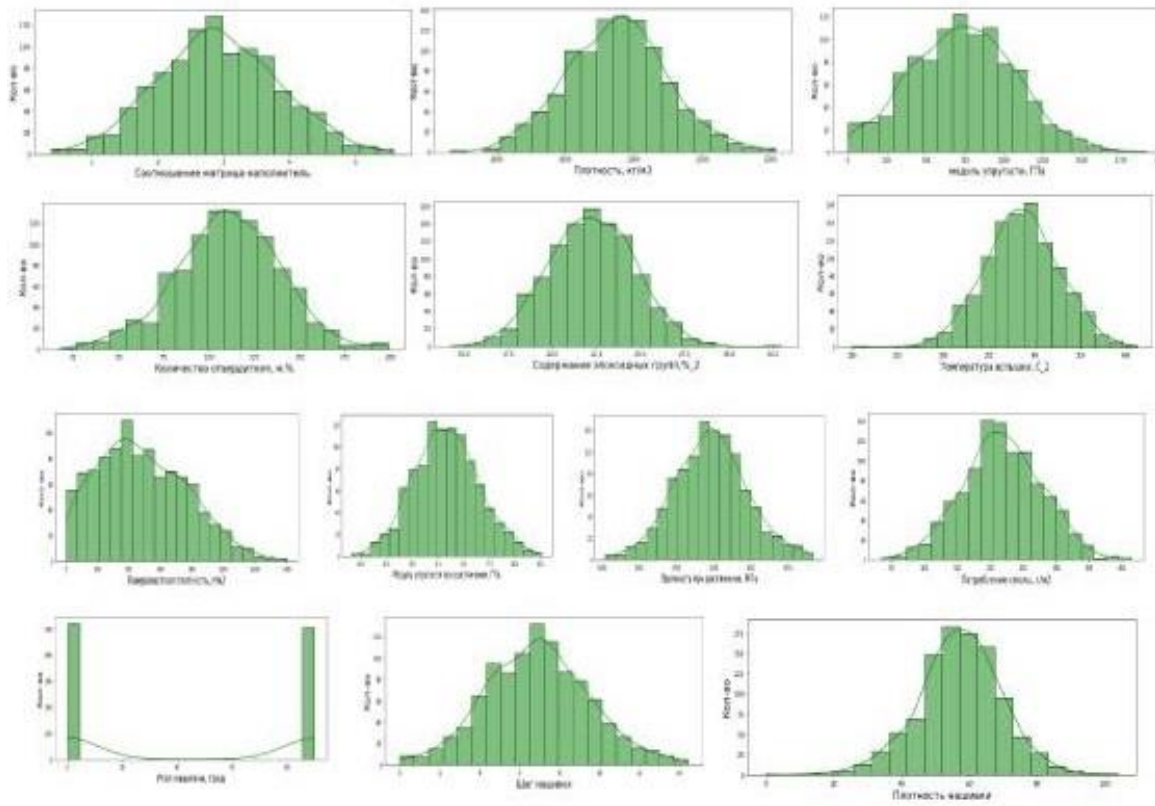
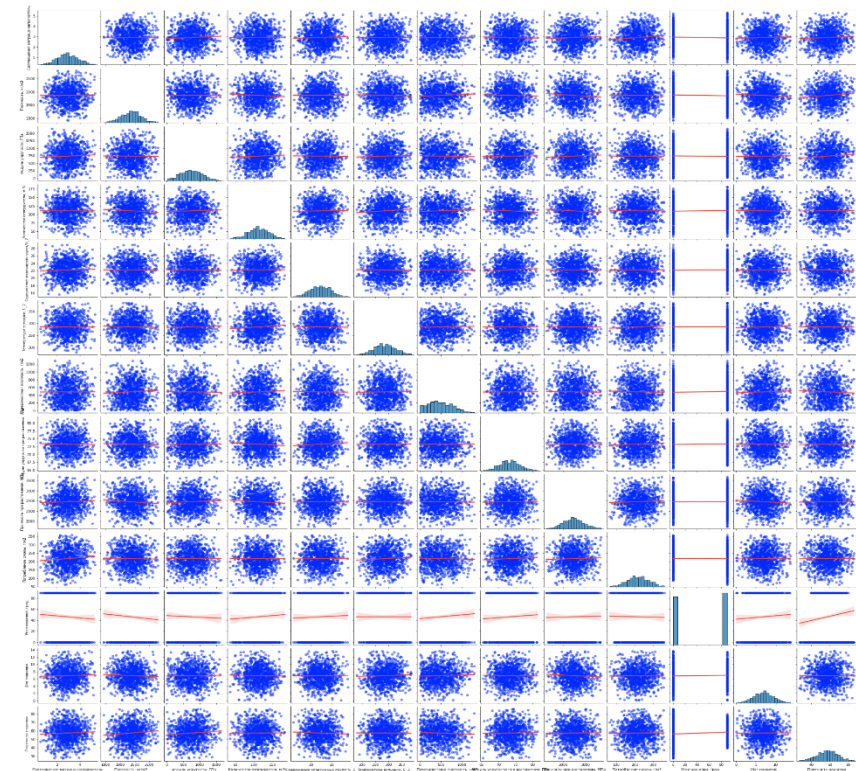
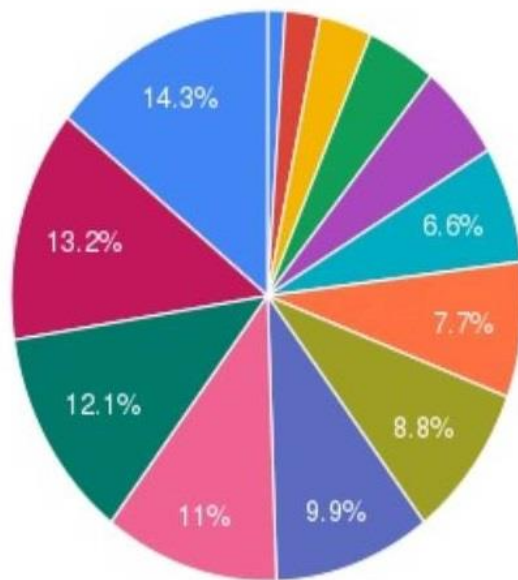


