

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

Дорохова Ольга Васильевна

Москва, 2022

Введение	3
1. Аналитическая часть	7
1.1. Постановка задачи	7
1.2. Описание используемых методов	8
1.3. Разведочный анализ данных	10
2. Практическая часть	12
2.1. Предобработка данных	12
2.2. Обучение моделей регрессии	20
2.3. Тестирование модели	22
2.4. Нейронная сеть для рекомендации соотношения матрица-наполнитель	23
2.5. Разработка приложения	26
2.6. Создание удаленного репозитория и выгрузка результатов работы	27
Заключение	29
Список использованной литературы	30

Введение

Композитные материалы (КМ) – это «супер-вещества», которые получают соединением двух и более компонентов. Компоненты в существенной степени отличаются друг от друга по своим свойствам. Сочетание их приводит к тому, что образуются новые материалы с уникальными свойствами, отличными от исходного сырья. То есть это суперновые материалы, вобравшие в себе, все лучшее от родителей.

Композитные соединения состоят из двух основных частей. Первая – это матрица, вторая – это наполнитель. Новые композитные вещества превосходят привычно используемые материалы по прочностным, механическим характеристиками и выгодно отличаются по массе, имея легкий вес.

Для примера композитных материалов можно привести клееную фанеру.

По своей структуре композитные материалы можно разделить на несколько групп:

- волокнистые;
- дисперсноупрочненные;
- упрочненные частицами;
- нанокompозиты.

Волокнистые композиты – это материалы, улучшение свойств которых производится волокнами или нитевидными кристаллами. Например, кирпич с соломой. Незначительное введение добавки наполнителя приводит к появлению новых, уникальных свойств.

А, к примеру, добавка электропроводящих волокон придает материалу новое свойство – проводить электрический ток.

Если композитный материал имеет слоистую структуру, то в нем матрица и наполнитель располагаются слоями. Примером можно привести стекло, покрытое слоями полимерных пленок.

Другие представители композитных материалов имеют структуру, которая представлена матрицей и наполненными ее частицами, отличающимися по размерам.

- упрочненные материалы имеют 20-25% частиц, размер которых составляет более 1 мкм,
- дисперсноупрочненные – 1-15% частиц, размер которых составляет 0,01-0,1 мкм.
- нанокompозитные материалы, имеют частицы, размер которых составляет – 10-100 нм.

Выделяют несколько видов композитных материалов:

1. Полимерные композитные материалы

Полимерные композитные материалы (ПКМ), имеют в качестве своей базы полимерную основу-матрицу. Это самый многочисленный вид КМ. Их применение позволило значительно снизить вес и улучшить эксплуатационные характеристики многих вещей. Так, к примеру, применение ПКМ при изготовлении искусственного спутника земли привело к снижению его веса, а облегчение на 1 кг дает экономию в 1000 долларов.

2. Стеклопластики

Полимерные композиты, армирование которых проводят стеклянными волокнами. Данные волокна получают формированием при нагревании неорганического стекла. В качестве матрицы выступают фенольные, эпоксидные смолы или термопластичные полимеры.

Материалы характеризуются: прозрачностью к радиоволнам, прочностью и электроизоляцией, низкой теплопроводностью.

Стеклопластики – это дешевый и доступный материал, его применяют в приборостроении, судостроении, строительстве, при изготовлении спортивных товаров.

3. Углепластики

В данных композитных материалах наполнителями служат углеродсодержащие волокна, которые «добывают» из натуральных или искусственно созданных волокон.

Матрицей также служит терморезактивный или термопластичный полимер.

Преимущества: низкая плотность, высокий коэффициент упругости, они легкие по массе, но в тоже время очень прочные, хорошо проводят электрический ток.

Применяются в авио-, машино- и ракетостроении, а также при производстве космической техники, спортивных товаров, медицинских протезов.

4. Боропластики

Матрица – терморезактивный полимер.

Наполнитель – борные волокна, борные жгуты.

Борные волокна имеют большую прочность при сжатии, чем волокна других материалов. Поэтому и получаемые с их использованием материалы обладают отличными прочностными, износостойкими характеристиками, а также характеризуются инертностью к агрессивным средам. Но в тоже время, им свойственна хрупкость, что вводит определенные ограничения на использование.

Есть у данных ПКМ еще один минус – высокая цена, по данной причине область их использования достаточно узкая. Применяются главным образом только при изготовлении деталей, которые используются при высокой нагрузке и требуют высокой надежности.

5. Органопластики

Основа – преимущественно используют эпоксидные, полимерные, фенольные смолы.

Наполнитель – искусственные или натуральные волокна. Волокна используются в виде нитей или жгутов, а также «полотен» - ткани, бумаги. Объем наполнителя составляет 40-70%.

Отличаются низкой плотностью, легкостью. Имеют высокую прочность. Отличное сопротивление к ударам, а также устойчивость к динамическому воздействию, нагрузкам. Но при всем при этом их прочность при изгибе и сжатии мала.

Применяются при машино-, авио-, судостроении, космической и авиационной технике, производстве спорт инвентаря и радиоэлектроники.

6. Полимеры наполненные порошками

В наше время применяют дешевые наполнители, такие как каолин, для термопластичной/термополимерной матрицы. Данные соединения используют для изготовления электроизоляционных материалов, труб. Сажу применяют для наполнения при изготовлении резин.

7. Текстолиты

Это полимерные материалы, имеющие слоистую структуру и применяемые для изготовления, например кухонных поверхностей, в качестве армирующего элемента используют ткани.

Композиционные материалы – это улучшенные соединения, это как валенки с калошами и тепло, и ноги не промокают.

Два ключевых момента, которые нужно знать о чудо композитных материалах:

- Материалы состоят из матрицы и наполнителя,
- Материалы после «модификации» приобретает новые свойства.

Существует определенный 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).

1 Аналитическая часть

1.1 Постановка задачи.

На основании имеющихся данных необходимо спрогнозировать следующие конечные свойства получаемых композиционных материалов:

- модуль упругости при растяжении;
- прочность при растяжении;
- соотношение матрица-наполнитель.

На входе имеются два датасета с данными о свойствах компонентов композиционных материалов. Основные сведения о данных датасетах представлены в Таблице 1 и Таблице 2.

Таблица 1 – Основные сведения датасета X_bp.xlsx

#	Column	Non-Null Count	Dtype
---	-----	-----	-----
0	Соотношение матрица-наполнитель	1023 non-null	float64
1	Плотность, кг/м3	1023 non-null	float64
2	модуль упругости, ГПа	1023 non-null	float64
3	Количество отвердителя, м.%	1023 non-null	float64
4	Содержание эпоксидных групп, %_2	1023 non-null	float64
5	Температура вспышки, С_2	1023 non-null	float64
6	Поверхностная плотность, г/м2	1023 non-null	float64
7	Модуль упругости при растяжении, ГПа	1023 non-null	float64
8	Прочность при растяжении, МПа	1023 non-null	float64
9	Потребление смолы, г/м2	1023 non-null	float64

Таблица 2 – Основные сведения датасета X_nup.xlsx

#	Column	Non-Null Count	Dtype
---	-----	-----	-----
0	Угол нашивки, град	1040 non-null	float64
1	Шаг нашивки	1040 non-null	float64
2	Плотность нашивки	1040 non-null	float64

Для удобства использования датасеты объединяются по индексу тип объединения INNER с отсечением лишних строк из датасета X_nup.xlsx.

Основная информация о полученном датасете представлена в Таблице 3.

Таблица 3 – Основные сведения объединенного датасета

#	Column	Non-Null Count	Dtype
---	-----	-----	-----
0	Соотношение матрица-наполнитель	1023 non-null	float64
1	Плотность, кг/м3	1023 non-null	float64
2	модуль упругости, ГПа	1023 non-null	float64
3	Количество отвердителя, м.%	1023 non-null	float64
4	Содержание эпоксидных групп, %_2	1023 non-null	float64
5	Температура вспышки, С_2	1023 non-null	float64
6	Поверхностная плотность, г/м2	1023 non-null	float64
7	Модуль упругости при растяжении, ГПа	1023 non-null	float64
8	Прочность при растяжении, МПа	1023 non-null	float64
9	Потребление смолы, г/м2	1023 non-null	float64
10	Угол нашивки, град	1023 non-null	float64
11	Шаг нашивки	1023 non-null	float64
12	Плотность нашивки	1023 non-null	float64

Данный датасет занимает 111.9 Кб.

