

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ**
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный технический университет имени Н.Э.
Баумана
(национальный исследовательский университет)»

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА
по курсу
«Data Science»

Слушатель

Кочетков Игорь Александрович

Москва, 2023

Оглавление

Введение.....	3
1. Анализ исходных данных и выбор методов решения	
1.1 Описание задачи и исходных данных.....	5
1.2 Разведочный анализ и визуализация данных.....	6
1.3 Методы решения задачи.....	13
2. Разработка моделей машинного обучения	
2.1 Предобработка данных.....	18
2.2 Разработка и обучение моделей.....	19
2.3 Тестирование моделей.....	21
2.4 Создание приложения с интерфейсом командной строки.....	26
2.5 Создание удалённого репозитория.....	27
3. Заключение.....	28
4. Список литературы.....	29

Введение

Развитие современной техники требует новых конструкционных материалов, превосходящих по своим прочностным, упругим и другим свойствам традиционные. Композиционный материал- искусственно созданный неоднородный сплошной материал, состоящий из двух и более композитов с чёткой границей раздела между ними. В большинстве композитов (за исключением слоистых) компоненты можно разделить на матрицу и включённые в нее элементы.

Машинное обучение (МО) воспринимается как многообещающий инструмент для разработки и открытия новых материалов для широкого спектра приложений. Разработка новых материалов с превосходными индивидуальными свойствами является конечной целью современных инженерных приложений. За последние несколько десятилетий, благодаря быстрому развитию высокопроизводительных параллельных вычислений, материаловедения и численного моделирования, многие важные свойства материалов теперь могут быть рассчитаны с помощью моделирования с достаточной точностью. По сравнению с простым прогнозированием свойств известных материалов разработка новых материалов для достижения регулируемых свойств является более важной задачей для научных и инженерных целей.

Эффективность и работоспособность материала зависят от правильного выбора исходных компонентов и технологии их совмещения, призванной обеспечить прочную связь между компонентами при сохранении их первоначальных характеристик.

Современные композиты изготавливаются из других материалов: полимеры, керамика, стеклянные и углеродные волокна, но данный принцип сохраняется. У такого подхода есть и недостаток: даже если мы знаем характеристики исходных компонентов, определить характеристики

композита, состоящего из этих компонентов, достаточно проблематично. Для решения этой проблемы есть два пути: физические испытания образцов материалов, или прогнозирование характеристик. Суть прогнозирования заключается в симуляции представительного элемента объема композита, на основе данных о характеристиках входящих компонентов (связующего и армирующего компонента).

В последние годы искусственный интеллект развивается очень быстрыми темпами, даже сама технология нейронных сетей стала более доступной обычным пользователям. Также появилось множество технологий, использующих нейронные сети для ускорения или упрощения процессов. Машинное обучение является самым простым вариантом искусственного интеллекта. Оно предполагает, что с помощью различных методов на основе большого количества «тренировочных» данных можно классифицировать или предсказать любой объект, явление или событие.

1. Анализ исходных данных и выбор методов решения

1.1 Описание задачи и исходных данных

Целью настоящей работы является прогнозирование характеристик «Модуля упругости при растяжении, ГПа» и «прочность при растяжении, МПа» и рекомендовать «соотношение матрица - наполнитель» материалов при помощи методов машинного обучения и построение моделей.

Данные предоставлены Центром НТИ «Цифровое материаловедение: новые материалы и вещества» (структурное подразделение МГТУ им. Н.Э. Баумана).

Имеются данные о начальных свойствах компонентов композиционных материалов (количество связующего, наполнителя, температурный режим и т.д.). Один набор данных «X_br» содержит (10 переменных вещественного типа) содержащей 1023 строки данных. Второй набор данных «X_pur» (3 переменные вещественного типа) содержащей 1040 строк данных. Данные представлены в Excel - файлах. Интересующие выходные характеристики (3 выходные переменные) содержатся в наборе «X_br». Для анализа данных будет использоваться язык программирования Python. Этому способствует простота языка, а также большое разнообразие открытых библиотек.

1.2 Разведочный анализ и визуализация данных

Предварительный анализ данных с целью выявления наиболее общих зависимостей, закономерностей и тенденций, характера и свойств анализируемых данных, законов распределения анализируемых величин. При разведочном анализе учитывается и сравнивается большое число признаков, а для поиска закономерностей используются самые разные методы. Результаты разведочного анализа помогут в разработке наилучшей стратегии углубленного анализа, выдвижение гипотез, уточнение особенностей применения тех или иных математических методов и моделей. Без разведочного анализа углубленный анализ данных будет производиться практически «вслепую».

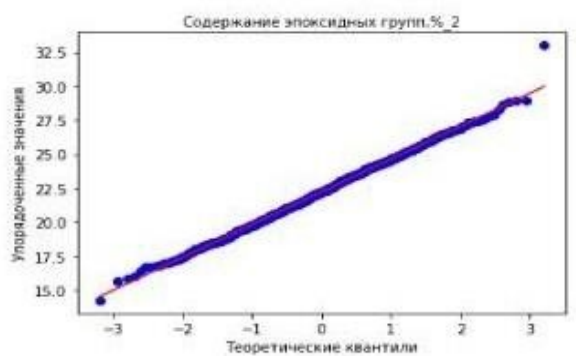
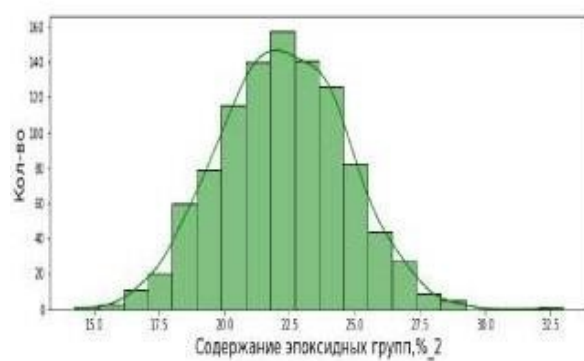
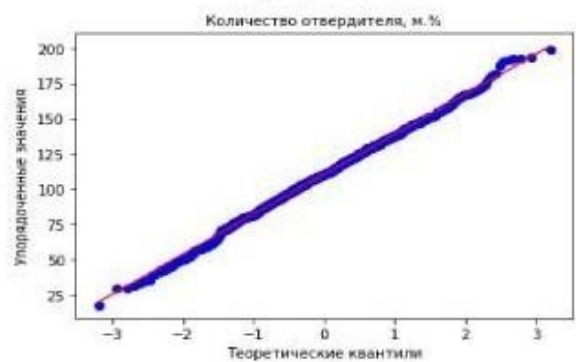
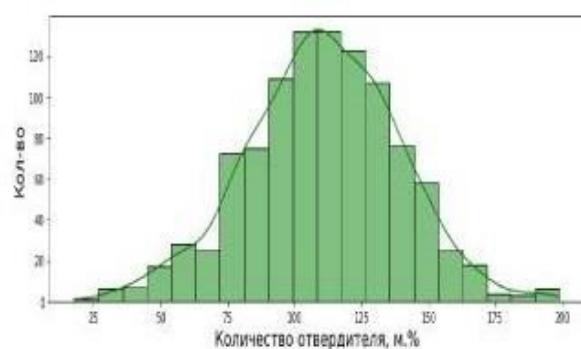
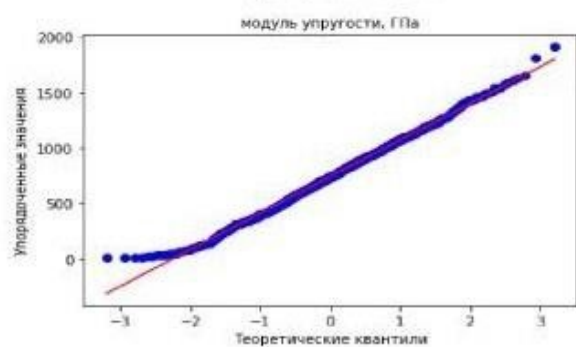
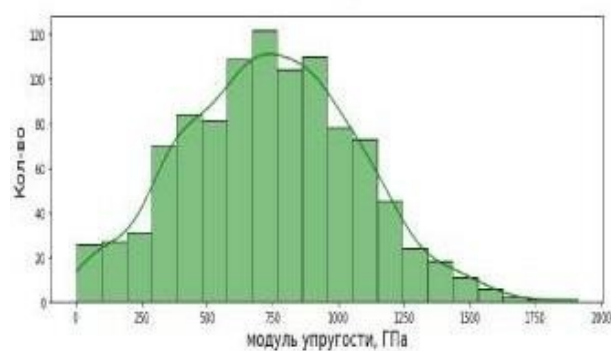
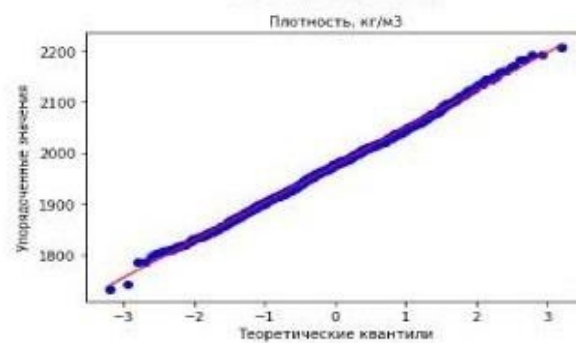
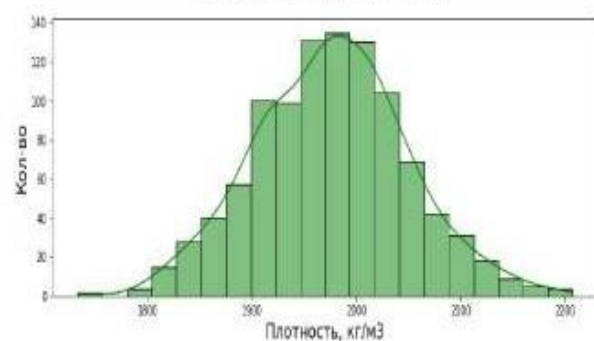
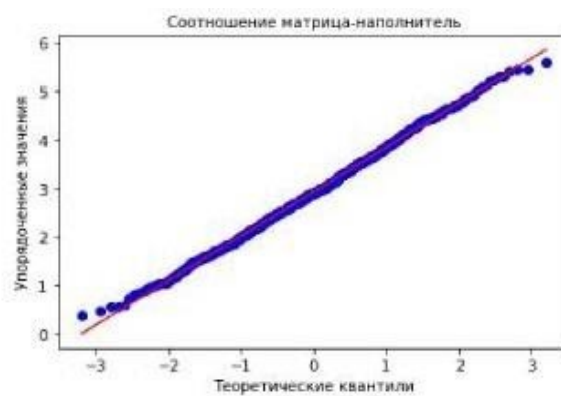
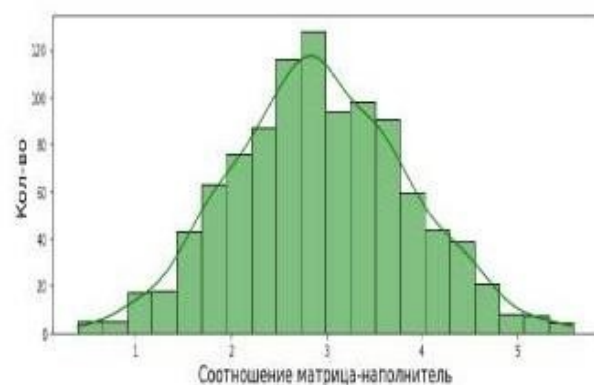
Одна из главных библиотек при разведочном анализе это Pandas. С помощью данной библиотеки мы загрузим наши данные (и объединим по типу INNER).