Диаграмма рассеивания



# Разработка и трансформация признаков



- Соотношение матрица-н...
- Плотность ,кг/м3
- Модуль упругости гпа
- Количество отвердителя %
- Содержание эпоксидных...
- Температура вспышки
- Поверхностная плотность
- Модуль упругости при ра...
- Прочность при растяжении
- Потребление смолы
- Угол нашивки
- Шаг нашивки
- Плотность нашивки

Шаг 1	Шаг 2
Новая переменная «Тотал»- суммарный вес всех параметров	Трансформация переменных. Отношение их к общему весу (доля)



# Этапы обработки данных

Корреляционная карта датасета после трансформации

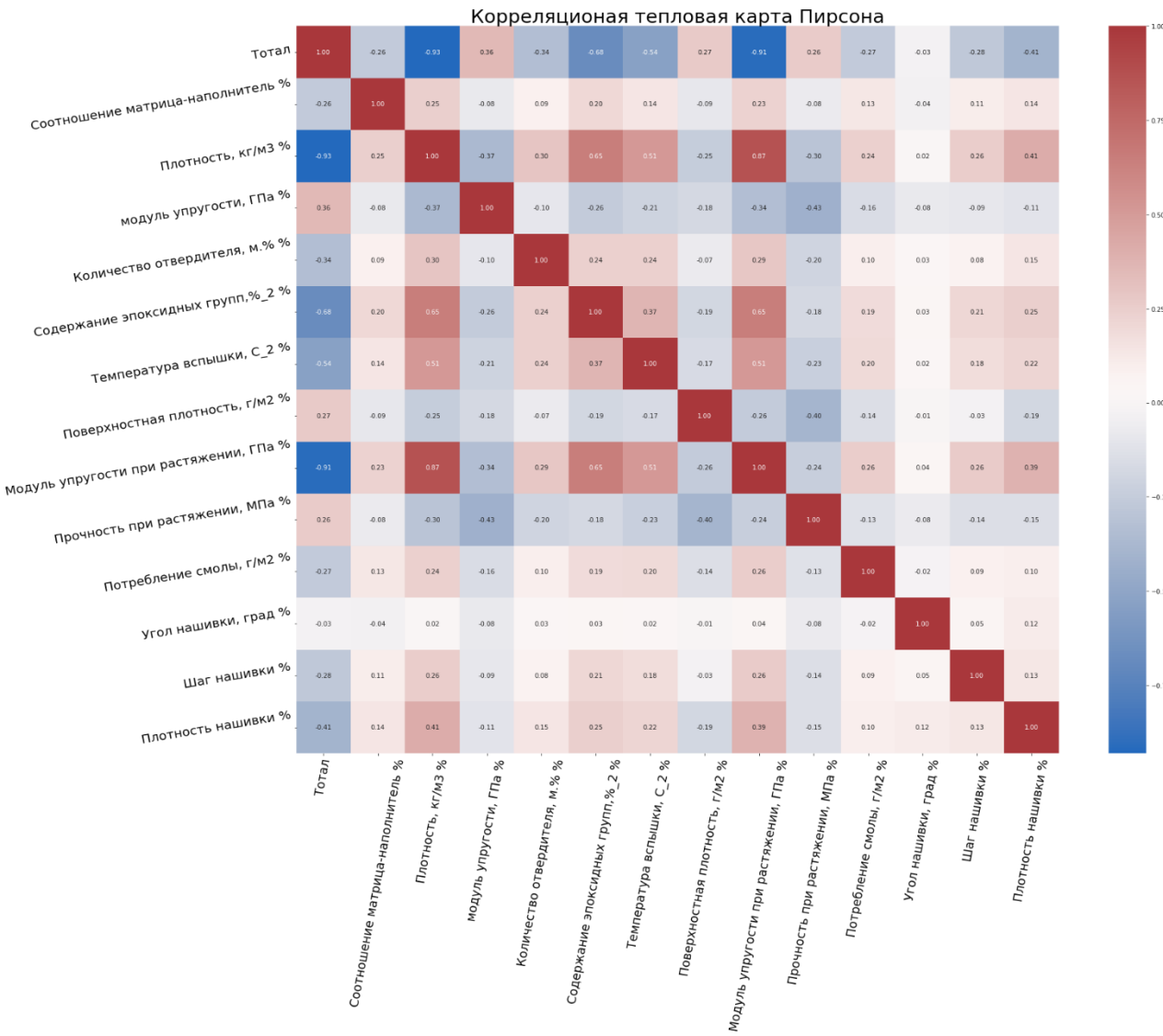
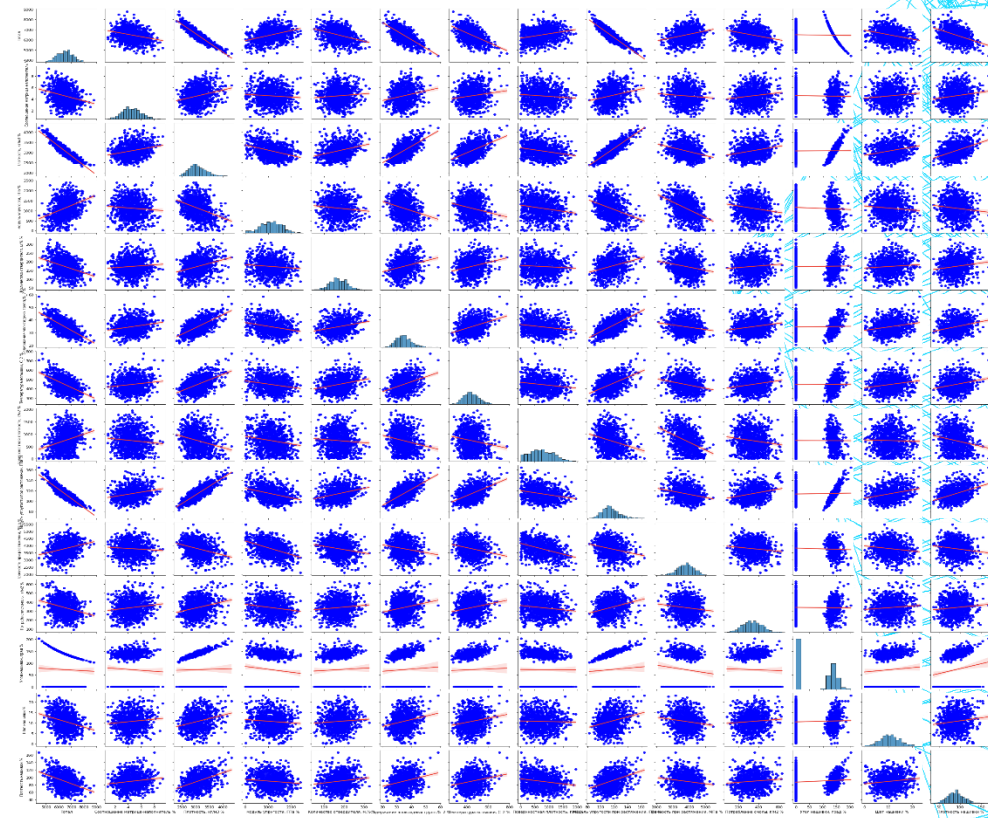
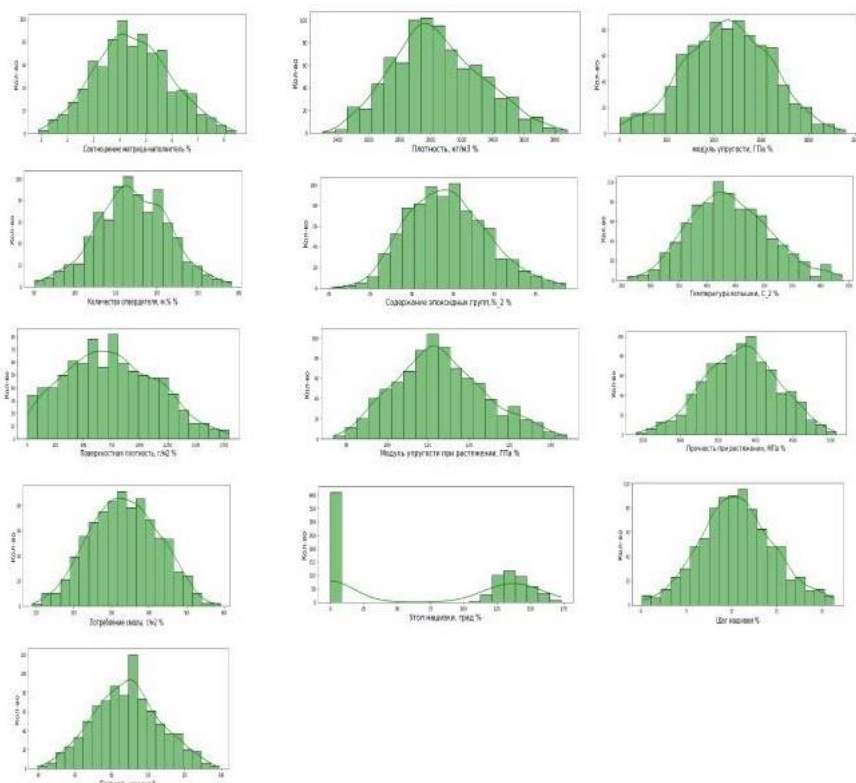