1.2 Описание используемых методов

Для прогнозирования модуля упругости при растяжении и прочности при растяжении используем следующие методы машинного обучения:

► Линейная регрессия (множественная)

Множественная линейная регрессия (MLR), также известная просто как множественная регрессия, – это статистический метод, который использует

несколько независимых переменных для прогнозирования результата переменной ответа. Цель множественной линейной регрессии – смоделировать линейную связь между независимыми переменными и ответной (зависимой) переменной. Множественная линейная регрессия вычисляется по формуле (1).

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \epsilon \quad (1)$$

где для i - n наблюдений:

y_i - зависимая переменная;

x_i - объясняющие переменные;

β_0 - y -интерцепт (постоянный член);

β_p - коэффициенты наклона для каждой объясняющей переменной;

ϵ - член ошибки модели (также известный как остатки).

Простая линейная регрессия – это функция, которая позволяет аналитику или статистику делать прогнозы относительно одной переменной на основе информации, которая известна о другой переменной. Линейная регрессия может быть использована только при наличии двух непрерывных переменных – независимой переменной и зависимой переменной. Независимая переменная – это параметр, который используется для расчета зависимой переменной или результата. Модель множественной регрессии распространяется на несколько объясняющих переменных.

► Случайный лес решений

Это алгоритм машинного обучения, заключающийся в использовании ансамбля решающих деревьев. Алгоритм сочетает в себе две основные идеи: метод бэггинга Бреймана, и метод случайных подпространств предложенный Тин Кам Хо. Алгоритм применяется для задач классификации, регрессии и кластеризации.

Основная идея заключается в использовании большого ансамбля решающих деревьев, каждое из которых само по себе даёт очень невысокое

качество классификации, но за счёт их большого количества результат получается хорошим.

Сама структура дерева представляет собой «листья» и «ветки». На рёбрах («ветках») дерева решения записаны признаки, от которых зависит целевая функция, в «листьях» записаны значения целевой функции, а в остальных узлах — признаки, по которым различаются случаи. Чтобы классифицировать новый случай, надо спуститься по дереву до листа и выдать соответствующее значение.

Подобные деревья решений широко используются в интеллектуальном анализе данных. Цель состоит в том, чтобы создать модель, которая предсказывает значение целевой переменной на основе нескольких переменных на входе.

Каждый лист представляет собой значение целевой переменной, изменённой в ходе движения от корня по рёбрам дерева до листа. Каждый внутренний узел сопоставляется с одной из входных переменных. Дерево может быть также «изучено» разделением исходных наборов переменных на подмножества, основанные на проверке значений признаков. Это действие повторяется на каждом из полученных подмножеств. Рекурсия завершается тогда, когда подмножество в узле имеет те же значения целевой переменной, таким образом, оно не добавляет ценности для предсказаний.

Для прогнозирования соотношения матрица-наполнитель воспользуемся построением нейронной сети.

1.3 Разведочный анализ данных

`.info()` Показывает количество признаков и количество показателей по ним, их тип данные, размер занимаемой памяти.

`.describe()` оценивает столбцы с числовым типом данных (int или float) и возвращает статистические данные, которые дают представление о распределении значений.

.isnull и .notnull – проверка на пропуски. При создании датасета из файла многие пустые столбцы импортируются как нулевое значение, что впоследствии создает проблемы при работе с данными. isnull() и notnull() используются для проверки и управления значениями NULL в датасете.

Гистограммы распределения показывают распределение переменной. Каждый столбик — частотность в определенном интервале. Применяются для визуализации, как часто какие значения встречаются в наборе данных.

Диаграммы рассеяния (также точечная диаграмма, англ. scatter plot) — математическая диаграмма, изображающая значения двух переменных в виде точек на декартовой плоскости. Могут использоваться и полярные координаты, особенно в случаях, когда одна из переменных представляет собой физическое направление или имеет циклический характер.

На диаграмме рассеяния каждому наблюдению (или элементарной единице набора данных) соответствует точка, координаты которой равны значениям двух каких-то параметров этого наблюдения. Если предполагается, что один из параметров зависит от другого, то обычно значения независимого параметра откладывается по горизонтальной оси, а значения зависимого — по вертикальной. Диаграммы рассеяния используются для демонстрации наличия или отсутствия корреляции между двумя переменными.

Ящички с усами – это такой вид диаграммы, который в удобной форме показывает медиану (или, если нужно, среднее), нижний и верхний квартили, минимальное и максимальное значение выборки и выбросы.

Корреляция - важнейший фактор, лежащий в основе анализа данных. Он сообщает, как переменные в наборе данных связаны друг с другом и как они перемещаются относительно друг друга. Значение корреляции колеблется от -1 до +1. 0 Корреляция означает, что две переменные не зависят друг от друга. Положительная корреляция указывает на то, что переменные движутся в одном направлении, а отрицательная корреляция указывает на противоположное.

Тепловая карта используется для графического представления значений матрицы с разными цветовыми оттенками для разных значений. Он очень четко визуализирует общую матрицу.

2 Практическая часть

2.1 Предобработка данных

Проверку на пропуски осуществляем с помощью `.isnull()`.

Соотношение матрица-наполнитель	0
Плотность, кг/м3	0
модуль упругости, ГПа	0
Количество отвердителя, м. %	0
Содержание эпоксидных групп, %_2	0
Температура вспышки, С_2	0
Поверхностная плотность, г/м2	0
Модуль упругости при растяжении, ГПа	0
Прочность при растяжении, МПа	0
Потребление смолы, г/м2	0
Угол нашивки, град	0
Шаг нашивки	0
Плотность нашивки	0
dtype: int64	

Пропусков нет.

Убеждаемся, что все ячейки в датасете заполнены.

Соотношение матрица-наполнитель	1023
Плотность, кг/м3	1023
модуль упругости, ГПа	1023
Количество отвердителя, м. %	1023
Содержание эпоксидных групп, %_2	1023
Температура вспышки, С_2	1023
Поверхностная плотность, г/м2	1023
Модуль упругости при растяжении, ГПа	1023
Прочность при растяжении, МПа	1023
Потребление смолы, г/м2	1023
Угол нашивки, град	1023
Шаг нашивки	1023
Плотность нашивки	1023
dtype: int64	

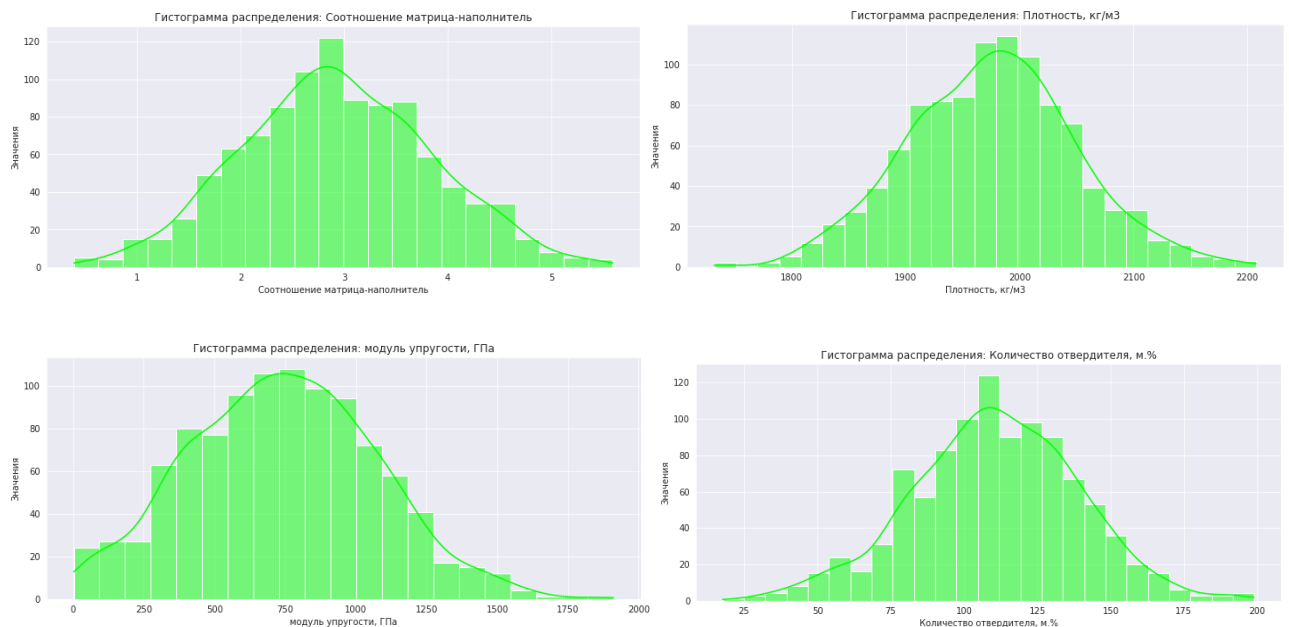
Общие сведения нашего датасета представлены в Таблице 4.