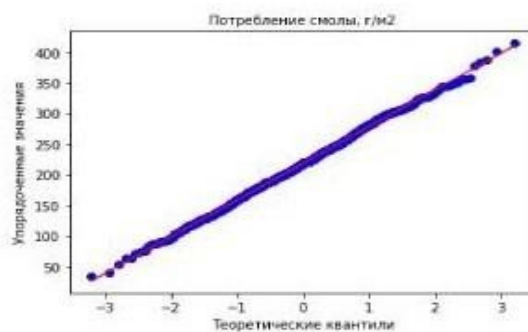
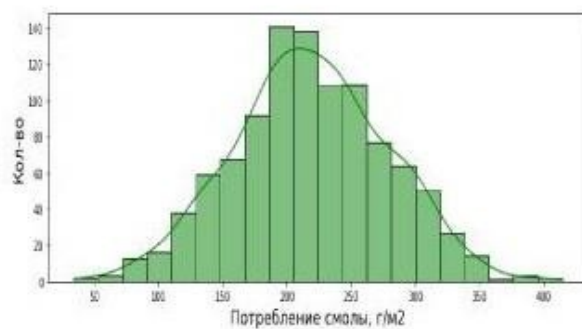
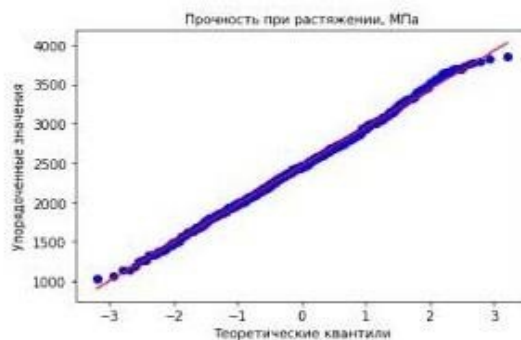
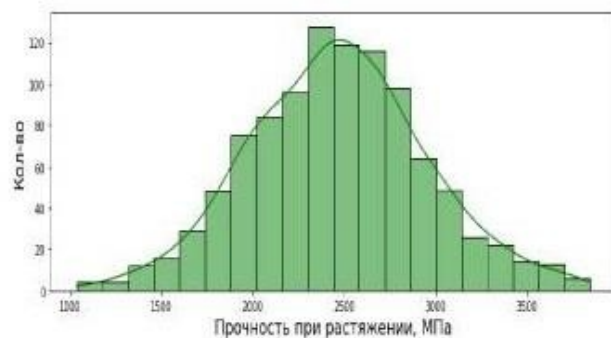
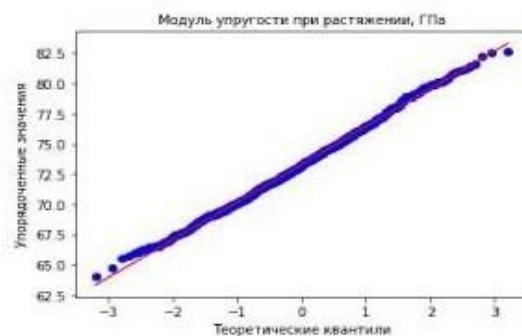
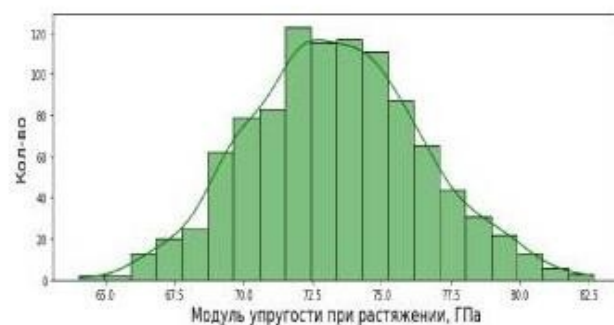
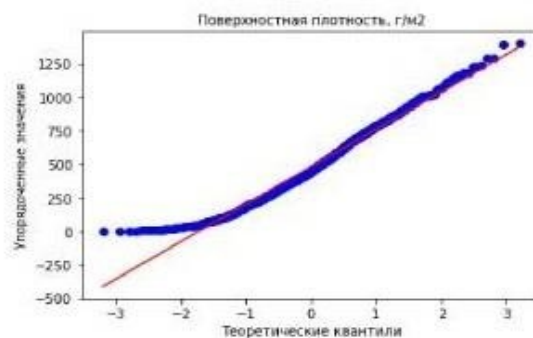
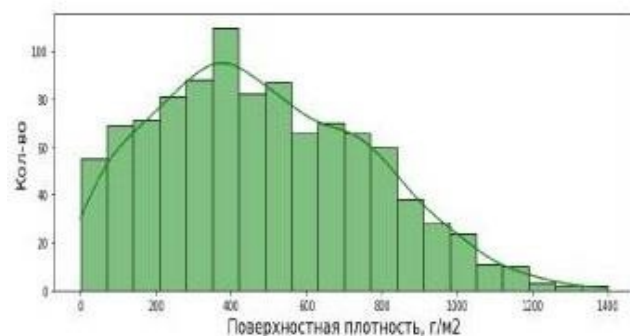
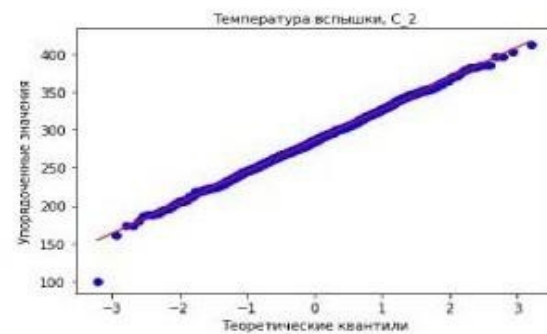
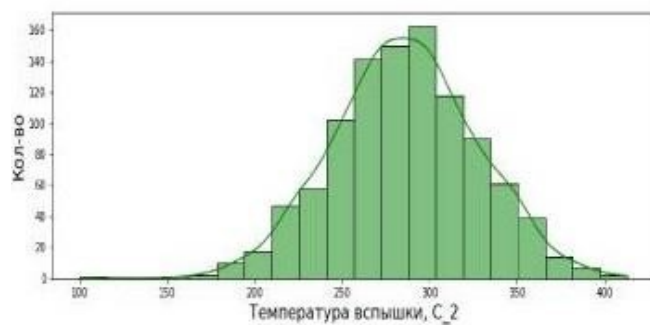
Таблица.1 Наименование параметров

	Соотношение матрица-наполнитель	Плотность, кг/м3	модуль упругости, ГПа	Количество отвердителя, м.%	Содержание эпоксидных групп, %_2	Температура вспышки, C_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	125.007689	90.0	9.076380	47.019770
1019	3.444022	2050.089171	444.732634	145.961978	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 x 13 columns

Часть данных 17 строк невозможно включить в таблицу из-за различия количества строк в наборах данных в процессе компоновки. Исходный датафрейм содержит 1023 строки с входными параметрами и 13 колонок переменных как показано в в таблице 1. Визуализация гистограмм, графиков квантиль-квантиль (QQ) и диаграмм («ящик с усами») позволяют наглядно увидеть характер распределения переменных.





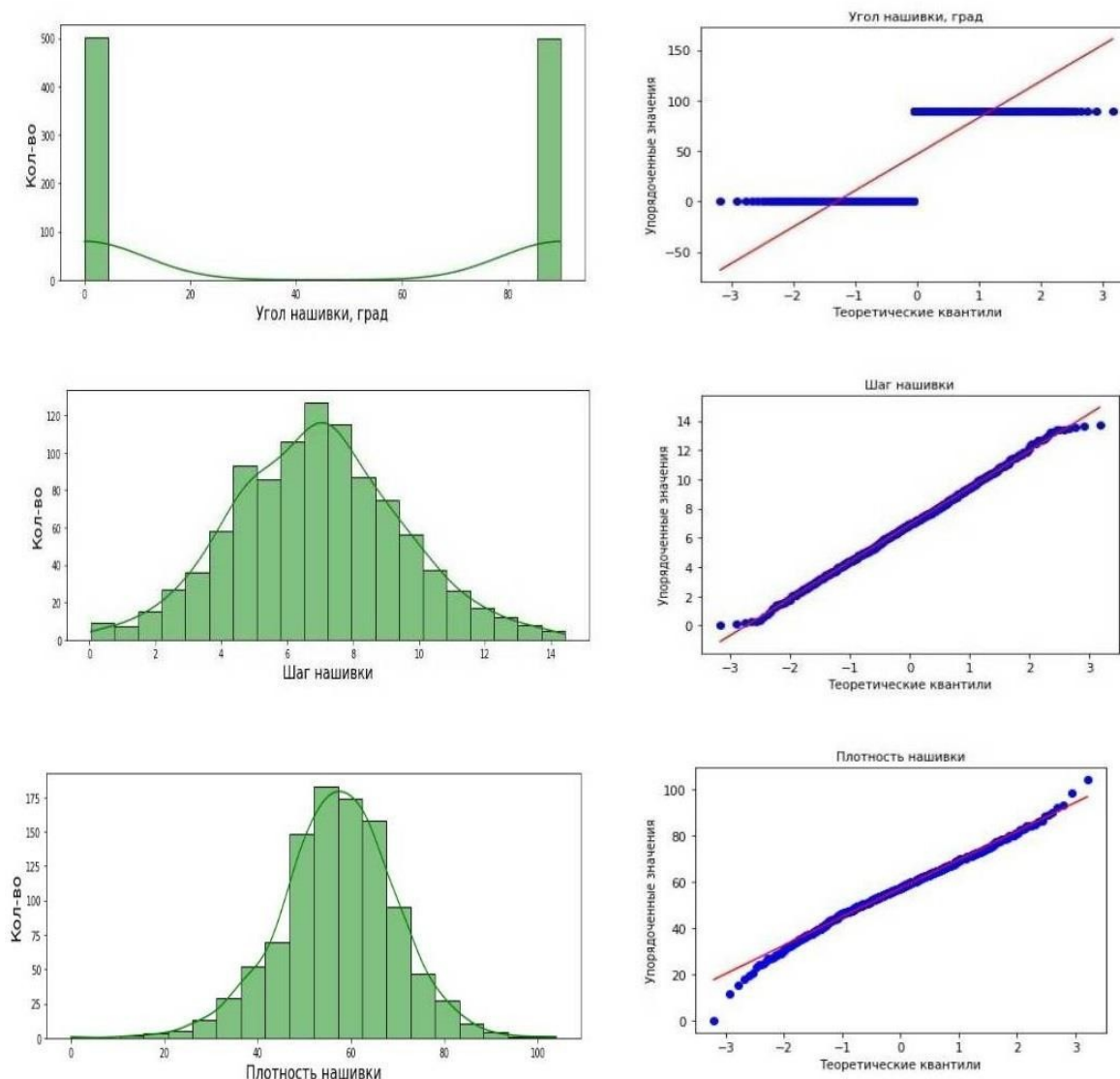


Рисунок 1 - Гистограммы распределения и QQ графики

Все переменные имеют нормальное распределение рисунок 1, кроме «Угол нашивки, град», где всего два значения, а также выделяется «Поверхностная плотность, г/м2» форма не похожа на нормальное распределение, медиана отличается от среднего значения.

