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»

**Тема:** «**Прогнозирование конечных свойств новых материалов**

**(композиционных материалов)»**

Слушатель Фаббро Александра Ивановна

# 

Москва, 2025

# **Содержание**

[Содержание 2](#_Toc106232838)

[Введение 3](#_Toc106232839)

[1. Аналитическая часть 5](#_Toc106232840)

[1.1. Постановка задачи 5](#_Toc106232841)

[1.2. Описание используемых методов 8](#_Toc106232842)

[1.3. Разведочный анализ данных 14](#_Toc106232843)

[2. Практическая часть 20](#_Toc106232844)

[2.1. Предобработка данных 20](#_Toc106232845)

[2.2. Разработка и обучение модели 20](#_Toc106232846)

[2.3. Тестирование модели 25](#_Toc106232847)

[2.4. Написать нейронную сеть, которая будет рекомендовать соотношение «матрица-наполнитель» 26](#_Toc106232848)

[2.5. Разработка приложения 29](#_Toc106232849)

[2.6. Создание удалённого репозитория и загрузка 32](#_Toc106232850)

[2.7. Заключение 33](#_Toc106232851)

[2.8. Список используемой литературы и веб ресурсы. 34](#_Toc106232852)

# **Введение**

Во второй половине XX века становление промышленного производства композиционных материалов способствовало прорыву в развитии ключевых отраслей — авиации, космонавтики, судо- и автомобилестроения, ракетной и ядерной техники, химической промышленности, медицины, строительства и средств индивидуальной баллистической защиты. С конца XX века, а особенно активно в XXI столетии, совершенствование теорий, методов и технологий производства композитов привело к значительному росту их выпуска, расширяя сферы применения и открывая новые возможности для инновационных решений.

Композиционный материал – это искусственно созданный неоднородный материал с монолитной структурой, состоящий из двух или более компонентов, разделённых чёткой границей. В большинстве композитов (кроме слоистых) выделяют матрицу (связующее) и армирующие элементы, распределённые в ней. Армирующие компоненты в конструкционных композитах отвечают за механические свойства (прочность, жёсткость и др.), а матрица обеспечивает согласованную работу армирующих элементов, защиту от механических повреждений и химических воздействий. Таким образом, матрица связывает и удерживает армирующие элементы, формируя единый материал с улучшенными характеристиками.

По структуре композиционные материалы делятся на несколько основных классов: волокнистые (армированы волокнами, нитями, нитевидными кристаллами), слоистые (армированы тканями, пленками, пластинами и другими слоистыми материалами), дисперсно-упрочненные (армированы материалами в виде дисперсных частиц) и нанокомпозиты.

Например, в защитных структурах средств индивидуальной баллистической защиты применяют все классы композитов. Но наибольшее распространение получили композиты, у которых в качестве связующего используются полимерные материалы, а в качестве армирующего материала – нити, жгуты, ткани или нетканые слоистые структуры из синтетических высокомодульных высокопрочных параарамидных или полиэтиленовых волокон. Такие композиты относятся к группе волокнистых или слоистых композитов, сокращенно их называют органопластиками. Органопластики применяются для изготовления шлемов и энергоемких жестких подложек комбинированной керамико-органопластиковой многослойной брони. С другой стороны, органопластики имеют низкую плотность, относительно высокую прочность на разрыв, ударную вязкость, но низкую прочность на сжатие и изгиб. Они легче, чем композиты, армированные стекловолокном и углеродным волокном.

На сегодняшний день современные достижения материаловедения, информационных технологий и наличие объемных баз данных позволяют автоматизировать проектирование полимерных композитов.

Дальнейшим развитием композитов, возможно, может стать создание «интеллектуальных» полимерных композитов, т.е. материалов, адекватно реагирующих на воздействия извне. Такие материалы будут способны не только противостоять внешним воздействиям, но и исправлять возникшие повреждения.