Таблица 4 - Общие сведения

	count	mean	std	min	25%	50%	75%	max
Соотношение матрица-наполнитель	1023.0	2.930366	0.913222	0.389403	2.317887	2.906878	3.552660	5.591742
Плотность, кг/м3	1023.0	1975.734888	73.729231	1731.764635	1924.155467	1977.621657	2021.374375	2207.773481
модуль упругости, ГПа	1023.0	739.923233	330.231581	2.436909	500.047452	739.664328	961.812526	1911.536477
Количество отвердителя, м.%	1023.0	110.570769	28.295911	17.740275	92.443497	110.564840	129.730366	198.953207
Содержание эпоксидных групп, % 2	1023.0	22.244390	2.406301	14.254985	20.608034	22.230744	23.961934	33.000000
Температура вспышки, С_2	1023.0	285.882151	40.943260	100.000000	259.066528	285.896812	313.002106	413.273418
Поверхностная плотность, г/м2	1023.0	482.731833	281.314690	0.603740	266.816645	451.864365	693.225017	1399.542362
Модуль упругости при растяжении, ГПа	1023.0	73.328571	3.118983	64.054061	71.245018	73.268805	75.356612	82.682051
Прочность при растяжении, МПа	1023.0	2466.922843	485.628006	1036.856605	2135.850448	2459.524526	2767.193119	3848.436732
Потребление смолы, г/м2	1023.0	218.423144	59.735931	33.803026	179.627520	219.198882	257.481724	414.590628
Угол нашивки, град	1023.0	44.252199	45.015793	0.000000	0.000000	0.000000	90.000000	90.000000
Шаг нашивки	1023.0	6.899222	2.563467	0.000000	5.080033	6.916144	8.586293	14.440522
Плотность нашивки	1023.0	57.153929	12.350969	0.000000	49.799212	57.341920	64.944961	103.988901

Выявляем дискретную величину «Угол нашивки, град».

Строим гистограммы распределения (Рисунок 1).



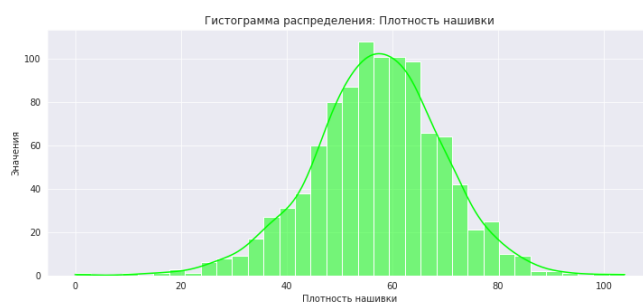
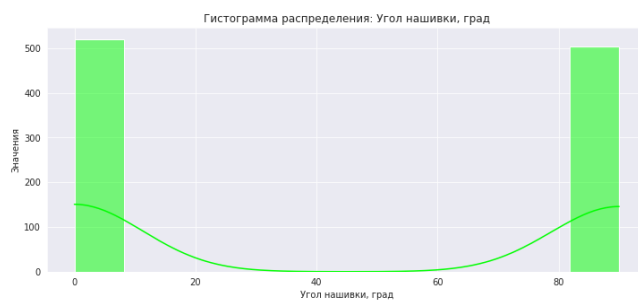
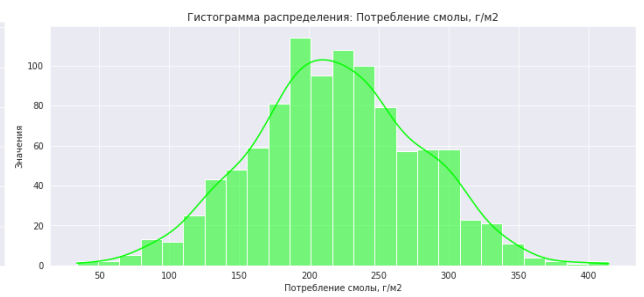
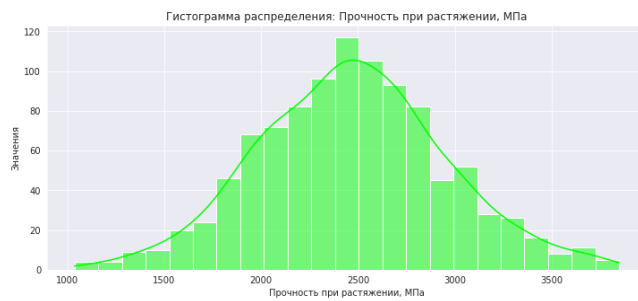
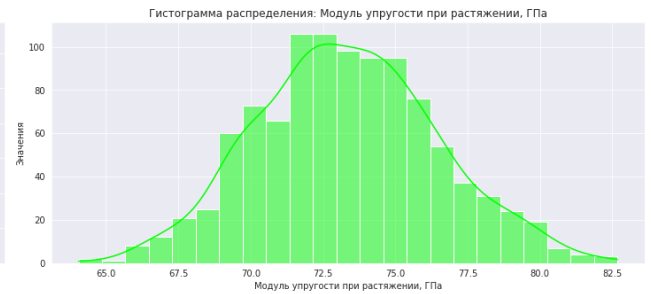
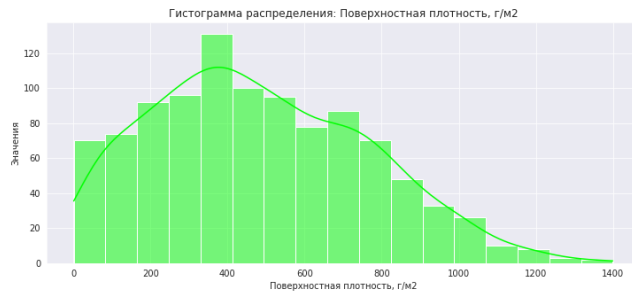
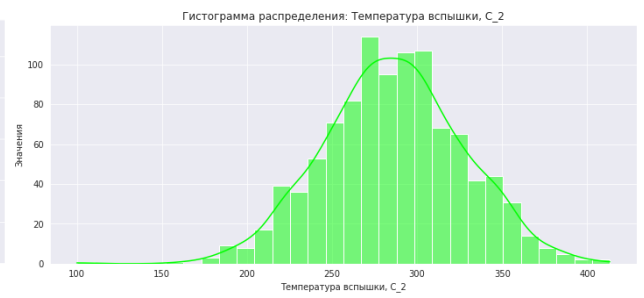
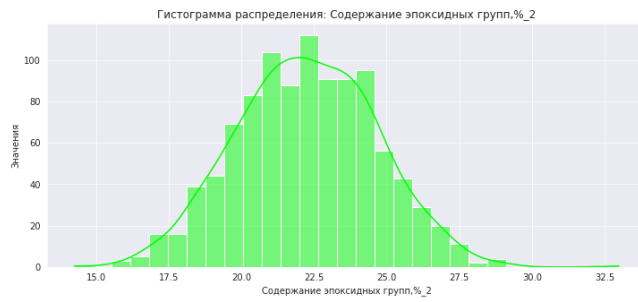


Рисунок 1 – Гистограммы распределения

Здесь уже наглядно убеждаемся в наличии дискретной величины «Угол нашивки, град». В основном все величины имеют нормальное распределение. Величина «Поверхностная плотность, г/м²» имеет распределение со смещением вправо.

Строим матрицу рассеивания (Рисунок 2), которая указывает нам об отсутствии линейной корреляции.