Диаграмма рассеивания после трансформации

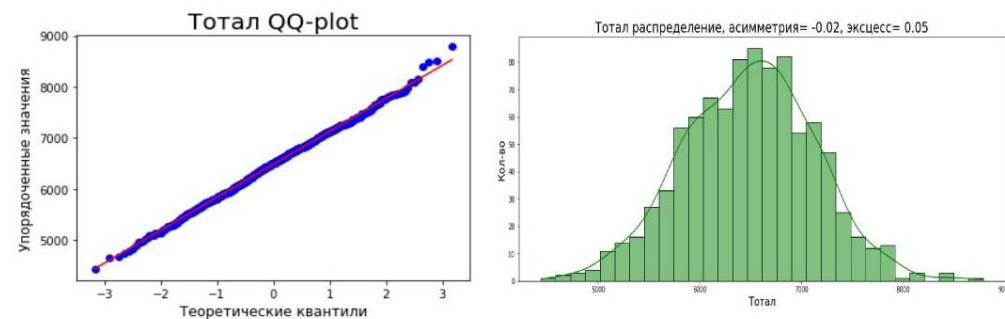


# Этапы обработки данных

Гистограммы распределения



Переменная «Тотал»





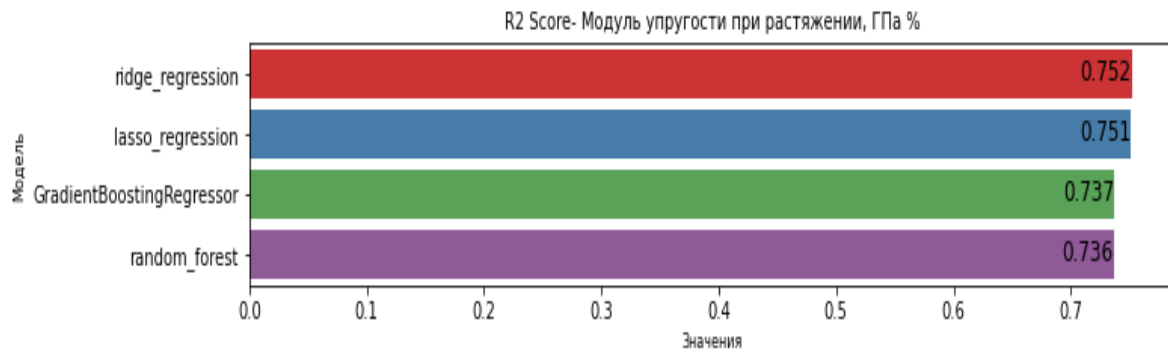
# Этапы разработки и обучения моделей

- Порядок разработки моделей для каждого параметра и для каждого выбранного метода соответствуют следующие этапы
- Разделение на обучающую тестовую выборки (соотношение 70% на 30%)
- Задание сетки гиперпараметров для оптимизации моделей
- Подстановка гиперпараметров в модель и обучение на тренировочных данных