Таблица 2. Описательная статистика исходных данных

	Соотношение матрица-наполнитель	Плотность, кг/м3	модуль упругости, ГПа	Количество отвердителя, м.%	Содержание эпоксидных групп, %_2	Температура вспыхи, 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
count	1023.000000	1023.000000	1023.000000	1023.000000	1023.000000	1023.000000	1023.000000	1023.000000	1023.000000	1023.000000	1023.000000	1023.000000	1023.000000
mean	2.930366	1975.734888	739.923233	110.570769	22.244390	285.882151	482.731833	73.328571	2466.922843	218.423144	44.252199	6.899222	57.153929
std	0.913222	73.729231	330.231581	28.295911	2.406301	40.943260	281.314690	3.118983	485.628006	59.735931	45.015793	2.563467	12.350969
min	0.389403	1731.764635	2.436909	17.740275	14.254985	100.000000	0.603740	64.054061	1036.856605	33.803026	0.000000	0.000000	0.000000
25%	2.317887	1924.155467	500.047452	92.443497	20.608034	259.066528	266.816645	71.245018	2135.850448	179.627520	0.000000	5.080033	49.799212
50%	2.906878	1977.621657	739.664328	110.564840	22.230744	285.896812	451.864365	73.268805	2459.524526	219.198882	0.000000	6.916144	57.341920
75%	3.552660	2021.374375	961.812526	129.730366	23.961934	313.002106	693.225017	75.356612	2767.193119	257.481724	90.000000	8.586293	64.944961
max	5.591742	2202.773481	1911.536477	198.953207	33.000000	413.273418	1399.542362	82.682051	3848.436732	414.590628	90.000000	14.440522	103.988901

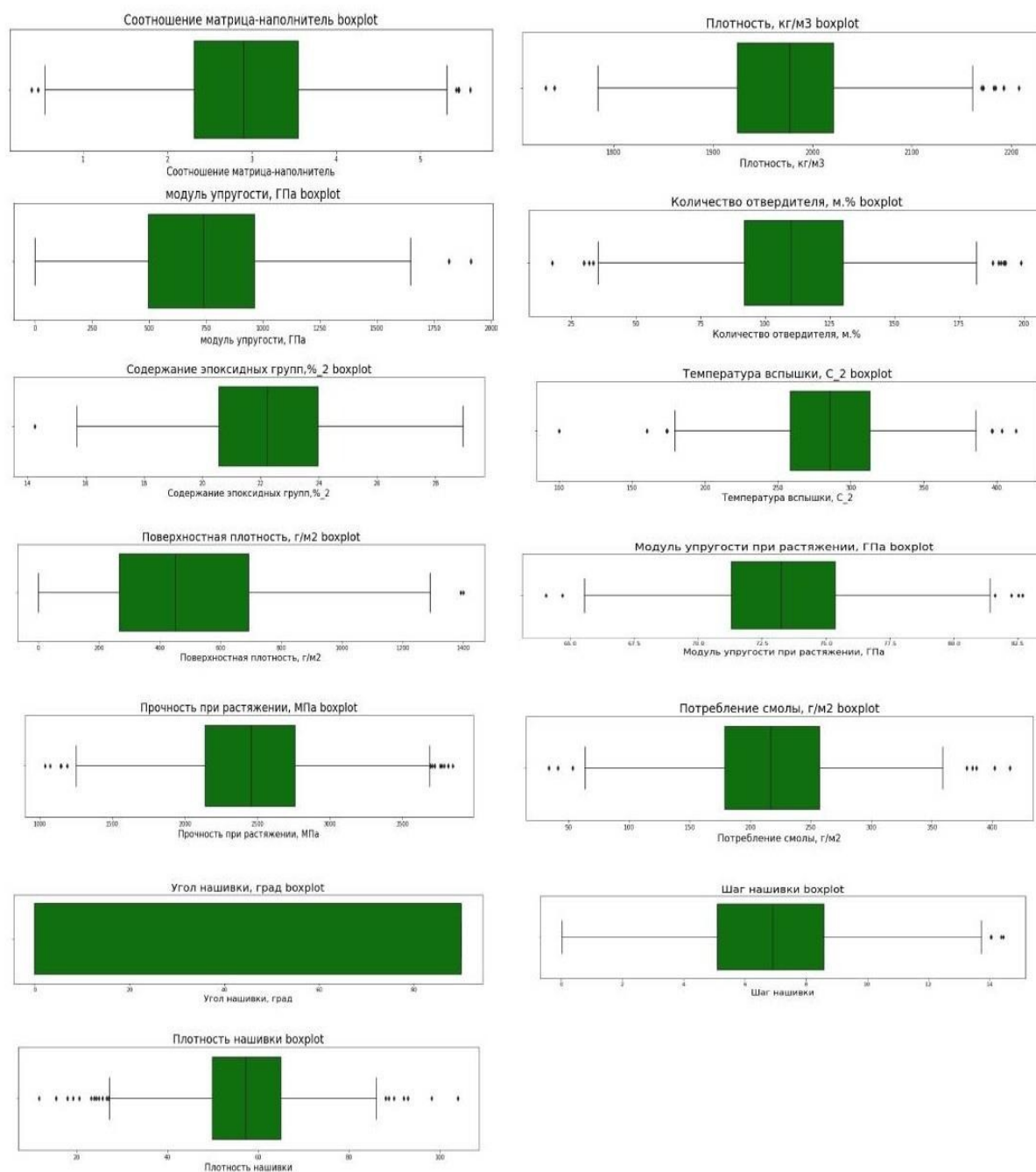


Рисунок 2 - Диаграммы размаха

В таблице 2 описательная статистика, также видно, что у переменной «Поверхностная плотность, г/м2» медианное значение меньше среднего. Диаграммы размаха («ящик с усами») кроме «Угол нашивки, град» показывают незначительные значения за пределами полутора межквартильных расстояний от первого и третьего квартиля на рисунке 2. Так же был проверен датафрейм на уникальные значения в каждой колонке и удалены дубликаты до колонки «Потребление смолы, г/м2» включительно. Для удаления выбросов был применён метод на основе межквартильного расстояния.

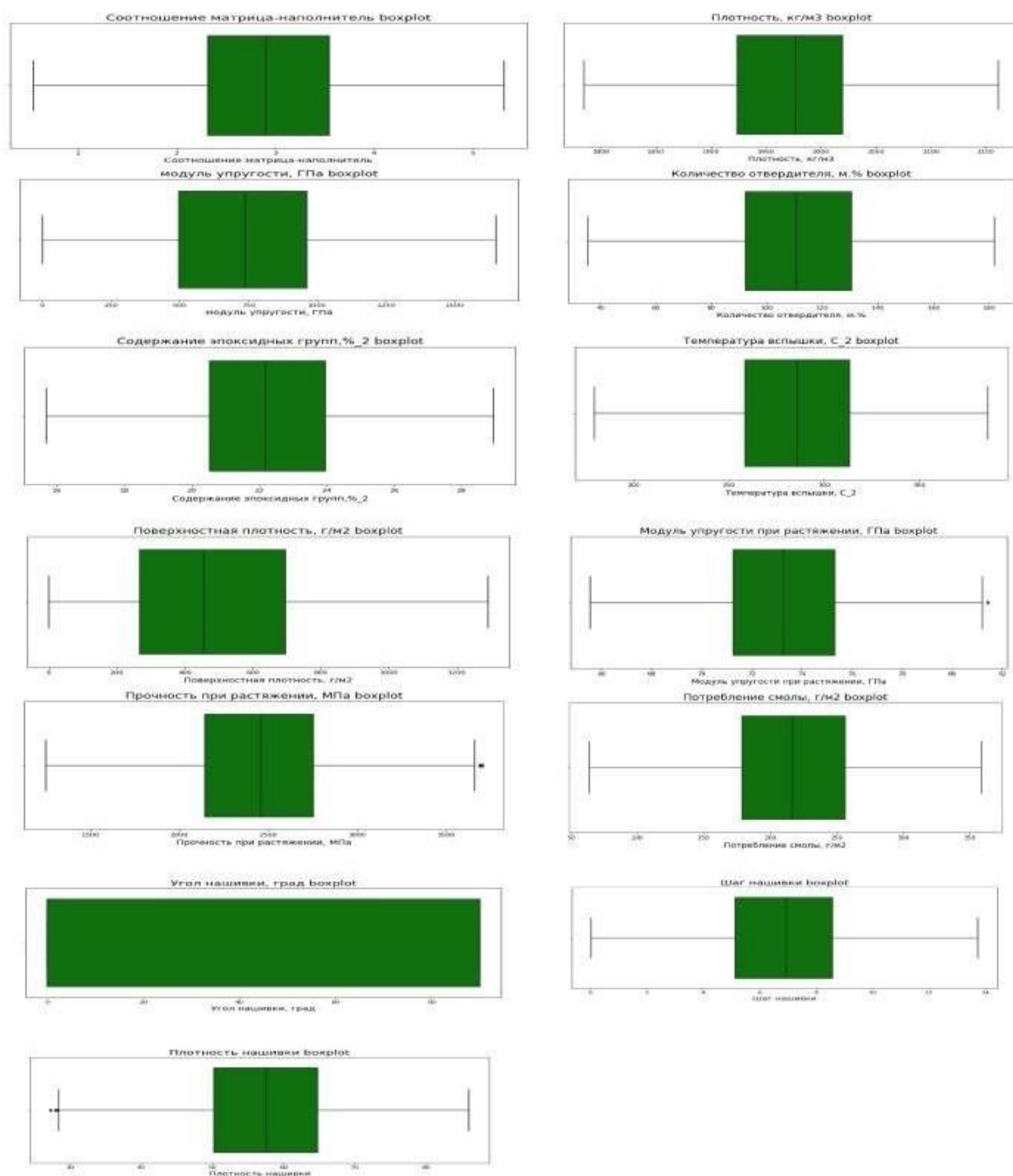


Рисунок 3 - Диаграммы размаха после удаления выбросов.

Как видно на рисунке 3 после удаления выбросов диаграммы подтянулись к медианному значению, кроме диаграммы «Поверхностная плотность г/м2» и «Угол нашивки, град»

Для анализа корреляционных признаков были построены корреляционная тепловая карта и диаграмма рассеивания.