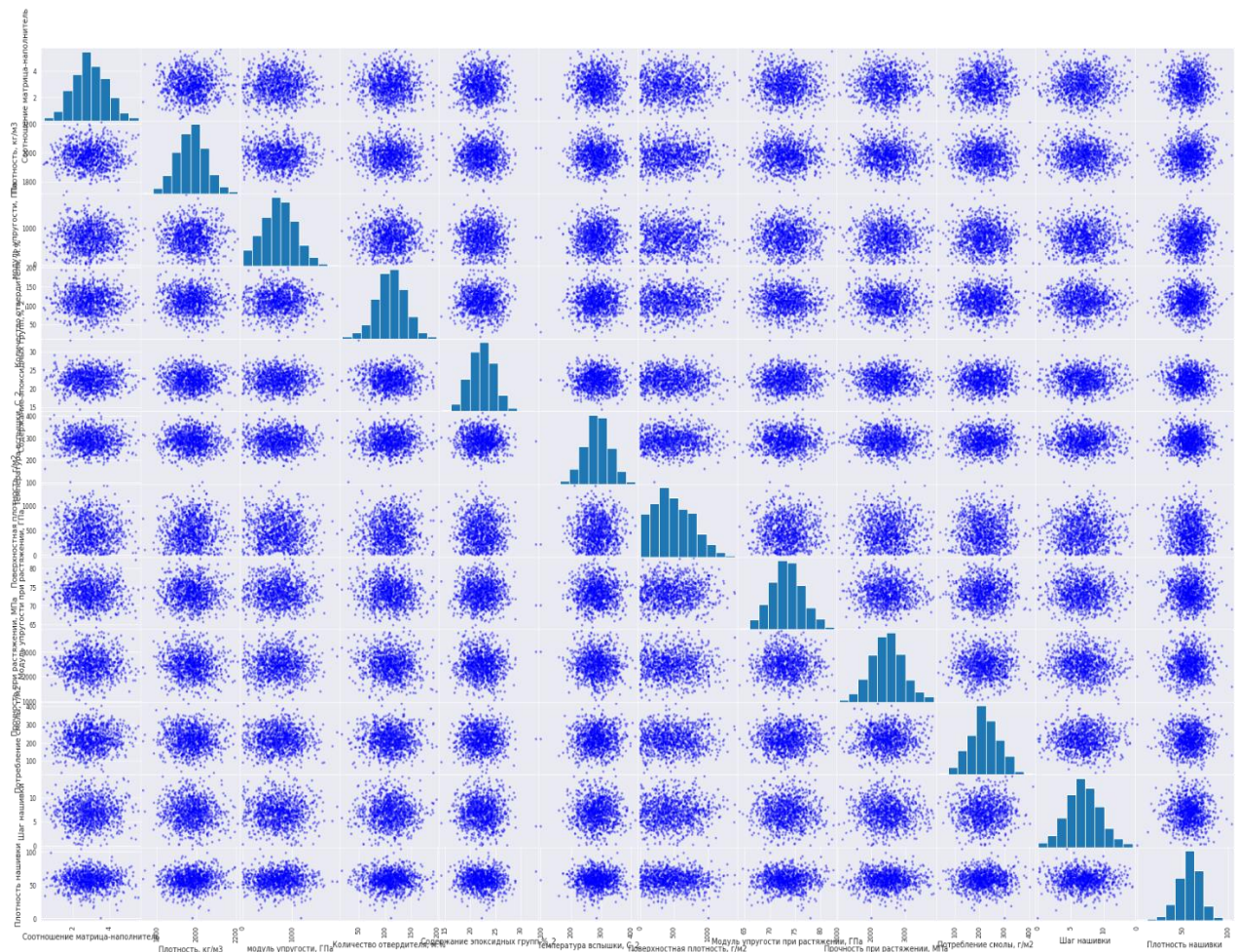


Рисунок 2 – Матричная диаграмма распределения

Тепловая карта корреляции (Рисунок 3) подтверждает отсутствие корреляции между признаками.

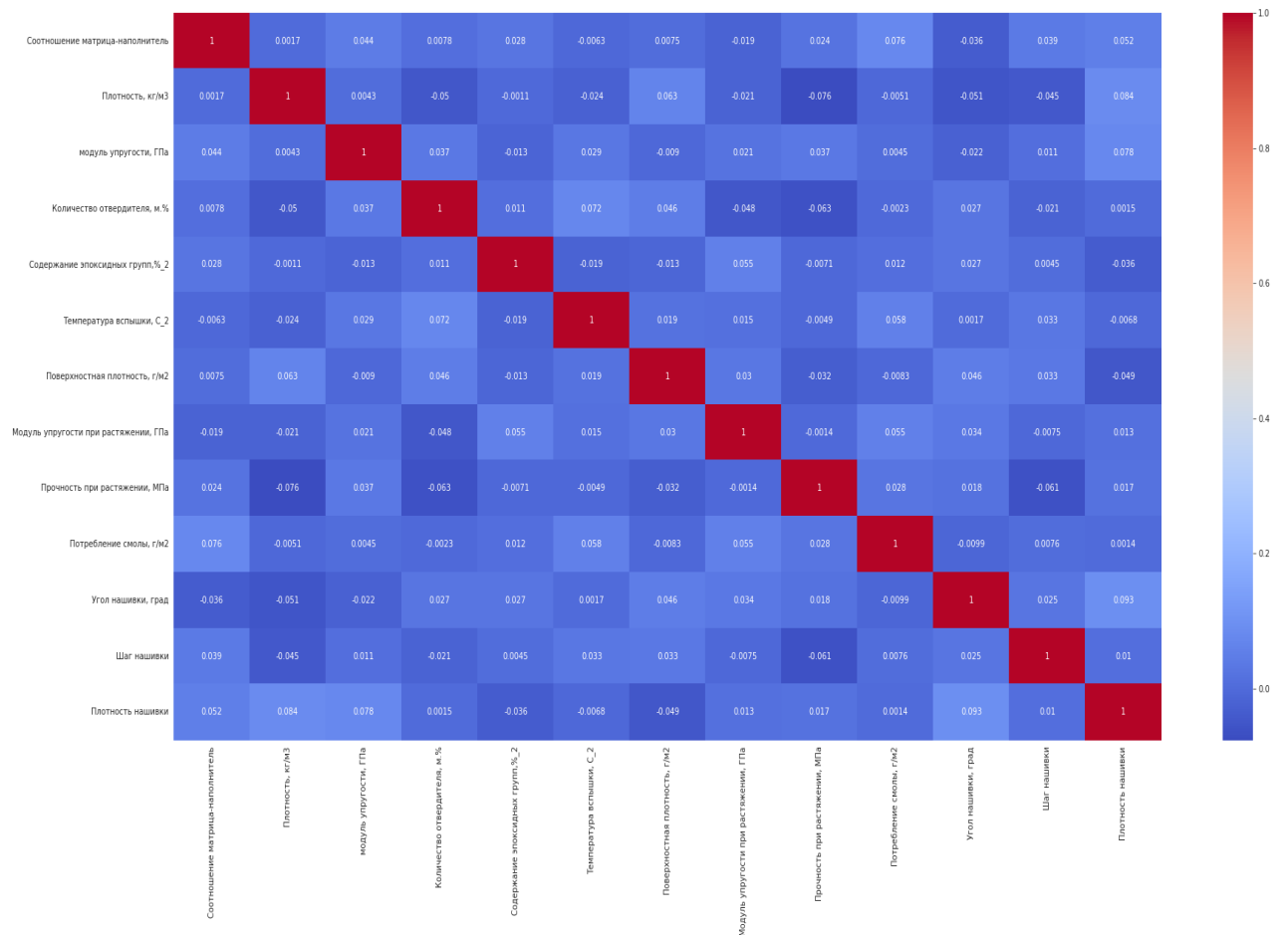


Рисунок 3 – Тепловая карта корреляции

Строим ящики с усами (Рисунок 4) для всех признаков.