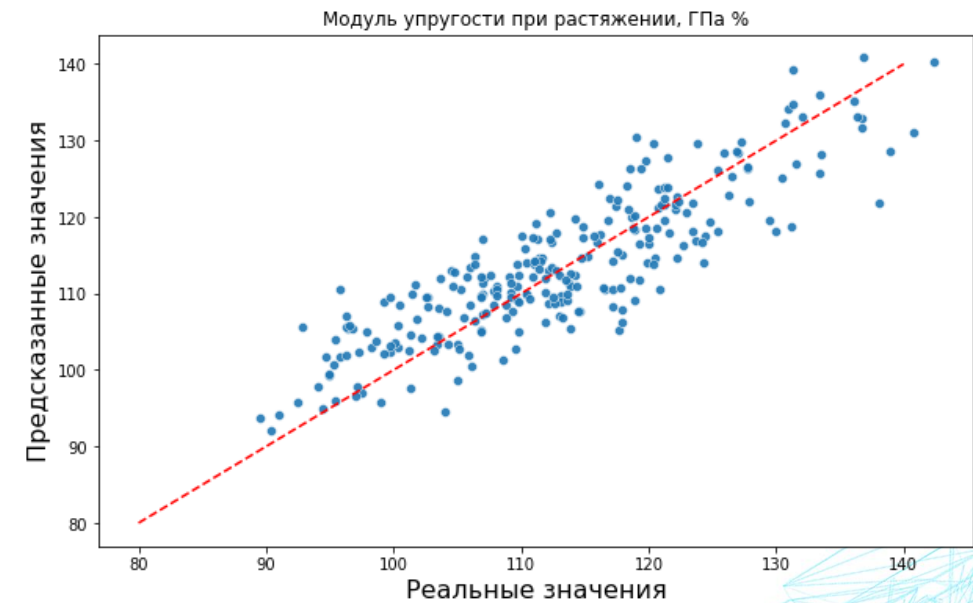
# Этапы разработки и обучения моделей

## ТЕСТИРОВАНИЕ МОДЕЛЕЙ



	Актуальные	Предсказанные
98	95.303293	100.808061
548	114.530026	107.744088
263	112.149924	112.991020
280	114.196785	119.764587
956	95.772454	101.665308
...	...	...
187	130.400799	125.164789
123	127.708729	126.613041
221	119.303526	116.467945
220	95.746823	110.590473
55	118.382868	121.056164

257 rows × 2 columns



Результаты моделей параметра «модуль упругости при растяжении, гпа %»