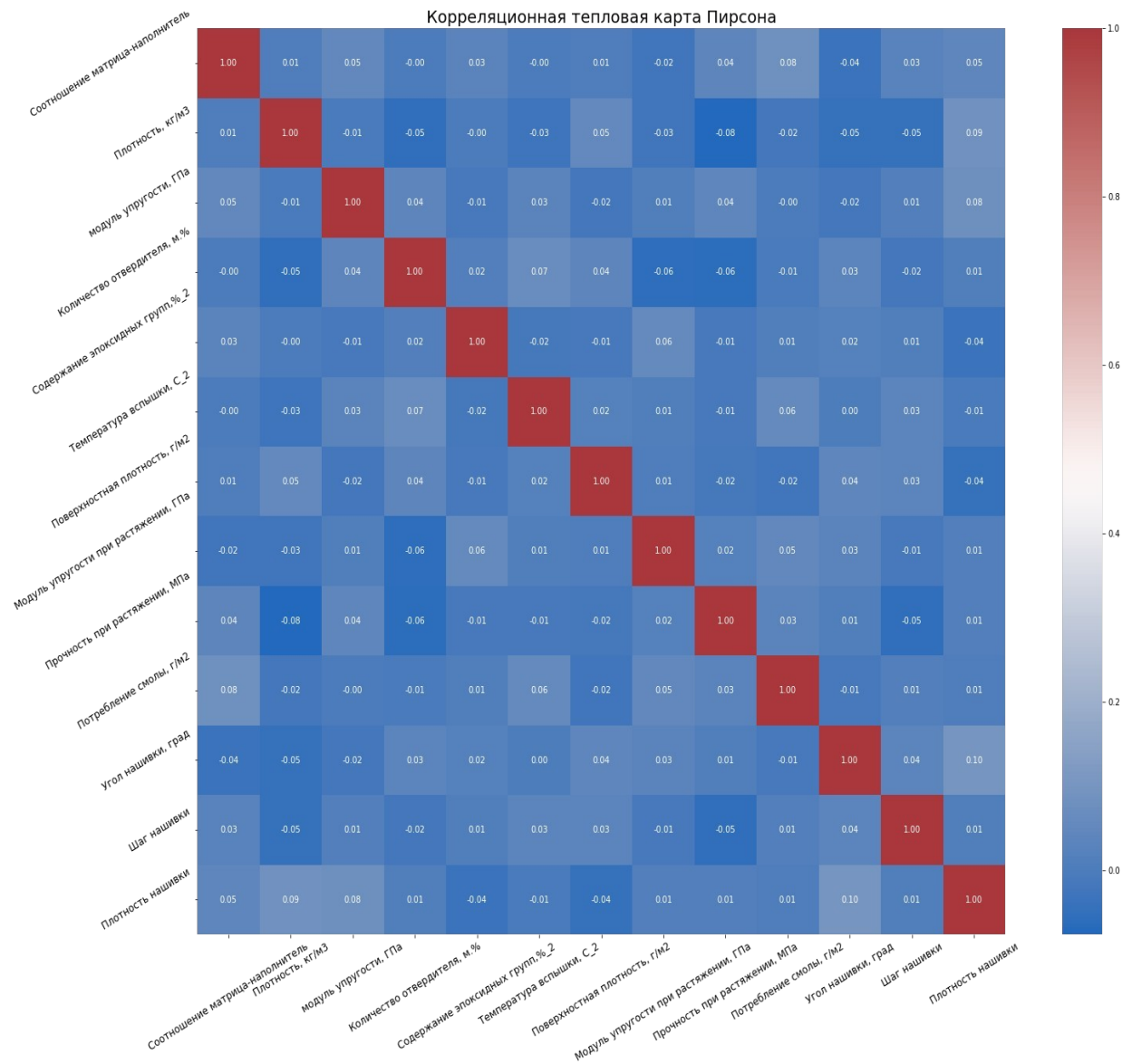


Рисунок 4 - Корреляционная тепловая карта

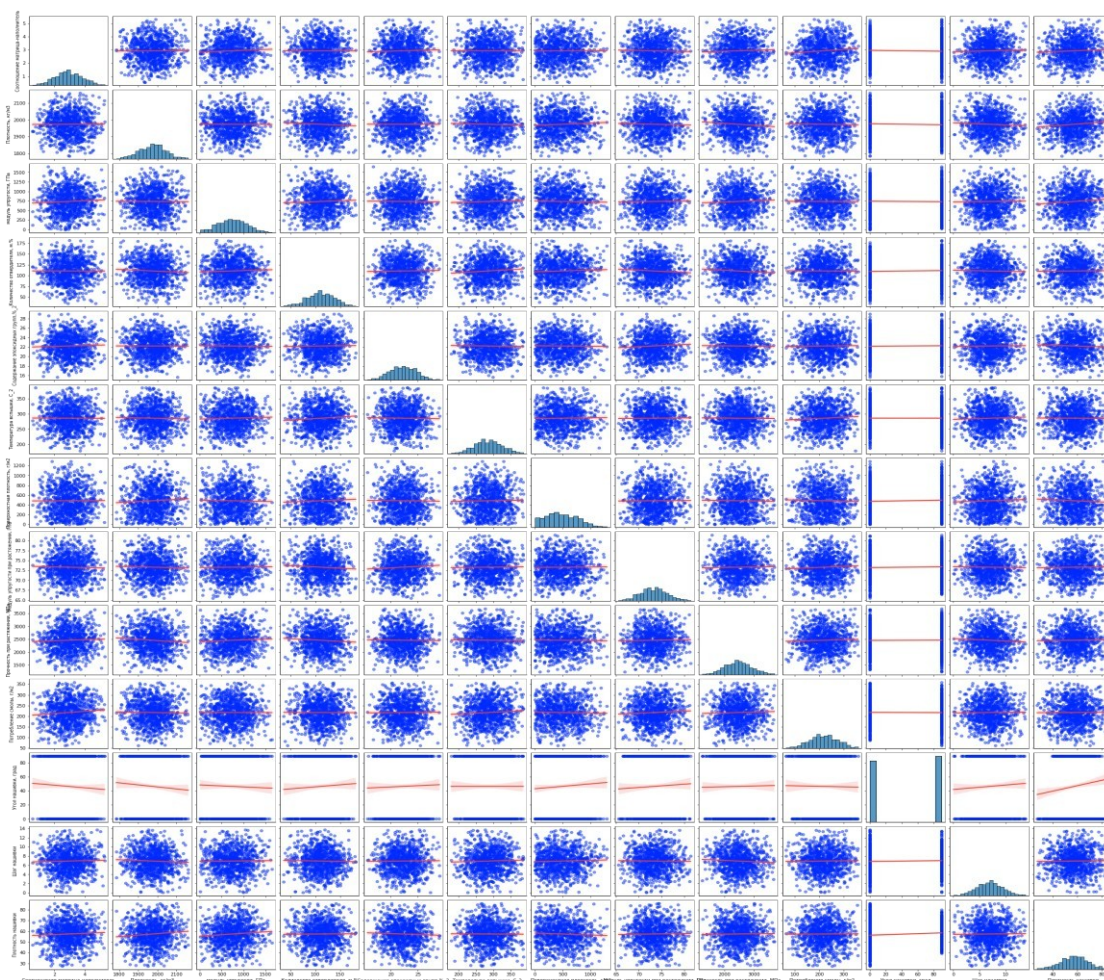


Рисунок 5 - Диаграмма рассеивания переменных

1.3 Методы решения задачи

Разведочный анализ данных показал, что линейной связи между любыми переменными нет, корреляция равна 0 это наглядно видно на рисунке 4 и 5. Для дальнейшего решения задачи виделось решение в конструирование признаков т.е. процесс трансформации существующих или создание на их основе новых с целью улучшения работы алгоритмов машинного обучения. В начале была создана новая переменная «Тотал» - общий вес всех ингредиентов (компонентов) материала. Наличие этой переменной показало положительную динамику в улучшении моделей. Поскольку совокупная плотность меняется, меняется и их отношение плотности каждого ингредиента (компонента) к общей плотности. Поэтому были трансформированы переменные в новые, для каждого ингредиента

(компонента), которые содержат информацию о его плотности, выраженная в долях (процентах) для масштабирования данных все значения умножены на 10000. Оказалось, что наличие этих колонок, и колонки «Тотал» положительно повлияло на точность.

Наименования новых 14 переменных:

- Тотал - общий вес переменных (ингредиентов, параметров)
- Соотношение матрица - наполнитель % - доля «Соотношение матрица - наполнитель» от «Тотал»
- Плотность, кг/м³ % - доля «Плотность, кг/м³» от «Тотал»
- Модуль упругости, гПа % - доля «Модуль упругости, гПа» от «Тотал»
- Количество отвердителя, гПа - % - доля «Количество отвердителя, гПа» от «Тотал»
- Содержание эпоксидных Групп, %_2 % - доля «Содержание эпоксидных групп, %_2» от «Тотал»
- Температура вспышки, С_2 % - доля «Температура вспышки, С_2» от «Тотал»
- Поверхностная Плотность, г/м² % - доля «Поверхностная Плотность, г/м²» от «Тотал»
- Модуль упругости при растяжении, гПа % - доля «Модуль упругости при растяжении, гПа» от «Тотал»
- Прочность при растяжении % - доля «Прочность при растяжении» от «Тотал»
- Потребление смолы, г/м² % - доля «Потребление смолы, г/м²» от «Тотал»
- Угол нашивки, град % - доля «Угол нашивки, град » от «Тотал»
- Шаг нашивки % - доля «Шаг нашивки » от «Тотал»
- Плотность нашивки % - доля «Плотность нашивки» от «Тотал»

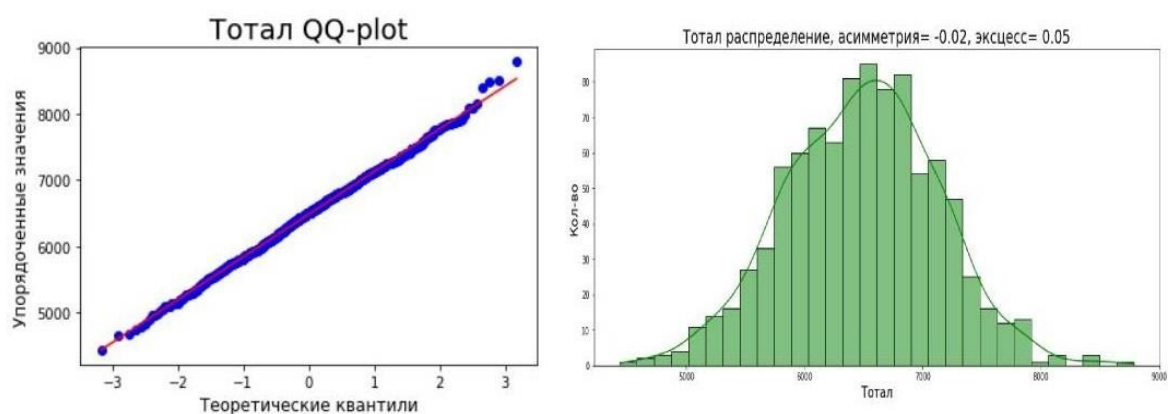


Рисунок 6 - Гистограмма распределения и QQ переменной «Тотал»

Переменная «Тотал» имеет нормальное распределение рисунок 6. График (QQ квантиль- квантиль) - это подтверждает, распределение вдоль линии точек под углом 45 градусов от нижнего левого угла до правого верхнего угла. Для удаления выбросов был применён метод на основе межквартильного расстояния.