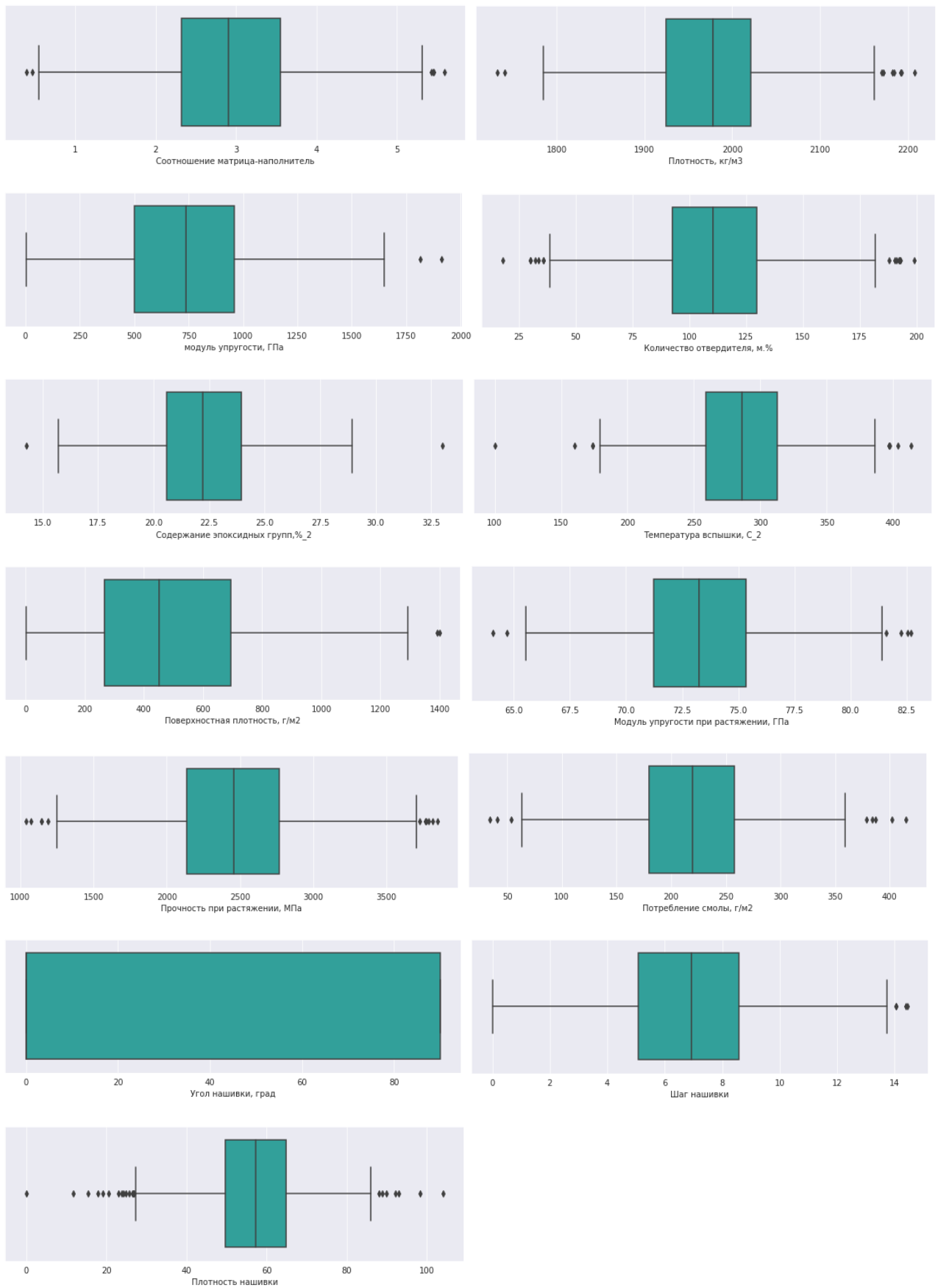


Рисунок 4 – Ящики с усами

Видим, что есть выбросы, от которых стоит избавиться

Избавляемся от выбросов методом межквартильного расстояния.

```
for p in data_nule:
    q75, q25 = np.percentile(data_nule.loc[:, p], [75, 25])
    iqr = q75 - q25
    max = q75 + (1.5 * iqr)
    min = q25 - (1.5 * iqr)
    data_nule.loc[data_nule[p] < min, p] = np.nan
    data_nule.loc[data_nule[p] > max, p] = np.nan
```

В результате избавляемся от 87 выбросов.

Соотношение матрица-наполнитель	6
Плотность, кг/м3	9
модуль упругости, ГПа	2
Количество отвердителя, м.%	14
Содержание эпоксидных групп, %_2	2
Температура вспышки, С_2	8
Поверхностная плотность, г/м2	2
Модуль упругости при растяжении, ГПа	6
Прочность при растяжении, МПа	11
Потребление смолы, г/м2	8
Угол нашивки, град	0
Шаг нашивки	4
Плотность нашивки	21

Проведем нормализацию данных с помощью метода MinMaxScaler() из библиотеки sklearn. После чего все значения признаков будут принадлежать диапазону от 0 до 1.

Основные параметры датасета после нормализации представлены в Таблице 5.

Таблица 5 – Общие сведения после нормализации

	count	mean	std	min	25%	50%	75%	max
Соотношение матрица-наполнитель	936.0	0.498933	0.187489	0.0	0.372274	0.494538	0.629204	1.0
Плотность, кг/м3	936.0	0.502695	0.187779	0.0	0.368517	0.511229	0.624999	1.0
модуль упругости, ГПа	936.0	0.446764	0.199583	0.0	0.301243	0.447061	0.580446	1.0
Количество отвердителя, м.%	936.0	0.504664	0.188865	0.0	0.376190	0.506040	0.637978	1.0
Содержание эпоксидных групп,%_2	936.0	0.491216	0.180620	0.0	0.367716	0.489382	0.623410	1.0
Температура вспышки, С_2	936.0	0.516059	0.190624	0.0	0.386128	0.515980	0.646450	1.0
Поверхностная плотность, г/м2	936.0	0.373733	0.217078	0.0	0.205619	0.354161	0.538683	1.0
Модуль упругости при растяжении, ГПа	936.0	0.488647	0.191466	0.0	0.359024	0.485754	0.615077	1.0
Прочность при растяжении, МПа	936.0	0.495706	0.188915	0.0	0.365149	0.491825	0.612874	1.0
Потребление смолы, г/м2	936.0	0.521141	0.195781	0.0	0.392067	0.523766	0.652447	1.0
Угол нашивки, град	936.0	0.511752	0.500129	0.0	0.000000	1.000000	1.000000	1.0
Шаг нашивки	936.0	0.502232	0.183258	0.0	0.372211	0.504258	0.624604	1.0
Плотность нашивки	936.0	0.513776	0.191342	0.0	0.390482	0.516029	0.638842	1.0

Посмотрим на ящики с усами (Рисунок 5) после нормализации.

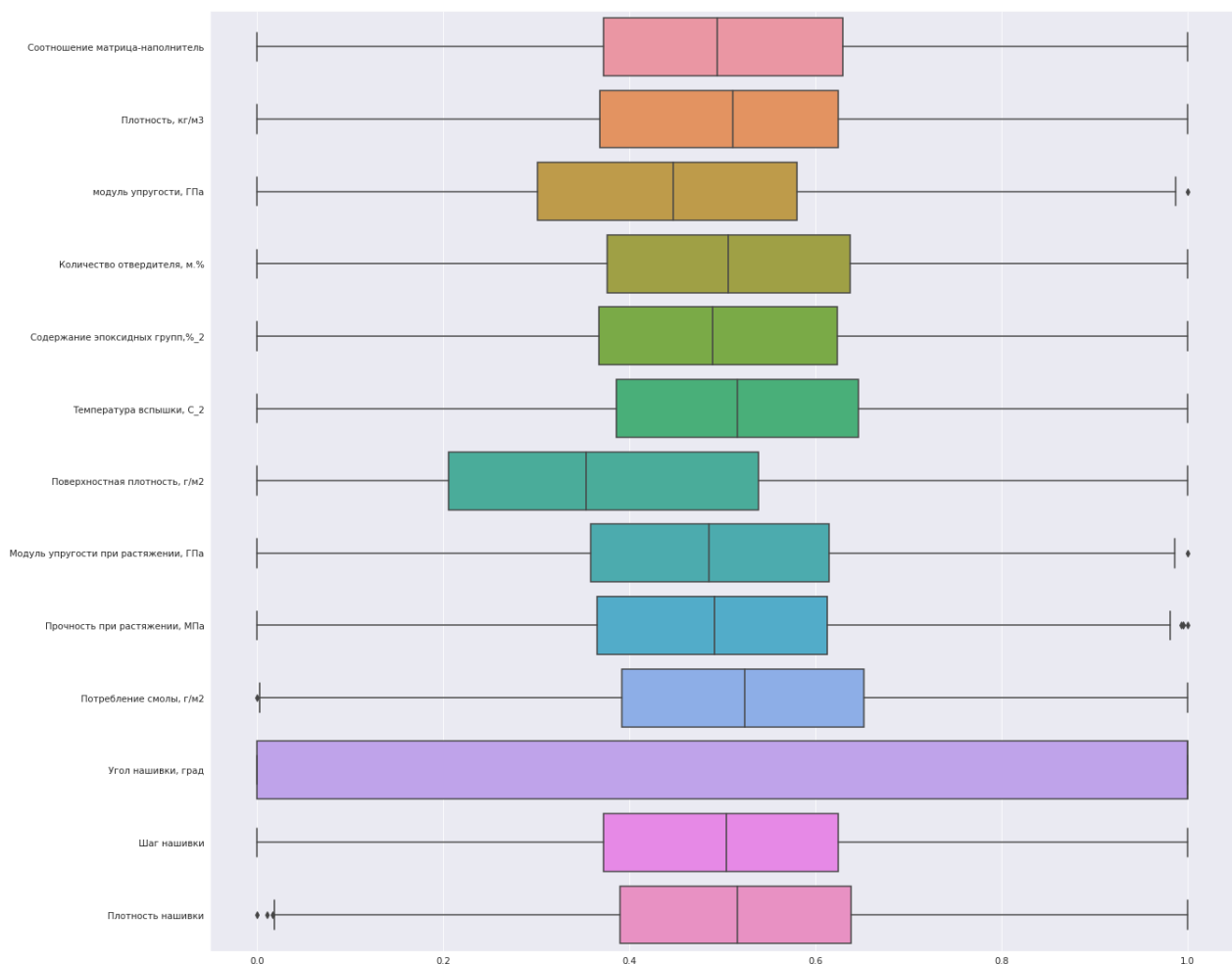


Рисунок 5 – Ящики с усами после нормализации

2.2 Разработка и обучение модели

Построение модели для модуля упругости при растяжении.