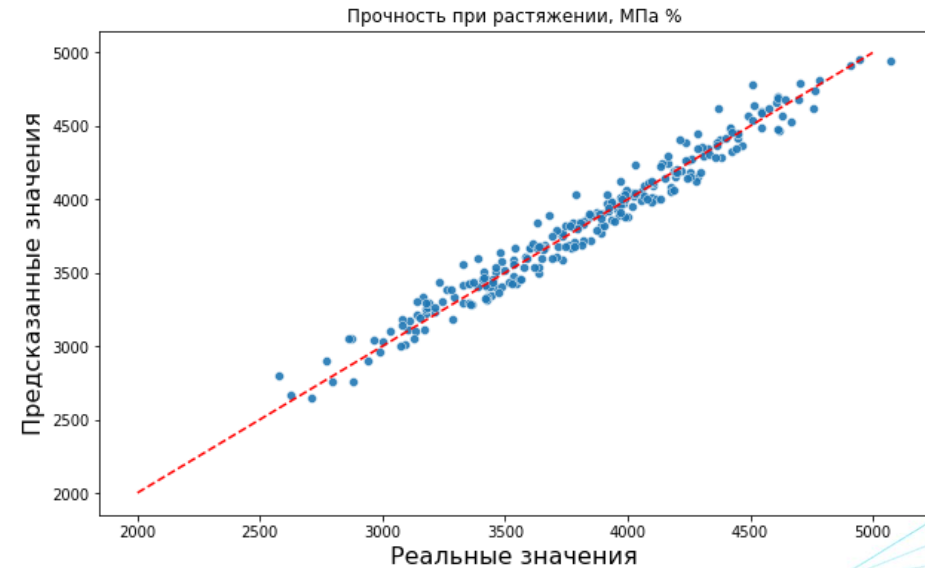
# Этапы разработки и обучения моделей

## ТЕСТИРОВАНИЕ МОДЕЛЕЙ



	Актуальные	Предсказанные
838	3838.330001	3868.416638
854	3721.639373	3772.856362
326	4762.439787	4746.201171
571	3663.306792	3685.718521
561	3499.185738	3515.113375
...	...	...
890	4365.196782	4369.705730
566	3388.291186	3400.893409
359	3126.192922	3048.276888
761	3373.007847	3436.692570
789	3633.791175	3684.651951

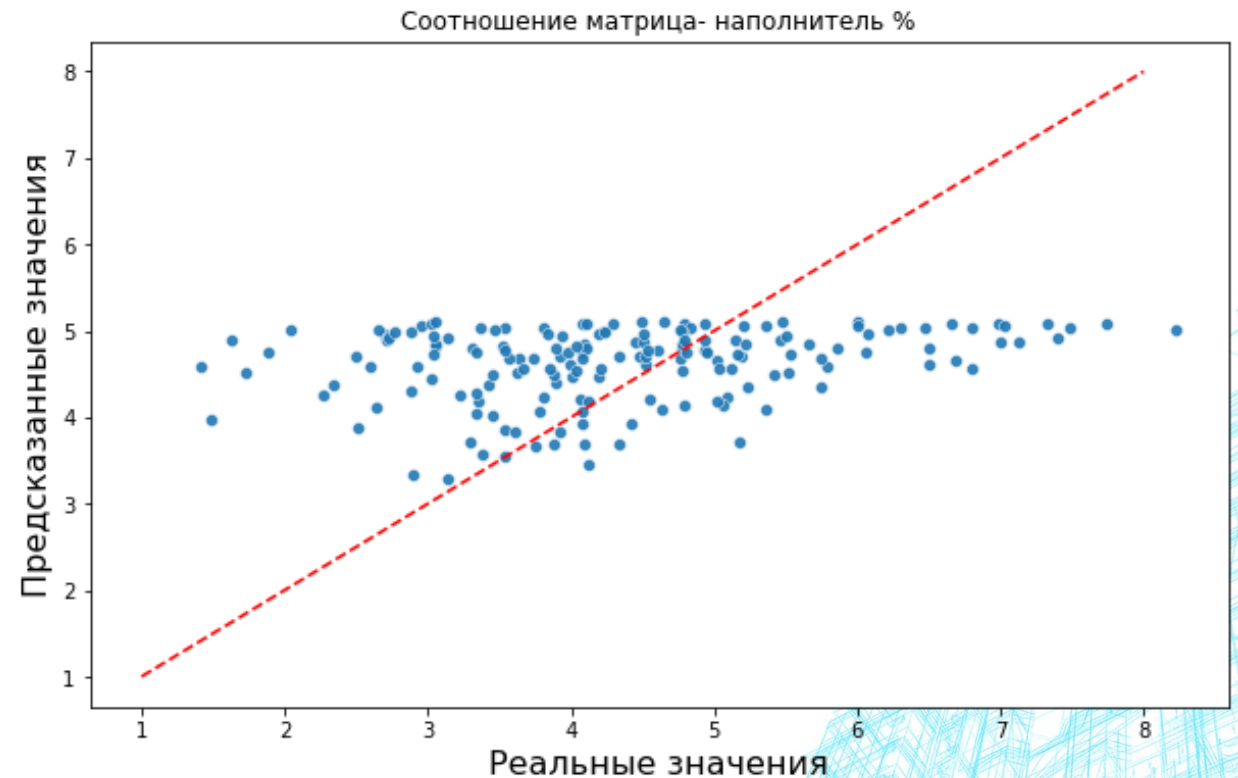
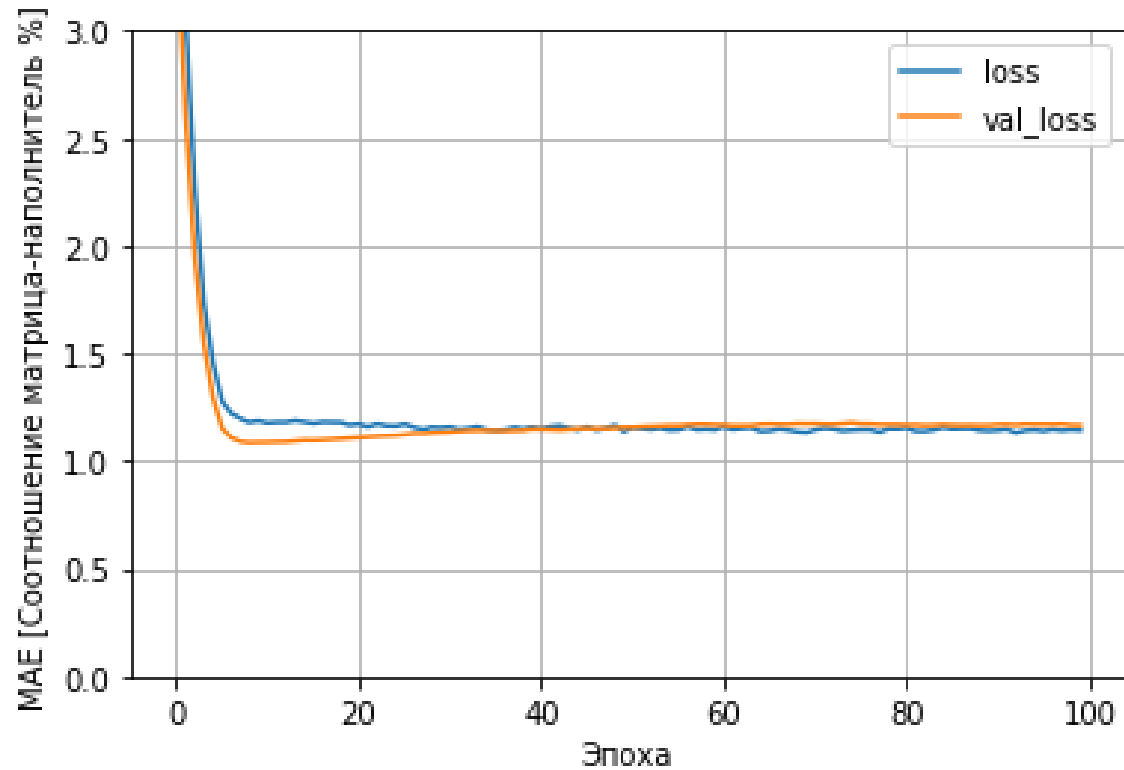
257 rows × 2 columns