Создание новых композиционных материалов традиционно представляет собой длительный и ресурсоемкий процесс. Это связано с тем, что конечные свойства композита не могут быть точно определены на основе характеристик его отдельных компонентов. Для получения материала с заданными параметрами необходимо проводить многочисленные комплексные испытания различных комбинаций. В связи с этим особую актуальность приобретает разработка методов прогнозирования результатов, которая позволит существенно сократить как финансовые затраты на создание новых материалов, так и трудозатраты исследователей.

Чтобы оптимизировать процессы в этом направлении были разработаны две модели, способные с высокой вероятностью прогнозировать модуль упругости при растяжении и прочности при растяжении, а также создана нейронная сеть, рекомендующая соотношение «матрица - наполнитель». На этой основе было создано веб-приложение на фреймворке Flask.

1. **Аналитическая часть**
   1. **Постановка задачи**

Для выполнения данной задачи были предоставлены 2 файла: X\_bp.csv (с данными о параметрах, состоящими из 1024 строки и 11 столбцов) и X\_nup.csv (c данными нашивок, состоящими из 1041 строки и 4 столбцов).

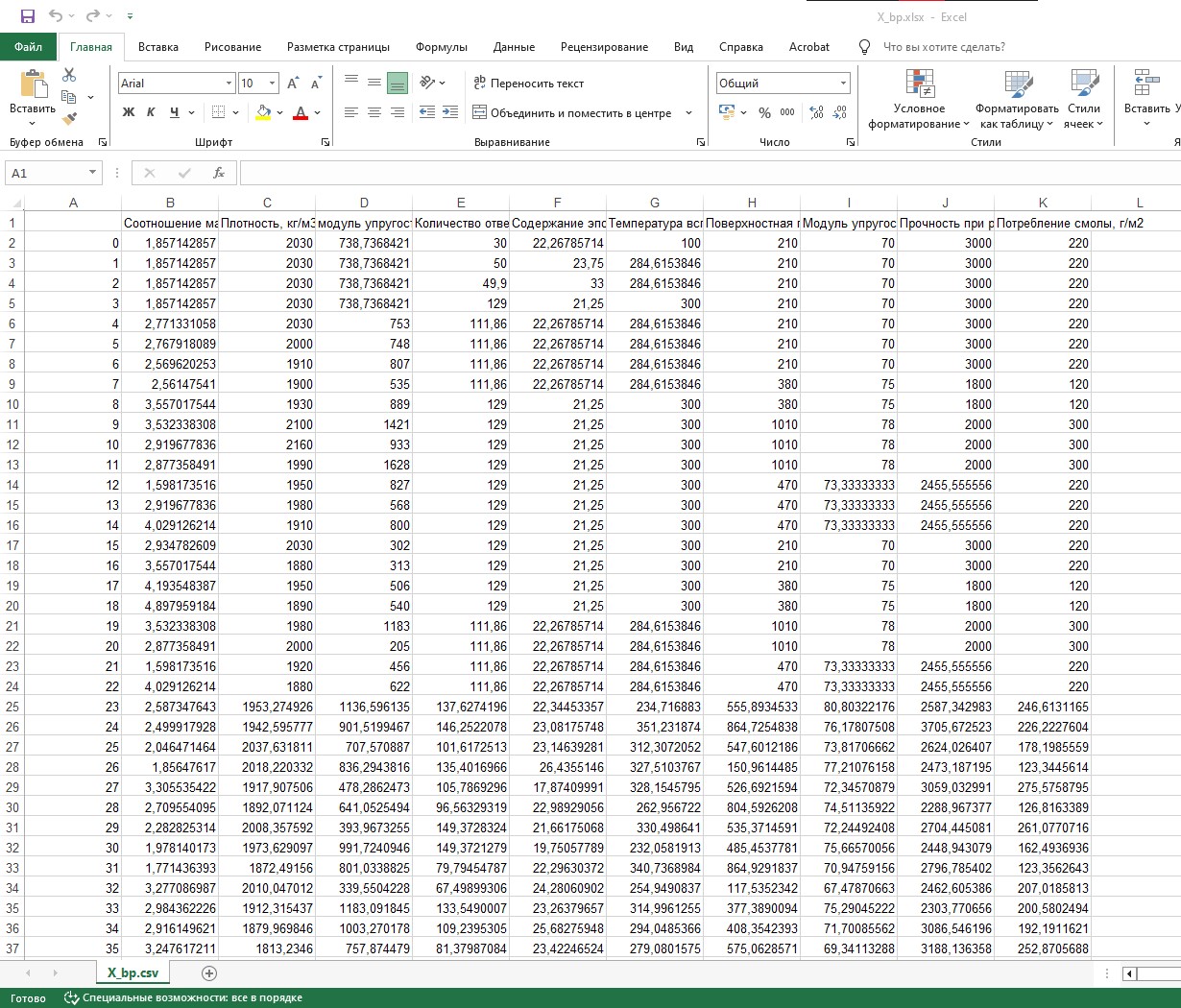


Рисунок 1 – файл X\_bp.csv (фрагмент)

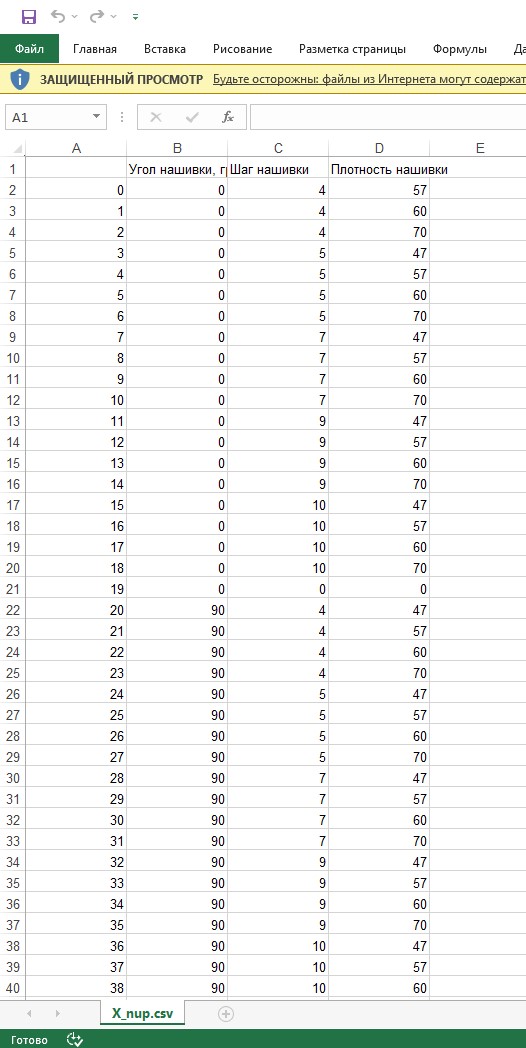


Рисунок 2 – файл X\_nup.csv (фрагмент)

Основная цель проекта заключается в разработке моделей для прогноза модуля упругости при растяжении и прочности при растяжении, а также в создании нейросети, которая будет рекомендовать соотношение «матрица-наполнитель».

Исходные данные, представленные в файлах X\_bp и X\_nup, на подготовительном этапе были интегрированы в единый датасет с целью их комплексного исследования.

Первым шагом работы стало создание датафрейма (X\_full) из импортированных данных.

Метод .shape показал, что датафрейм включает 1023 строки и 13 столбцов.

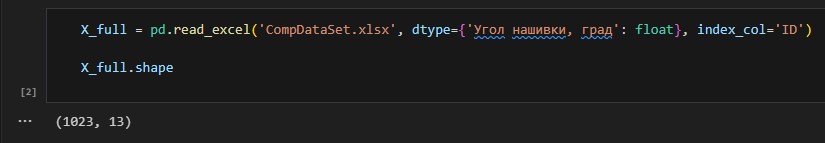


Рисунок 3 – Вывод размерности датафрейма.

В процессе загрузки данных значения параметра "угол нашивки, град" были приведены из типа int к типу float, в качестве индекса был взят столбец «ID».

Метод .info() вывел отсутствие пропусков по каждому столбцу и принадлежность всех столбцов к одному типу данных.