Линейная модель

```
from sklearn.linear_model import LinearRegression

model_LRypr = LinearRegression().fit(X_tr_ypr, y_tr_ypr)
y_pred_LRypr = model_LRypr.predict(X_test_ypr)
R2_LRypr = model_LRypr.score(X_test_ypr, y_test_ypr)
MSE_LRypr = mean_squared_error(y_test_ypr, y_pred_LRypr)
MAE_LRypr = mean_absolute_error(y_test_ypr, y_pred_LRypr)
print(f'R2: {R2_LRypr}\nMSE: {MSE_LRypr}\nMAE: {MAE_LRypr}')
```

Метрики

```
R2: -0.015054283902918408
MSE: 0.037862780315206186
MAE: 0.15856359645664184
```

RandomForestRegressor

```
from sklearn.ensemble import RandomForestRegressor

model_RFRypr = RandomForestRegressor().fit(X_tr_ypr, y_tr_ypr)
y_pred_RFRypr = model_RFRypr.predict(X_test_ypr)
R2_RFRypr = model_RFRypr.score(X_test_ypr, y_test_ypr)
MSE_RFRypr = mean_squared_error(y_test_ypr, y_pred_RFRypr)
MAE_RFRypr = mean_absolute_error(y_test_ypr, y_pred_RFRypr)
print(f'R2: {R2_RFRypr}\nMSE: {MSE_RFRypr}\nMAE: {MAE_RFRypr}')
```

Метрики

```
R2: -0.04559121165283897
MSE: 0.03900184549155408
MAE: 0.16102249841963617
```

RandomForestRegressor + GSCV

```
from sklearn.model_selection import GridSearchCV

param_grid = {'n_estimators': [30, 40, 50], 'max_features': [1, 'sqrt', 'log2'], 'min_samples_leaf': [10, 20], 'min_samples_split': [2, 4]}
model_RFypr = RandomForestRegressor()
```

```
GSCVypr = GridSearchCV(model_RFypr, param_grid, cv=10)
GSCVypr.fit(X_tr_ypr, y_tr_ypr)
y_pred_GSCVypr = GSCVypr.predict(X_test_ypr)
R2_GSCVypr = GSCVypr.score(X_test_ypr, y_test_ypr)
MSE_GSCVypr = mean_squared_error(y_test_ypr, y_pred_GSCVypr)
MAE_GSCVypr = mean_absolute_error(y_test_ypr, y_pred_GSCVypr)
print(f'R2: {R2_GSCVypr}\nMSE: {MSE_GSCVypr}\nMAE: {MAE_GSCVypr}')
```

Метрики

```
R2: -0.019642106547438143
MSE: 0.038033911774547124
MAE: 0.15803636806677887
```

Лучшее решение:

```
{'max_features': 1,
 'min_samples_leaf': 10,
 'min_samples_split': 4,
 'n_estimators': 40}
```

Построение модели для прочности при растяжении

Линейная модель

```
model_LRpr = LinearRegression().fit(X_tr_pr, y_tr_pr)
y_pred_LRpr = model_LRpr.predict(X_test_pr)
R2_LRpr = model_LRpr.score(X_test_pr, y_test_pr)
MSE_LRpr = mean_squared_error(y_test_pr, y_pred_LRpr)
MAE_LRpr = mean_absolute_error(y_test_pr, y_pred_LRpr)
print(f'R2: {R2_LRpr}\nMSE: {MSE_LRpr}\nMAE: {MAE_LRpr}')
```

Метрики

```
R2: 0.010451043105166313
MSE: 0.035203482767514595
MAE: 0.14757801797050743
```

RandomForestRegressor

```
model_RFRpr = RandomForestRegressor().fit(X_tr_pr, y_tr_pr)
y_pred_RFRpr = model_RFRpr.predict(X_test_pr)
R2_RFRpr = model_RFRpr.score(X_test_pr, y_test_pr)
MSE_RFRpr = mean_squared_error(y_test_pr, y_pred_RFRpr)
MAE_RFRpr = mean_absolute_error(y_test_pr, y_pred_RFRpr)
print(f'R2: {R2_RFRpr}\nMSE: {MSE_RFRpr}\nMAE: {MAE_RFRpr}')
```

Метрики

R2: -0.024529043370392856
MSE: 0.03644790919317894
MAE: 0.15056755206683722

RandomForestRegressor + GSCV

```
param_grid = {'n_estimators': [30, 40, 50], 'max_features': [1, 'sqrt', 'log2'], 'min_samples_leaf': [10, 20], 'min_samples_split': [2, 4]}  
model_RFpr = RandomForestRegressor()  
GSCVpr = GridSearchCV(model_RFpr, param_grid, cv=10)  
GSCVpr.fit(X_tr_pr, y_tr_pr)  
y_pred_GSCVpr = GSCVpr.predict(X_test_pr)  
R2_GSCVpr = GSCVpr.score(X_test_pr, y_test_pr)  
MSE_GSCVpr = mean_squared_error(y_test_pr, y_pred_GSCVpr)  
MAE_GSCVpr = mean_absolute_error(y_test_pr, y_pred_GSCVpr)  
print(f'R2: {R2_GSCVpr}\nMSE: {MSE_GSCVpr}\nMAE: {MAE_GSCVpr}')
```

Метрики

R2: 0.0008553681446396011
MSE: 0.03554485160607674
MAE: 0.14862216023731364

Лучшее решение

```
{'max_features': 1,  
 'min_samples_leaf': 10,  
 'min_samples_split': 4,  
 'n_estimators': 30}
```

2.3 Тестирование модели

Сравнительный анализ моделей по метрикам представлен в Таблице 6.

Таблица 6 – Сравнительный анализ моделей

Модель регрессии	R2	MSE	MAE
Линейная регрессия МУ	-0.015054	0.037863	0.158564
Случайный лес МУ	-0.045591	0.039002	0.161022
Случайный лес + GSCV МУ	-0.019642	0.038034	0.158036
Линейная регрессия ПР	0.010451	0.035203	0.147578
Случайный лес ПР	-0.024529	0.036448	0.150568
Случайный лес + GSCV ПР	0.000855	0.035545	0.148622

2.4 Нейронная сеть для рекомендации соотношения матрица-наполнитель

Построим три нейронные сети для рекомендации соотношения матрица-наполнитель

Первая с одним слоем

```
linear_model = tf.keras.Sequential([
    normalizer,
    layers.Dense(units=1)
])
linear_model.compile(
    optimizer=tf.optimizers.Adam(learning_rate=0.1),
    loss='mean_absolute_error')
```

Архитектура модели:

Model: "sequential"

Layer (type)	Output Shape	Param #
normalization (Normalization)	(None, 12)	25
dense (Dense)	(None, 1)	13

=====
Total params: 38
Trainable params: 13
Non-trainable params: 25

Посмотрим график уменьшения ошибок (Рисунок 6)