Результаты моделей прогноза параметра «Прочность при растяжении, МПа %»

# Этапы разработки и обучения моделей

## ТЕСТИРОВАНИЕ МОДЕЛЕЙ

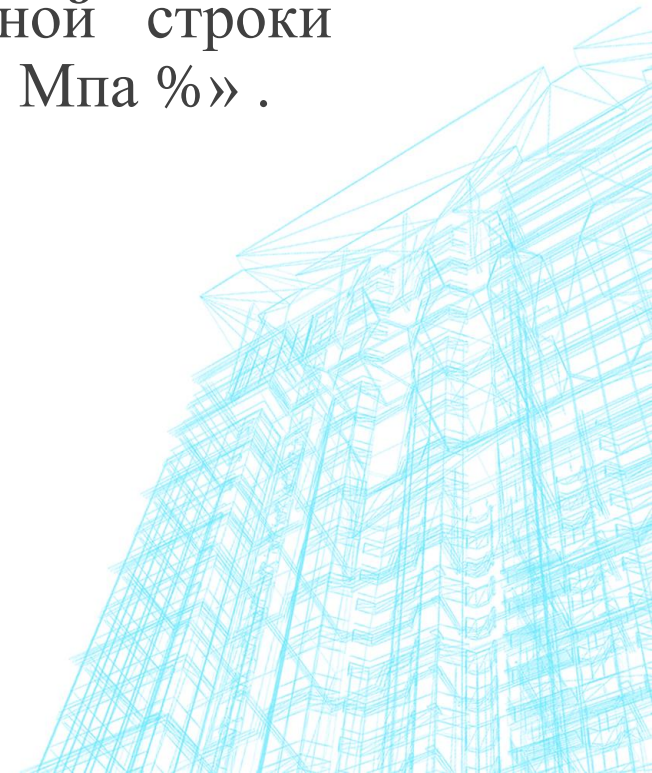


Результаты модели нейронной сети «Соотношение матрица-наполнитель %»