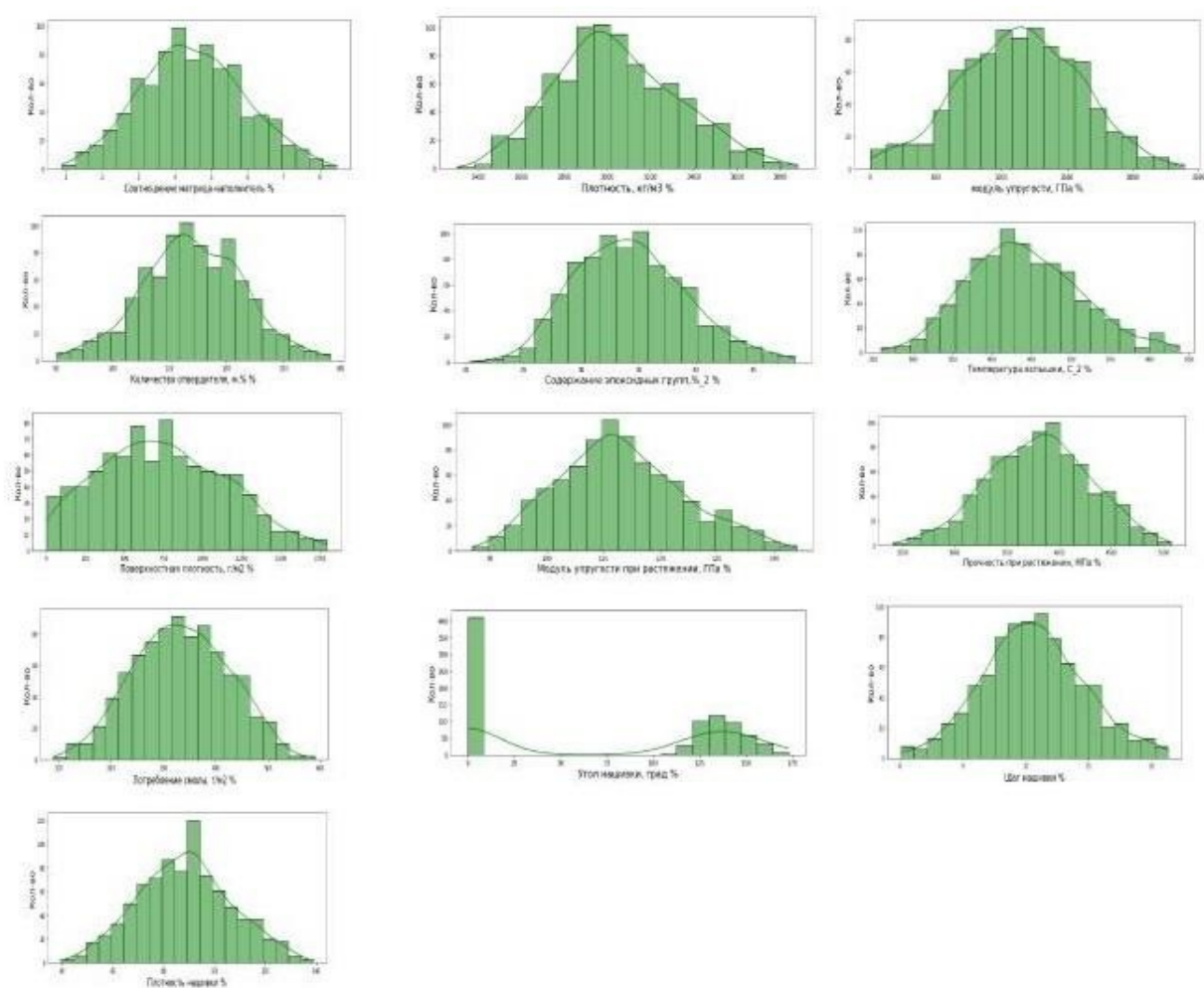


Рисунок 7 - Гистограммы распределения переменных

Стоит отметить на рисунке 7 все переменные имеют относительно нормальное распределение, Гистограммы - это подтверждают, а вот переменная «Угол нашивки, град %», где раньше было всего два значения 0 и 90 при трансформации поменяла распределение.

Для построения прогноза «Модуль упругости при растяжении, гПа %», «Прочность при растяжении %» будут применены следующие модели:

- Гребневая регрессия (ridge regression)
- Лассо регрессия
- RandomForestRegressor
- GradientBoostingRegressor

Для рекомендации «Соотношение Матрица - накопитель %» будет построена нейронная сеть с использованием Tensorflow Keras, будут протестированы линейная регрессия и регрессия DNN.

Теперь построим новую корреляционную тепловую карту и диаграмму рассеивания нового трансформированного датасета.

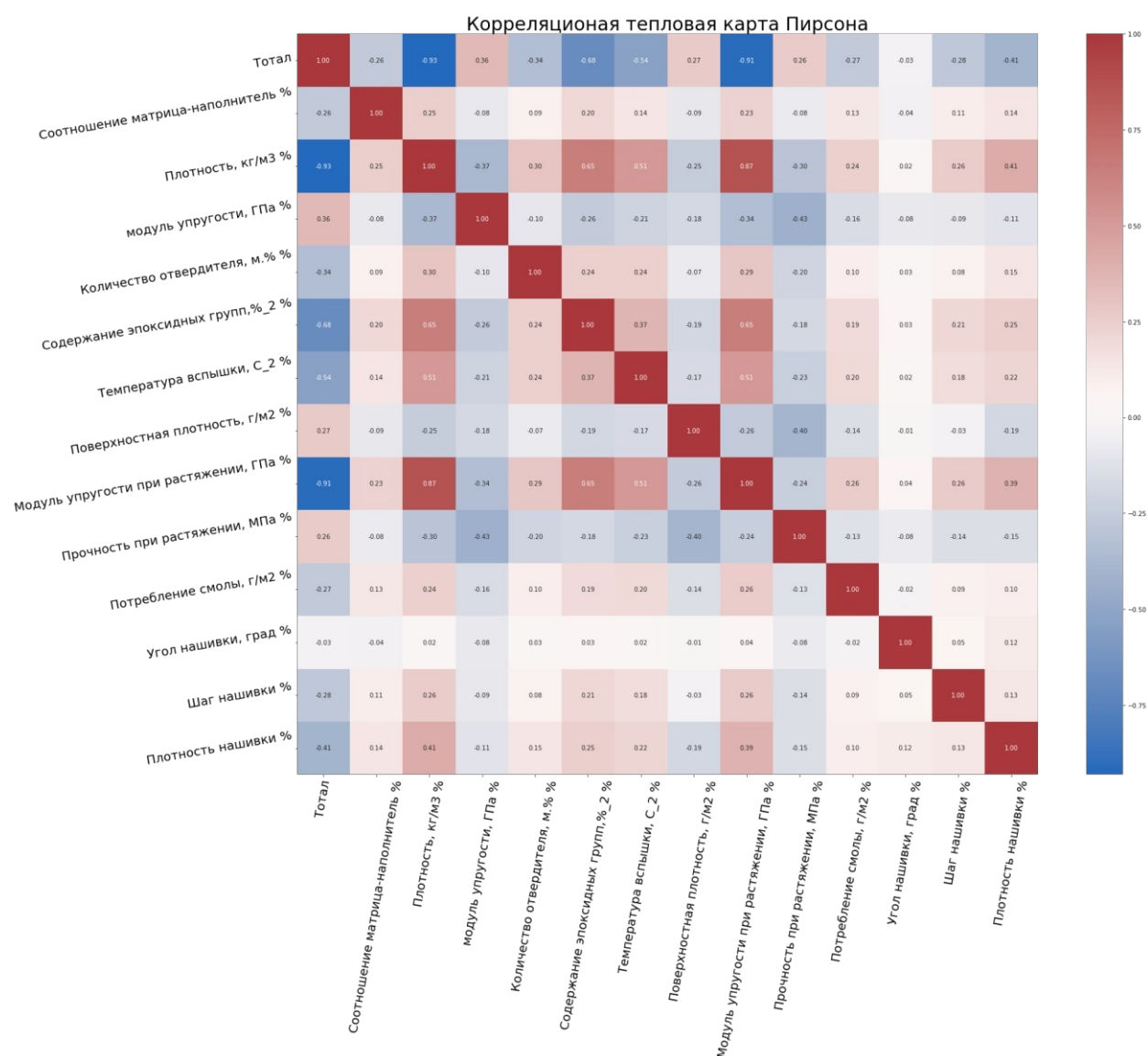


Рисунок 8 - Корреляционная тепловая карта

Теперь достаточно коррелирующих переменных это видно на рисунке 8 для прогнозирования «Модуль упругости при растяжении, гпа %», «Прочность при растяжении %» и чтобы рекомендовать «Соотношение матрица - наполнитель %», был произведен отбор признаков.

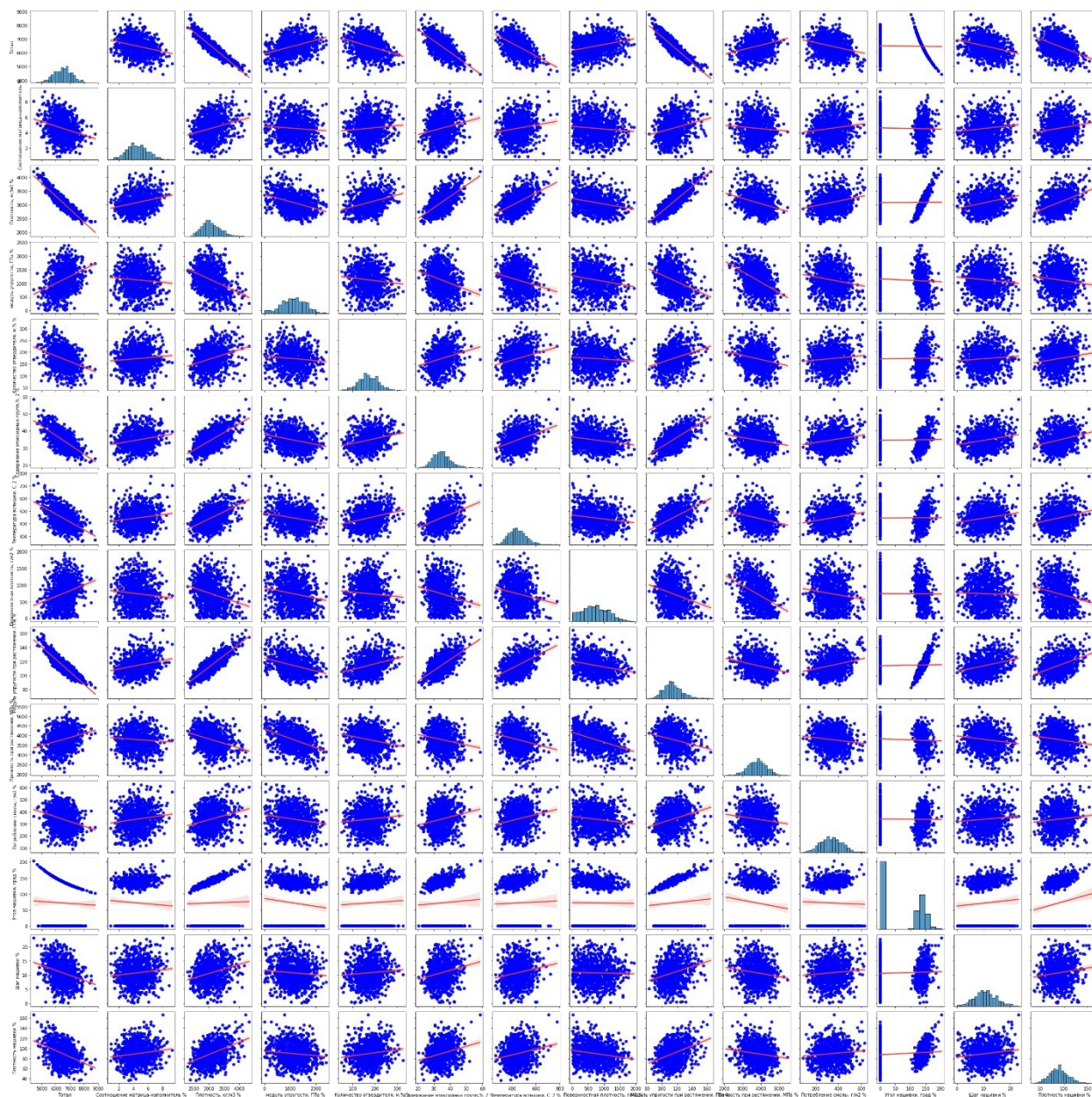


Рисунок 9. Диаграмма рассеивания нового датасета

2. Разработка моделей машинного обучения

2.1 Предобработка данных

Благодаря широкому использованию в предварительной обработке данных, анализе данных и машинном обучении Pandas в сочетании с Numpy, Scikit-Learn и Matplotlib фактически становится стеком данных в Python.