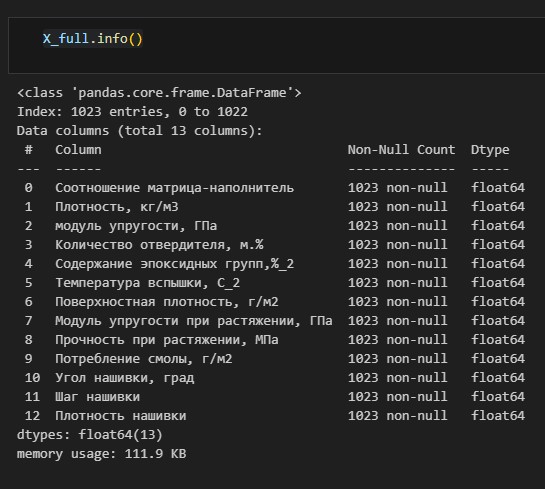


Рисунок 4 – вывод общей информации о датафрейме.

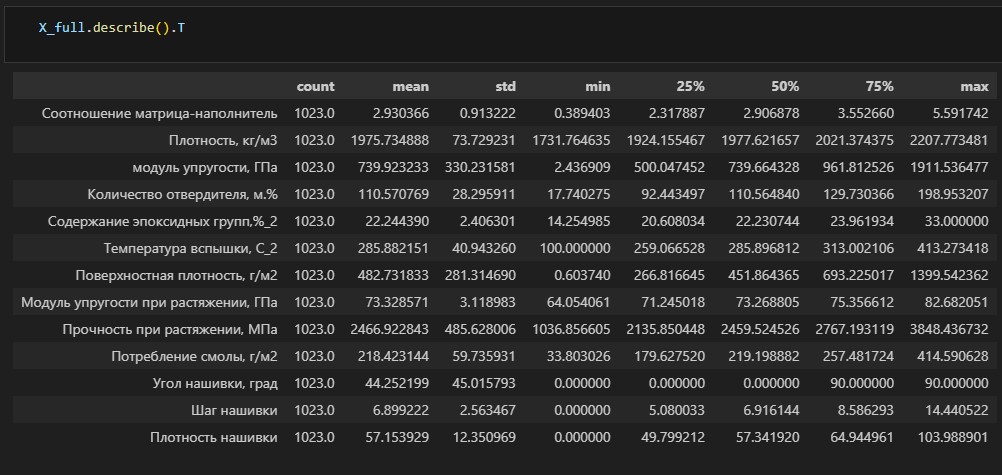
Поскольку все столбцы имеют числовой тип данных, была применена функция .decsribe().Т, выводящая сводную статистику, которая дает представление о распределении значений.  


Рисунок 5 – сводная статистика.

Свойство объекта .Т здесь использовано для удобства интерпретации данных. Полученные значения позволяют увидеть, есть ли в столбцах какие-либо выбросы, т.е. значения, которые сильно отличаются от большинства значений в наборе данных. Например, если для столбца «Модуль упругости» максимальное значение равно 1911.5, а 75% квартиль только 962, то значение 1911.5 считается выбросом.

* 1. **Описание используемых методов**

Первичное изучение данных показало необходимость проведения детального разведочного анализа, включающего визуализацию распределений через гистограммы, анализ выбросов с помощью Boxplot и изучение взаимосвязей переменных посредством попарных диаграмм рассеивания.

В данном проекте были применены следующие методы:

1. метод межквартильного размаха IQR для обработки выбросов;

2. метод K-means++ для кластеризации;

3. DBSCAN для кластеризации;

4. AdaBoost Regressor для обучения первой модели;

5. Decision Tree Regressor для обучения второй модели.

1. Метод межквартильного размаха IQR – устойчивая к выбросам половина выборки относительно медианы. Это удобный показатель изменчивости признаков для асимметричных распределений и данных с аномальными значениями.  
 Преимущества:

а) Робастность к выбросам — в отличие от стандартного отклонения, IQR не зависит от экстремальных значений.