# Разработка приложения

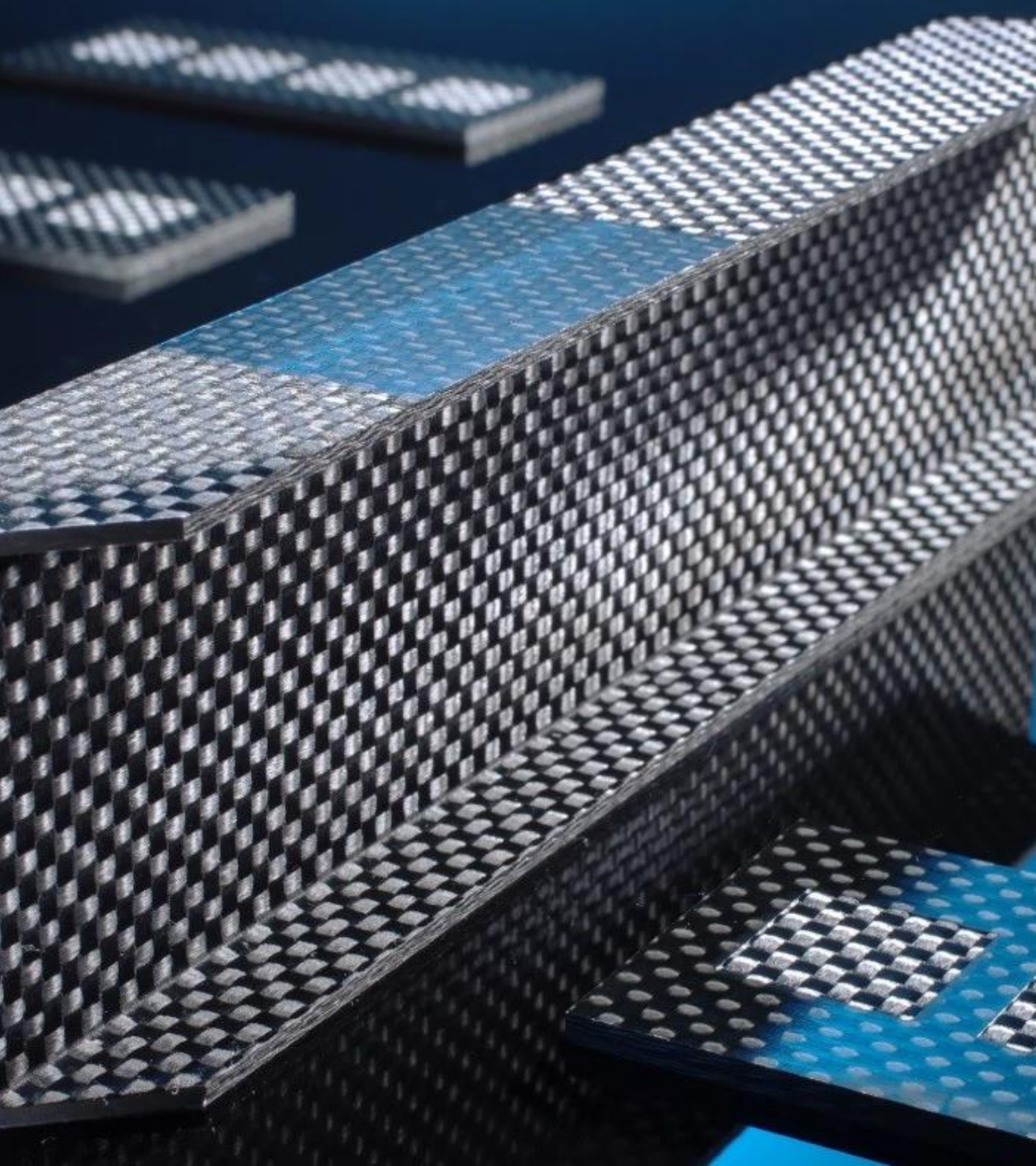
- Разработано приложение с интерфейсом командной строки которое выдает прогноз «Прочность при растяжении, Мпа %» .





# Удалённый репозиторий

- Страница создана на GitHub
- Адрес страницы: <https://github.com/rinabiu/DataScienceVKR>
- В репозитории находятся:
  - ✓ Файл тетрадки Jupyter Notebook
  - ✓ Набор данных, модель
  - ✓ Приложение
  - ✓ ВКР в тестовом формате



СПАСИБО ЗА ВНИМАНИЕ