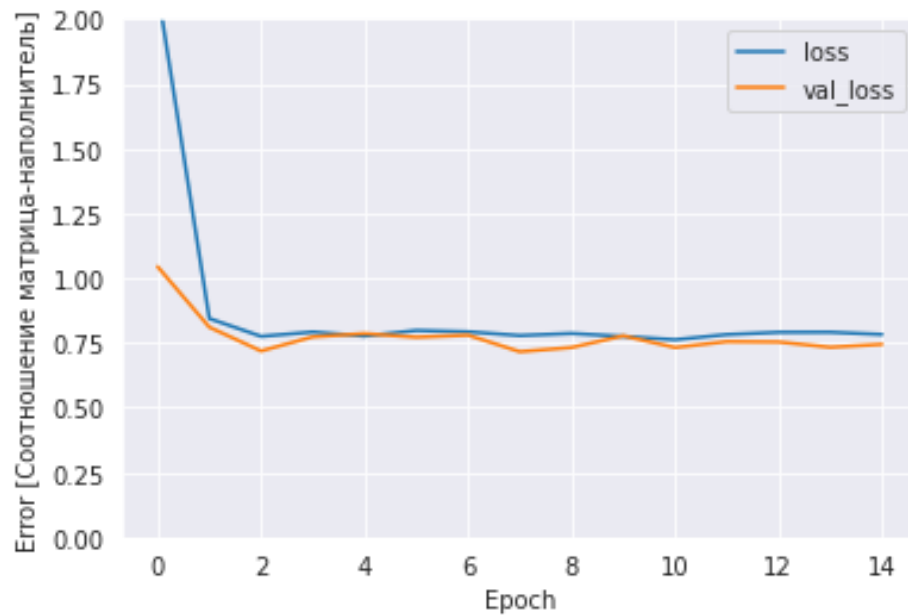


Рисунок 6 _График уменьшения ошибки

Вторая модель.

```
DNN_model = tf.keras.Sequential([
    normalizer,
    layers.Dense(18, activation='relu'),
    layers.Dense(1)
])
DNN_model.compile(loss='mean_absolute_error',
                  optimizer=tf.keras.optimizers.Adam(0.001))
```

Архитектура модели:

Model: "sequential_1"

Layer (type)	Output Shape	Param #
normalization (Normalizatio n)	(None, 12)	25
dense_1 (Dense)	(None, 32)	416
dense_2 (Dense)	(None, 1)	33

```
=====
Total params: 474
Trainable params: 449
Non-trainable params: 25
```


Посмотрим график уменьшения ошибок этой модели (Рисунок 7)

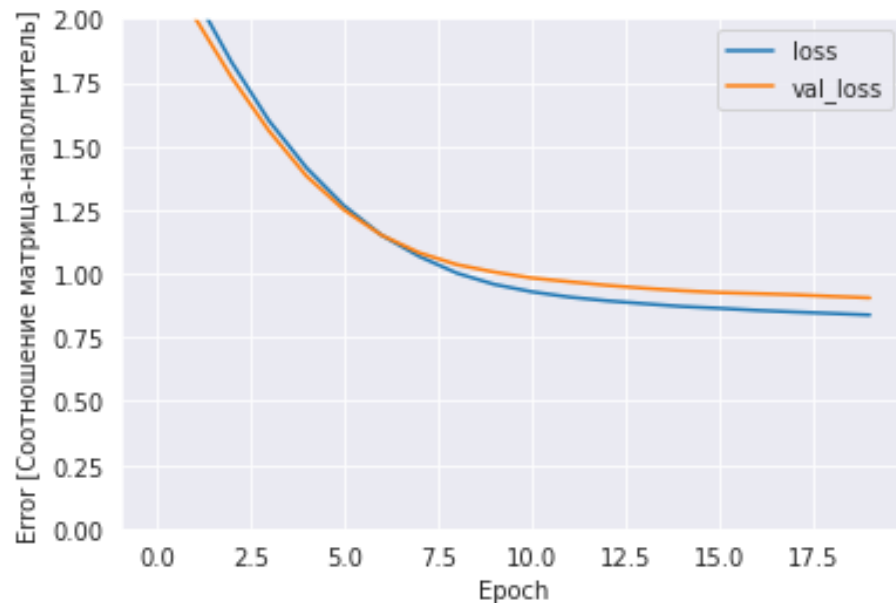


Рисунок 7 _График уменьшения ошибки

Третья модель

```
DNNN_model = tf.keras.Sequential([
    normalizer,
    layers.Dense(18, activation='relu'),
    layers.Dense(18, activation='relu'),
    layers.Dense(1)
])
DNNN_model.compile(loss='mean_absolute_error',
                    optimizer=tf.keras.optimizers.Adam(0.001))
```

Архитектура модели:

Model: "sequential_2"

Layer (type)	Output Shape	Param #
normalization (Normalizatio n)	(None, 12)	25
dense_3 (Dense)	(None, 32)	416
dense_4 (Dense)	(None, 32)	1056
dense_5 (Dense)	(None, 1)	33

=====

Total params: 1,530
 Trainable params: 1,505
 Non-trainable params: 25

Посмотрим график уменьшения ошибок этой модели (Рисунок 8)

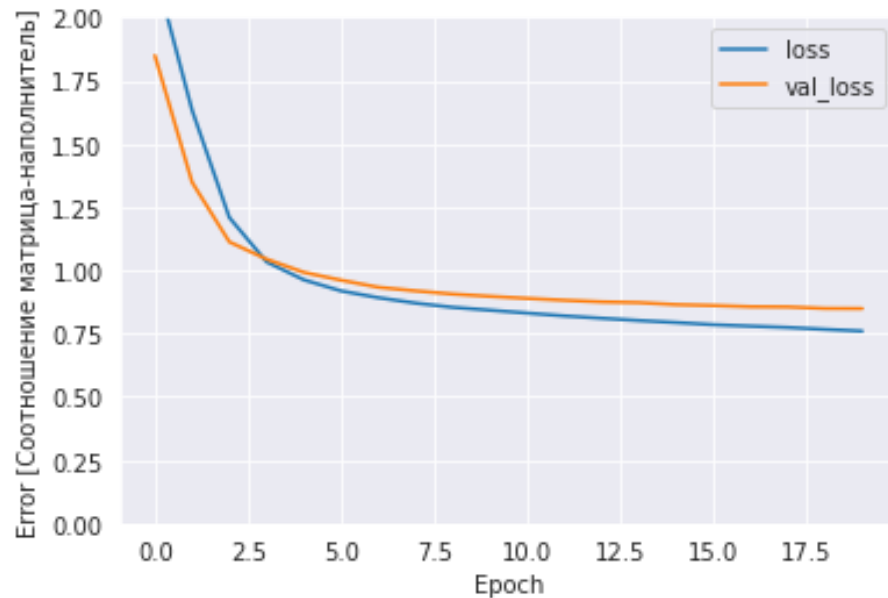


Рисунок 8 _График уменьшения ошибки

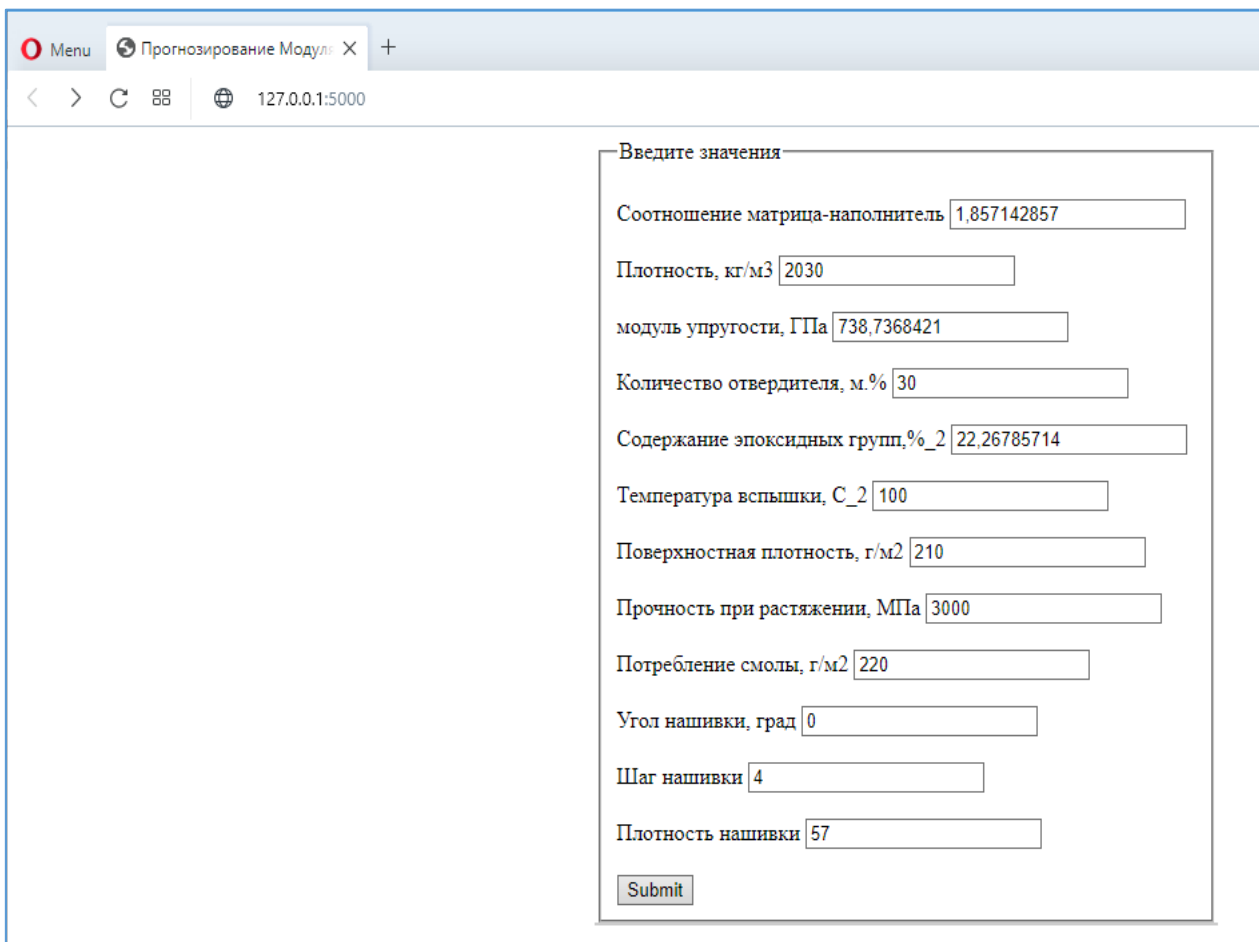
Сравнительный анализ моделей по метрикам представлен в Таблице 7