- 1) Загрузка исходных данных в Excel файлах, объединение по типу INNER в исходный датафрейм с удалением поля индекса.
- 2) Удаление дубликатов (до колонки «Потребление смолы, г/м2»)
- 3) Удаление выбросов методом межквартильного расстояния
- 4) Создание новых признаков из существующих данных и трансформация переменных для обучения моделей машинного обучения
- 5) Разбиение тренировочной и валидационной выборок на датафреймы, содержащие входные и целевые переменные.
- 6) Шкалирование данных (трансформация, нормализация)

После предобработки исходный датасет стал содержать 854 строк, которые были разделены на 597 тренировочных и 257 проверочных, в соответствии с условием задачи. Для нормализации использовался `MinMaxScaler()` в библиотеке `Sklearn`. Таким образом мы преобразовали значения в диапазон от 0 до 1.

2.2 Разработка и обучение моделей

Для прогнозирования «Модуль упругости при растяжении, гПа %» и «Прочность при растяжении %». Основной акцент сделан на гребневую регрессию и лассо регрессию. Для понимания гребневой (ридж) регрессии используют методы регуляризации для того чтобы удостовериться, что модель принимает решение на основании независимых переменных, которые оказывают значительное влияние на результирующую переменную, т.е. регуляризация помогает модели обращать внимание лишь на те признаки, которые действительно имеют значение и избавляются от шумов. В Лассо регрессии модель может полностью избавиться от шумов. Для сравнения были использованы еще модели `RandomForestRegressor` и `GradientBoostingRegressor`.

Реализации в специализированных библиотеках машинного обучения метода поиска гиперпараметров по сетке позволяют по заданным различным значениям гиперпараметров выбранного класса моделей протестировать возможные сочетания. Для поиска гиперпараметров по сетке использовался метод GridSearchCv библиотеки Sklearn.

Для прогнозирования «Модуля упругости при растяжении, гПа %» поиск по сетке с перекрестной проверкой показал:

Гребневая регрессия

- Сила регуляризации (alpha)- 0.0004
- Решатель (solver)- lsqr

Лассо регрессия

- Сила регуляризации (alpha)- 0.0006

GradientBoostingRegressor, RandomForestRegressor настройки по умолчанию

Для прогнозирования «Прочности при растяжении поиск %» по сетке с перекрестной проверкой показал:

Гребневая регрессия

- Сила регуляризации (alpha)- 0.0001
- Решатель (solver)- sag

Лассо регрессия

- Сила регуляризации (alpha)- 0.0006

GradientBoostingRegressor, RandomForestRegressor настройки по умолчанию

Для рекомендации «Соотношение матрица-наполнитель %» был построен регрессор DNN. Архитектура нейронной сети: последовательная входной слой содержит 64 слоя и активационную функцию «sigmoid», слой Dropout между входным и скрытым слоем с частотой 50 процентов, второй скрытый слой содержит 32 слоя и активационную функцию «sigmoid» и один

выходной слой. С помощью метода `compile()` настроен процесс компиляции это необходимо перед обучением. В качестве функции ошибки будет использоваться средняя абсолютная ошибка (MAE) и оптимизатор Adam. Данные поделены на тренировочные и тестовые (20%).

2.3 Тестирование моделей

Обучение моделей регрессии подобранных с использованием поиска гиперпараметров по сетке, проводились в среде Jupyter Notebook. Оценка точности предсказаний проводилась с использованием оставшихся 30 % тестовых данных.

Визуализация соотношений, тестированных данных и предсказанных моделями для «Модуля упругости при растяжении, гПа %» и «Прочности при растяжении %»

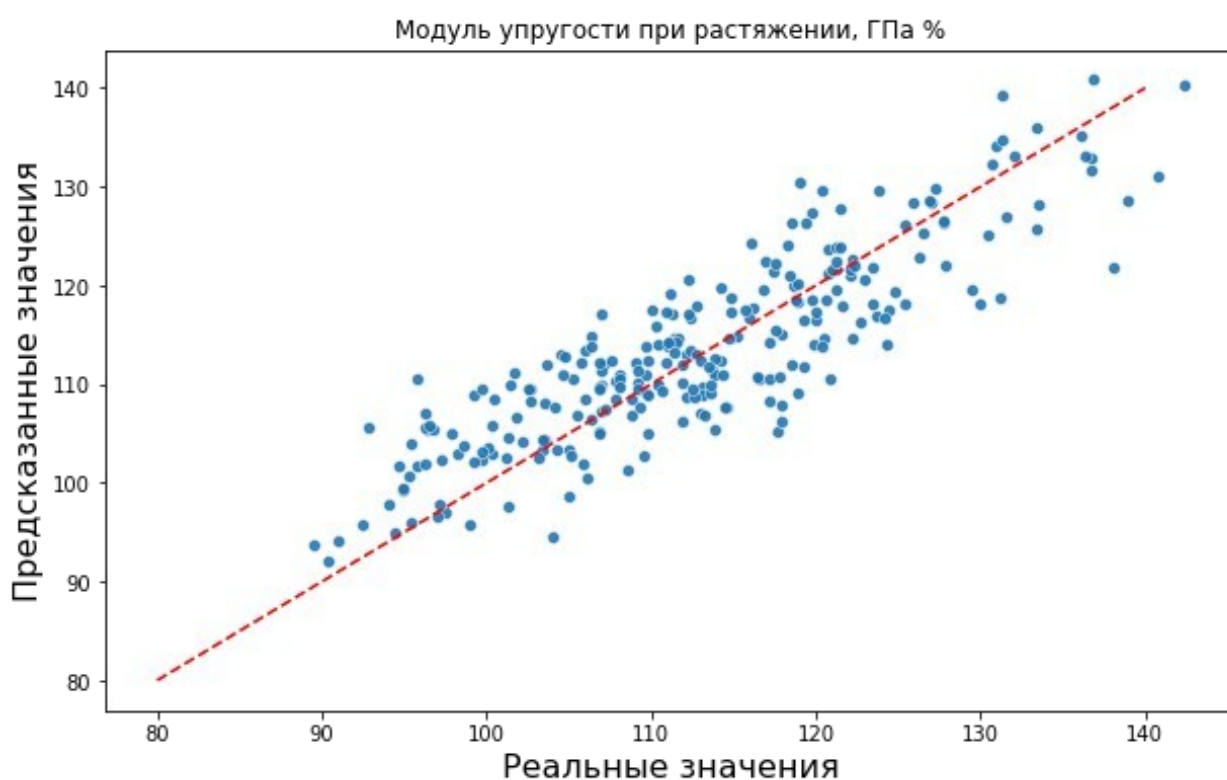


Рисунок 10. Гребневая регрессия «Модуль упругости при растяжении, гПа %»

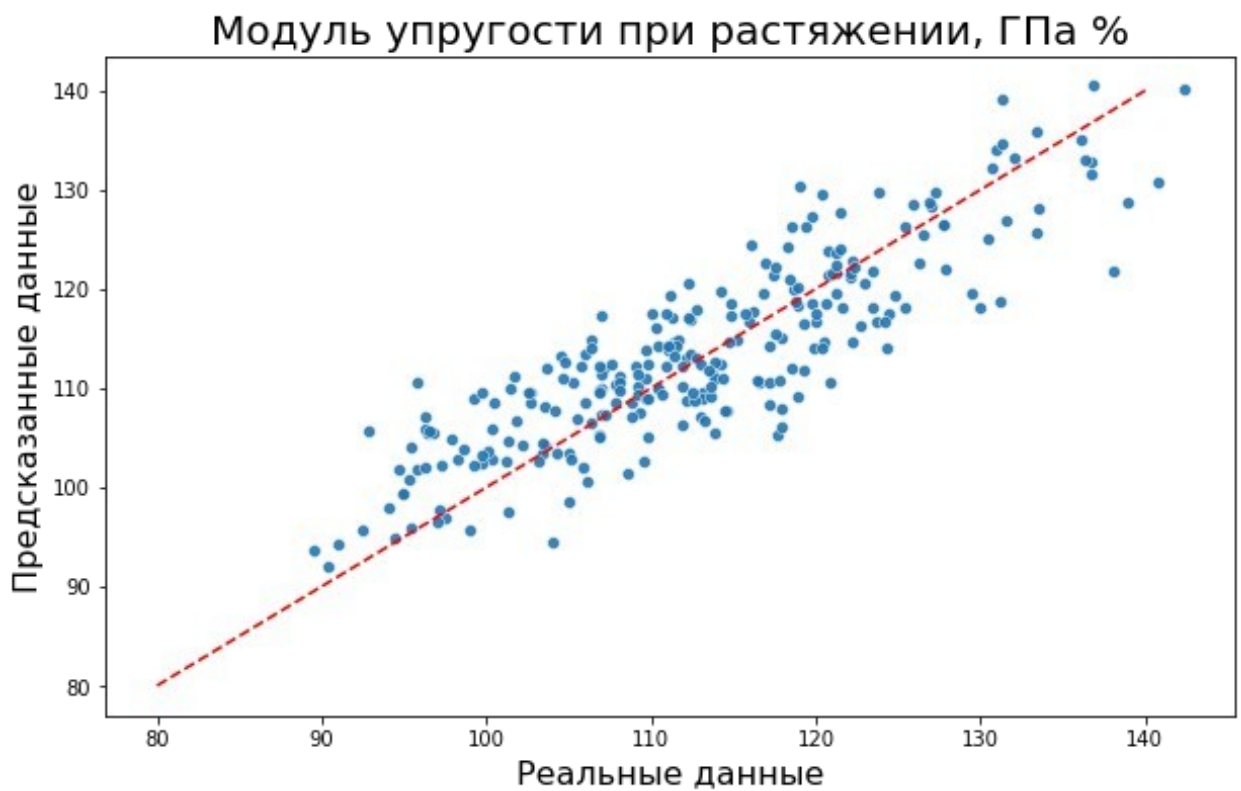


Рисунок 11. Лассо регрессия «Модуль упругости при растяжении, гпа %»

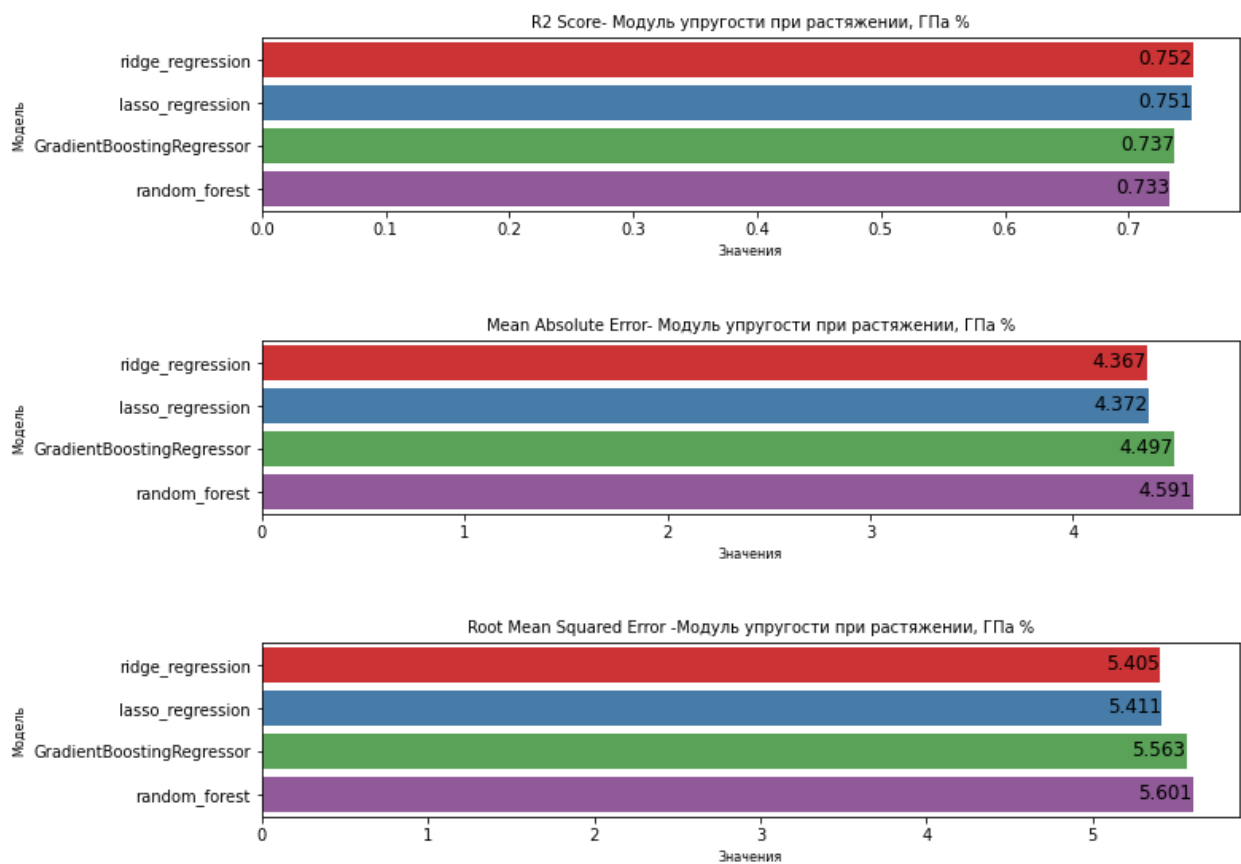


Рисунок 12. Метрика и ошибки «Модуль упругости при растяжении, гпа %»

Для оценки качества моделей регрессии использовались специальные показатели.

- 1) R^2_score (коэффициент детерминации) принимает значение от 0 до 1 и показывает долю объяснённой дисперсии объясняемого рода. Чем ближе R^2 к 1, тем меньше доля необъяснённого.
- 2) Mean Absolut Error (MAE) - средняя абсолютная ошибка, показывает среднее значение абсолютных отклонений между наблюдаемыми и прогнозируемыми значениями
- 3) Root Mean Squared Error (RMSE) - среднеквадратичная ошибка, показывает расстояние между двумя точками.

Теперь про результаты, как видно из рисунка 10 и 11 модели Лассо регрессии и Ридж регрессии неплохо спрогнозировала «Модуль упругости при растяжении, ГПа %» это подтверждает и коэффициент детерминации R^2 который составил 0.75 (рисунок 12), и небольшой средней абсолютной ошибкой. Модели удовлетворительно справились со своей задачей уменьшения размерности. Остальные две модели GradientBoostingRegressor и RandomForestRegressor использовались с настройками по умолчанию для сравнения.

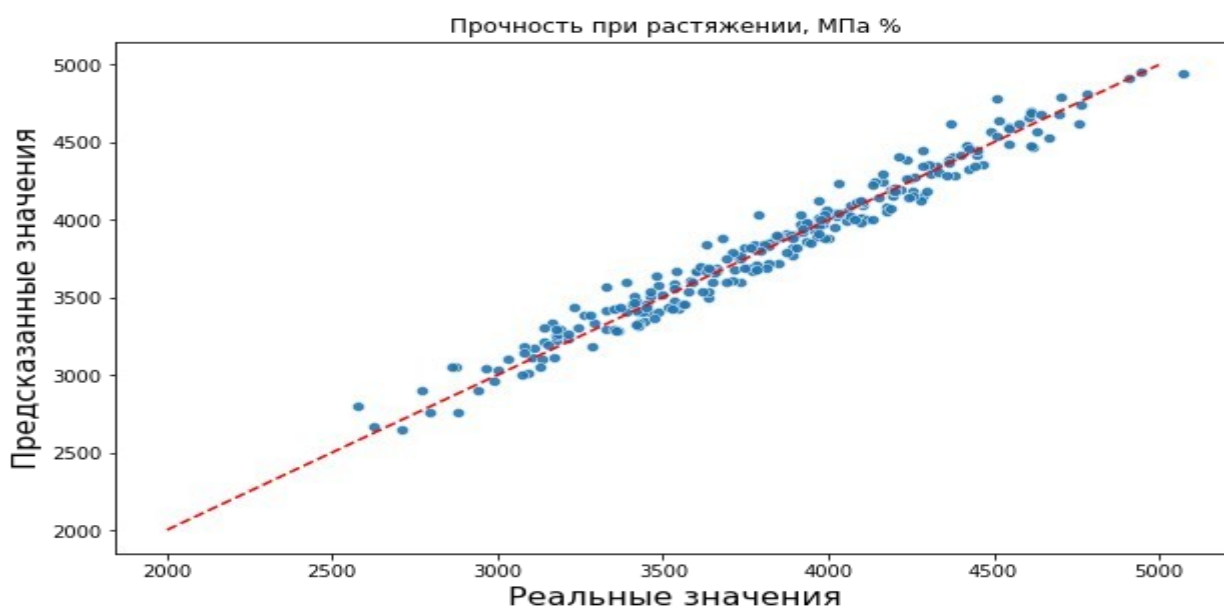


Рисунок 13. Гребневая регрессия «Прочность при растяжении, МПа %»

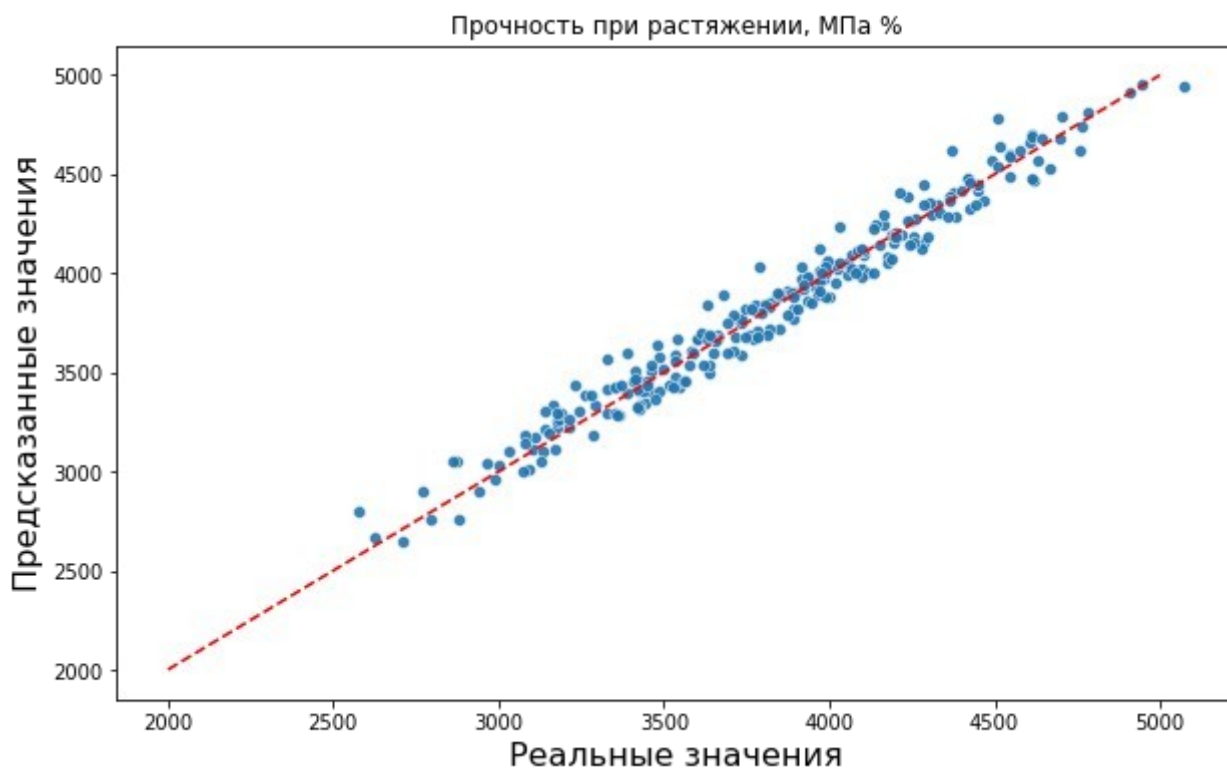


Рисунок 14. Лассо регрессия «Прочность при растяжении, МПа %»