Таблица 7 – Сравнительный анализ моделей

	Mean absolute error [Соотношение матрица-наполнитель]
linear_model	0.7213618755340576
DNN_model	0.8182805776596069
DNNN_model	0.8055660128593445

2.5 Разработка приложения

Приложение было разработано для предсказания модуля упругости при растяжении с помощью FLASK. Скриншот интерфейса приложения можно посмотреть на рисунке 9.



Введите значения

Соотношение матрица-наполнитель

Плотность, кг/м3

модуль упругости, ГПа

Количество отвердителя, м.%

Содержание эпоксидных групп, %_2

Температура вспышки, С_2

Поверхностная плотность, г/м2

Прочность при растяжении, МПа

Потребление смолы, г/м2

Угол нашивки, град

Шаг нашивки

Плотность нашивки

Рисунок 9 – Интерфейс приложения

Код приложения расположен в репозитории на GitHub.

2.6 Создание удаленного репозитория и загрузка результатов работы на него.

Скриншот страницы удаленного репозитория на GitHub представлен на Рисунке 10.

Ссылка на репозиторий: <https://github.com/OlgaD2021/Zadacha>

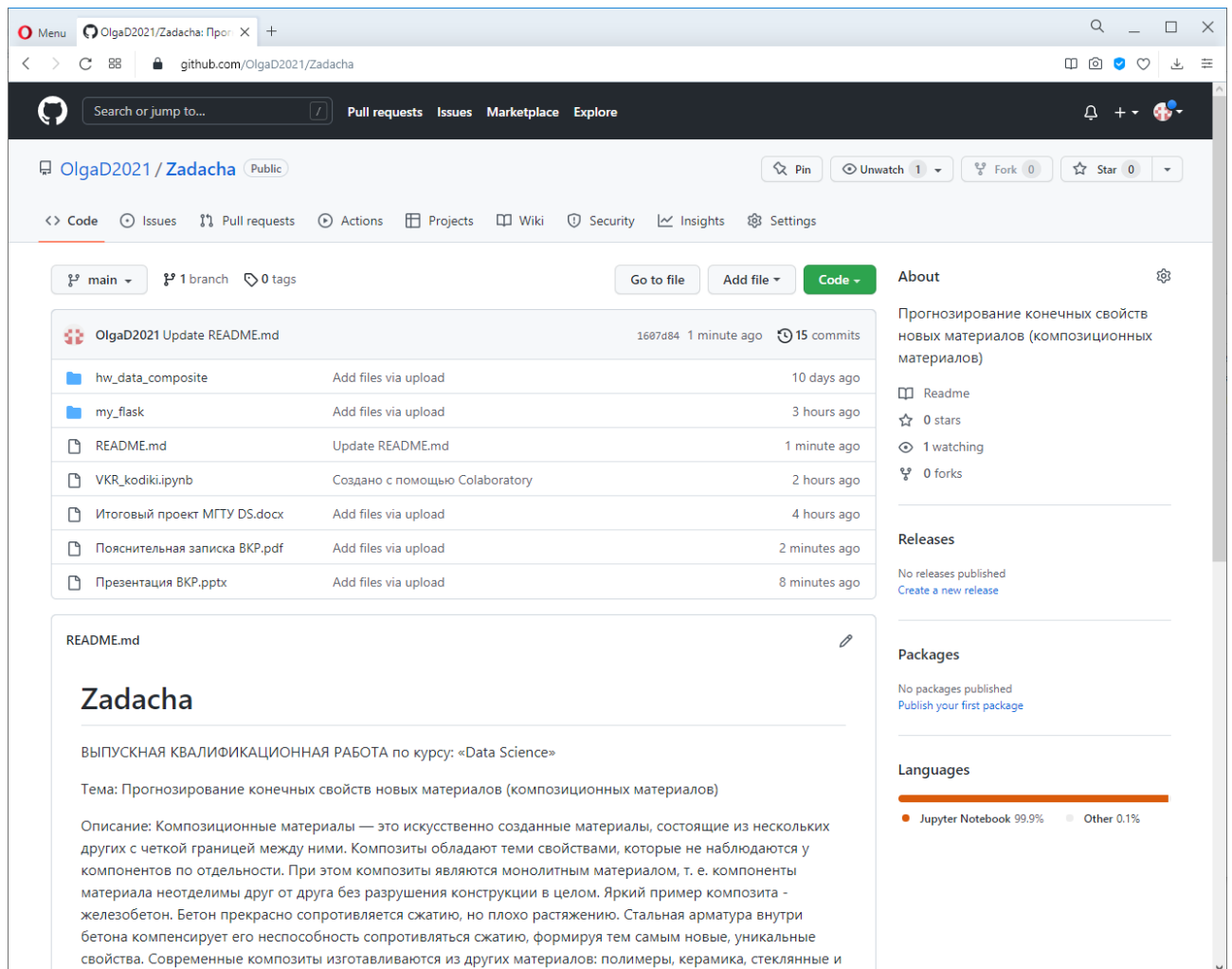


Рисунок 10 – Страница проекта на GitHub

Структура удаленного репозитория на GitHub:

- **hw_data_composite** - папка с входными файлами (X_bp.xlsx, X_nup.xlsx);
- **my_flask** - папка с файлами для приложения;
- **Итоговый проект МГТУ DS требования.docx** - задачи и требования к оформлению работы;
- **VKR_kodiki.ipynb** - код работы;
- **Презентация ВКР.pptx** - презентация ВКР;
- **Пояснительная записка ВКР.pdf** - описание работы.

Заключение

После проведения анализа входных данных и построения моделей машинного обучения можно сделать вывод, что подобранные модели плохо работают с предоставленными данными. Для построения лучшей модели машинного обучения необходимо предоставление дополнительных измерительных показателей, глубокое изучение теоретической и практической темы в области композиционных материалов, пересмотр методов анализа предоставленных данных.

Список использованной литературы:

1. Полимерные композиционные материалы на основе волокон различной химической природы / Ю.А. Кадыкова, А.Н. Леонтьев, О.Г. Васильева, С.Е. Артеменко // Строительные материалы, оборудование, технологии XXI века. 2002. № 6. С. 10-11. 2
2. Сравнительные характеристики базальто-, стекло- и углепластиков, сформованных методом поликонденсационного наполнения / Ю.А. Кадыкова, О.Г. Васильева, С.Е. Артеменко, А.Н. Леонтьев // Пластические массы. 2003. № 5. С. 37-38
3. В.А.Головко. От многослойных персептронов к нейронным сетям глубокого доверия: парадигмы обучения и применение. В сб.: Нейроинформатика-2015. XVII Всероссийская научно-техническая конференция с международным участием. Лекции по нейроинформатике, с.47-84. НИЯУ МИФИ, 2015
4. Перевод: А.Мюллер, С.Гвидо. Введение в машинное обучение с помощью Python. Руководство для специалистов по работе с данными. Вильямс, 2017, 480 с.
5. Перевод: О.Жерон. Прикладное машинное обучение с помощью Scikit-Learn и TensorFlow. Концепции, инструменты и техники для создания интеллектуальных систем. Вильямс, 2018, 688 с.
6. Виды и применение композитных материалов. Режим доступа: <https://epolymer.ru/article/vidy-i-primeneniye-kompozitnyh-materialov>. (Дата обращения: 15.06.2022).
7. Random_forest. Режим доступа: https://ru.wikipedia.org/wiki/Random_forest. (Дата обращения: 15.06.2022).