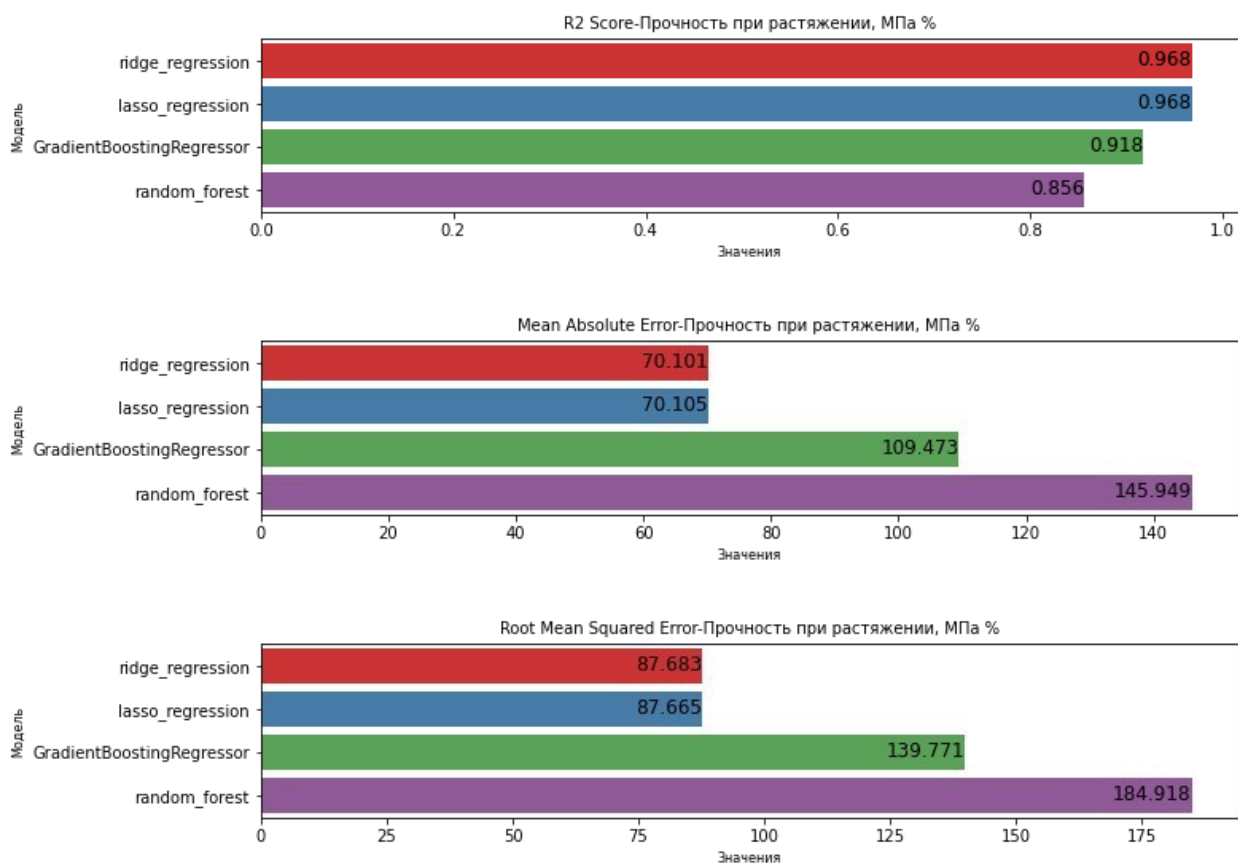


Рисунок 15. Метрика и ошибки «Прочность при растяжении, МПа %»

Результаты прогнозирования «Прочность при растяжении, ГПа %» гораздо лучше, как видно на рисунках 13 и 14 модели лассо регрессия и гребневой регрессия отлично справились с задачей, предсказанные значения и реальные проецируются вдоль линии. Коэффициент детерминации R^2 очень близок к 1 (модель имеет высокую значимость), и столь небольшой средней абсолютной ошибкой. Стоит отметить RandomForestRegressor показал почти такой же R^2 , но средняя абсолютная ошибка была значительно хуже.

Визуализация соотношений значений тестовых данных и значений, полученных обучением модели регрессии DNN для «Соотношение матрица-наполнитель %» [9].

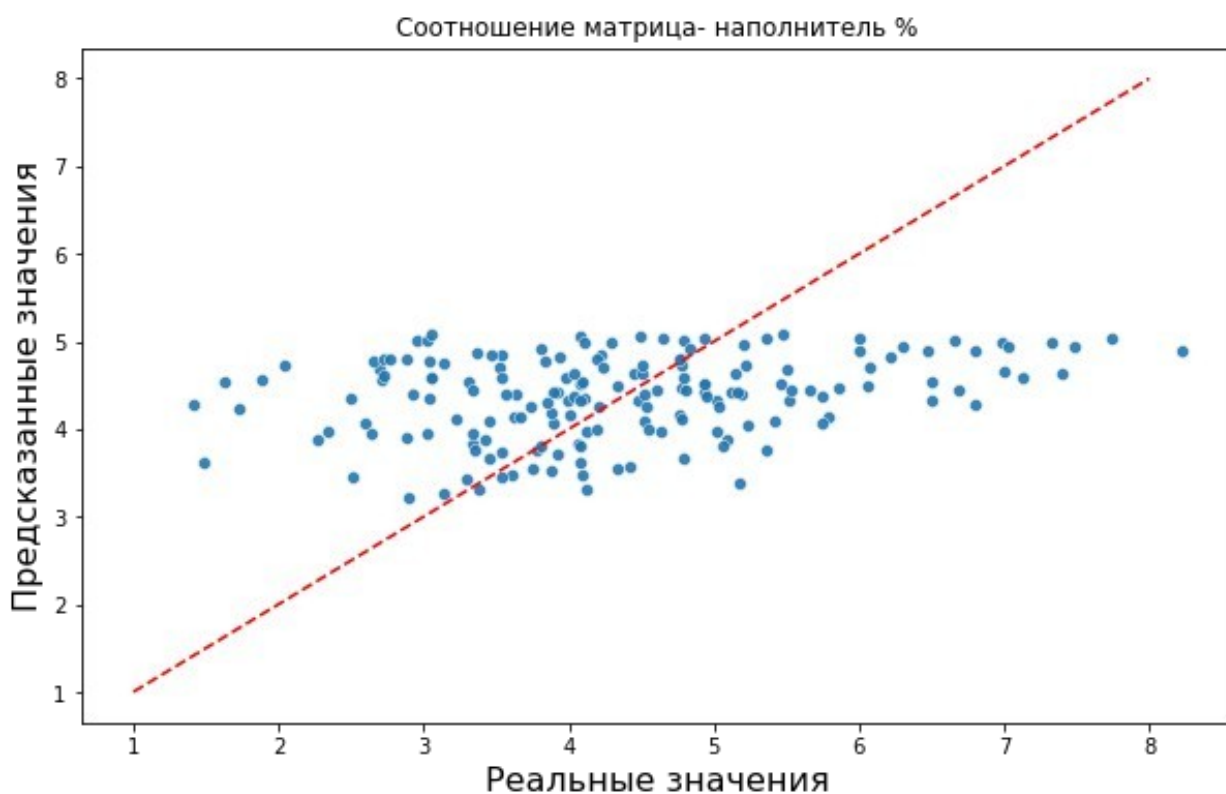


Рисунок 17 - DNN результат

А вот нейронная сеть не справилась с поставленной задачей рекомендовать соотношение матрица-наполнитель это подтверждает рисунок 17 ошибка точки распределены горизонтально.

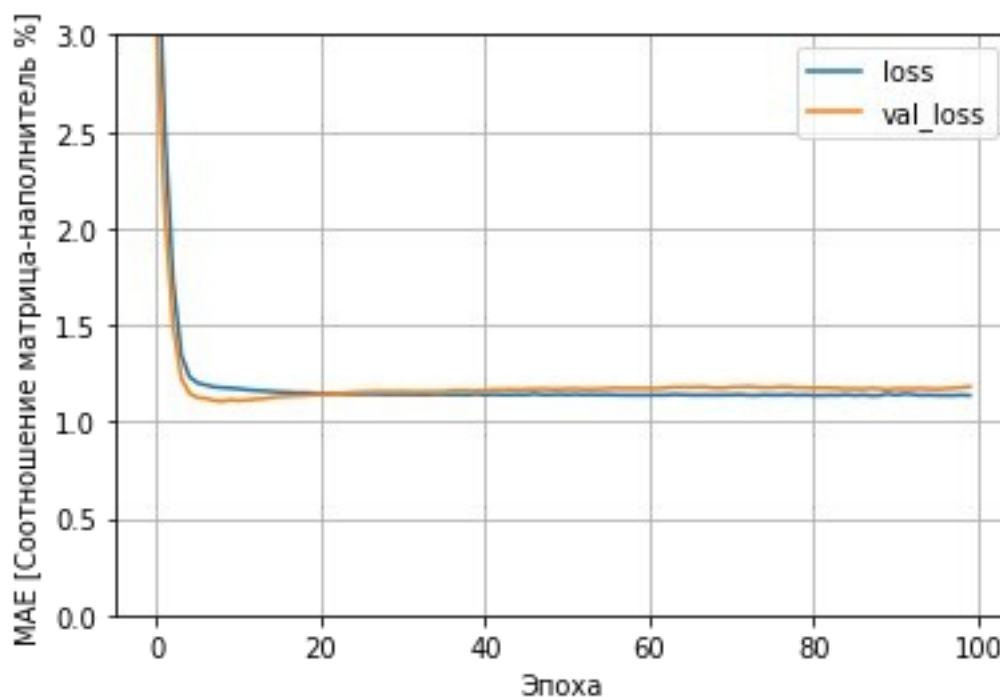


Рисунок 18 - Обучение модели

Как видно на рисунке 18 на 20 итерациях (эпохах) ошибка стремительно уменьшилась, а дальше если и уменьшалась, то совсем незначительно. Нейронная сеть DNN показала плохой результат реальные значения и прогнозируемые значения не проецируются вдоль линии.

2.4 Создание приложения

Как видно обученные регрессионные модели довольно неплохо прогнозируют «Модуль упругости при растяжении, гПа %» и очень хорошо «Прочность при растяжении %» и с хорошей точностью. А вот нейронная сеть очень плохо рекомендовала «Соотношение матрица-наполнитель %» может повлиять неправильный отбор признаков или неправильно настроенная нейронная сеть.

По имеющемуся набору данных прогнозировать «Прочность при растяжении %» разработано приложение.

Приложение в качестве аргументов принимает параметры, соответствующие входным параметрам исходных данных («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 %», «Потребление смолы, г/м2 %», «Угол нашивки, град %», «Шаг нашивки %», «Плотность нашивки %») итого 12 параметров и возвращает параметр соответствующий выходному параметру («Прочность при растяжении, Мпа %»)

2.4 Создание удаленного репозитория.

Для проекта выпускной квалификационной работы был создан отдельный репозиторий на GitHub. Ссылка на репозиторий:

https://github.com/Greceia/BMSTU_VKR_Kochetkov_I_A.git

- ✓ model - (папка с моделью нейронной сети для прогноза «Прочность при растяжении, Мпа %», нужна для приложения).
- ✓ doc - пояснительная записка.
- ✓ Презентация - презентация ВКР.
- ✓ Main_book.ipynb - основная тетрадь Jupyter Notebook с исследованиями.
- ✓ app.py - приложение с интерфейсом командной строки.
- ✓ App_book.ipynb - приложение в Jupyter
- ✓ X_br - исходный датасет.
- ✓ X_nur - исходный датасет.
- ✓ README - описание проекта.

Заключение

В ходе выполнения данной работы был выполнен пайплайн построения моделей машинного обучения. Все построенные регрессионные модели для прогнозирования «Модуль упругости, гПа %» и «Прочность при растяжении %» показали хорошие результаты. Ключевым фактором в решении поставленной задачи, сыграла разработка и трансформация признаков, которые значительно улучшили все протестированные модели. В случае с нейронной сетью для рекомендации «соотношение матрица-наполнитель %» результат нельзя назвать удовлетворительным, можно улучшить результат

4. Список используемой литературы и веб ресурсы.

1. Д. А. Иванов, А. И. Ситников, С. Д. Шляпин - «Композиционные материалы: учебное пособие для вузов» 2019.
2. А. А. Миронов. «Машинное обучение часть I.» 2019.
3. Язык программирования Python - <https://www.python.org/>.
4. Библиотека Pandas - <https://www.python.org/>.
5. Библиотека Matplotlib - <https://matplotlib.org/>.
6. Библиотека Seaborn - <https://seaborn.pydata.org/>.
7. Библиотека Sklearn - <https://scikit-learn.org/>.
8. Среда разработки Jupyter Notebook - <https://jupyter.org/>.
9. Библиотека Tensorflow - <https://www.tensorflow.org/>.
10. Чун-Те Чен и Грейс Х. Гу. - «Машинное обучение для композитных материалов» 2019.
11. Жерон, Орельен. «Прикладное машинное обучение с помощью Scikit-Learn и TensorFlow: концепции, инструменты и техники для создания интеллектуальных систем» 2018.