б) Простота интерпретации — чётко показывает, где сосредоточено 50% центральных данных.

в) Работает с ненормальными распределениями — не требует симметричности данных.

г) Универсальность — применим для любых непрерывных данных (финансы, медицина, инженерия).

д) Используется в box-plot — визуализация разброса и выбросов становится наглядной.

Недостатки:

а) Игнорирует форму распределения — не различает симметричные и асимметричные данные.

б) Может пропускать выбросы — если аномалии находятся близко к Q1/Q3, но не за границами ±1.5×IQR.

в) Не учитывает все точки данных — использует только 25-й и 75-й процентили, "теряя" остальную информацию.

г) Эмпирический коэффициент (1.5) — границы выбросов выбраны условно и могут не подходить для некоторых задач.

2. K-means++ — это улучшенная версия алгоритма K-means для кластеризации данных, которая предлагает более эффективный способ выбора начальных центроидов (начальных точек кластеров). Основная цель — избежать плохой инициализации, которая может привести к неоптимальным результатам в стандартном K-means.

Ключевая идея:

-Первый центроид выбирается случайно из всех точек данных.

-Каждый следующий центроид выбирается с вероятностью, пропорциональной квадрату расстояния до ближайшего уже выбранного центроида.

-Процесс повторяется, пока не будут выбраны все k центроидов.

Преимущества:

а) Более быстрая сходимость - уменьшает количество итераций, необходимых для стабилизации кластеров.

б) Устойчивость к начальным условиям – позволяет лучше распределить центроиды на начальном этапе и уменьшает вероятность попадания алгоритма в локальные минимумы.

Недостатки:

а) Не решает все проблемы K-means:

-чувствительность к форме кластеров (лучше работает с выпуклыми кластерами одинакового размера),

-не подходит для данных с шумом или выбросами (можно комбинировать с методами вроде DBSCAN).

б) Зависимость от метрики расстояния — как и K-means, использует евклидово расстояние, что может быть неудачным выбором для данных сложной структуры.

в) Требует заданного числа кластеров (k) — как и оригинальный K-means, нуждается в предварительной оценке k (например, через "метод локтя").

г) Вычислительная нагрузка при инициализации — выбор центроидов сложнее, чем случайный, но всё равно быстрее, чем, например, иерархическая кластеризация.

3. DBSCAN – алгоритм кластеризации, основанный на концепции плотности точек данных в пространстве. Точки, окруженные областями высокой плотности объединяются в один кластер.

Преимущества:

а) Автоматическое определение числа кластеров — в отличие от K-means или иерархической кластеризации.

б) Находит кластеры произвольной формы — подходит для данных с нелинейной структурой (например, спирали).

в) Устойчив к выбросам — автоматически помещает шумовые точки в отдельную категорию.

г) Эффективно взаимодействует с разномасштабными данными — может работать с использованием пространственных индексов на больших наборах данных.

д) Стабильность результатов – не использует случайную инициализацию, поэтому выдает одинаковые результаты при одинаковых параметрах.

Недостатки:

а) Проблемы с кластерами разной плотности — если плотность сильно варьируется, может потребоваться оптимизация (например, HDBSCAN).

б) Чувствительность к параметрам — eps и min\_samples сильно влияют на результат.

в) Не работает для высокоразмерных данных — "проклятие размерности" снижает эффективность метрик расстояния.

г) Зависит от выбора метрики расстояния — евклидово расстояние может быть неудачным для категориальных данных.

д) Вычислительная сложность — для больших наборов данных без соответствующих индексных структур алгоритм может работать медленно.

е) Сложность интерпретации шумовых точек – не всегда очевидно, являются ли точки, классифицированные как шум действительно выбросами.

4. AdaBoost Regressor – ансамблевый метод, в основе которого лежит процедура бустинга, предполагающая итеративное построение ансамбля, где на каждом шаге новая слабая модель оптимизируется относительно остаточных ошибок, полученных от композиции предыдущих. Это обеспечивает адаптивное улучшение совокупного прогноза.

Преимущества:

а) Улучшение точности - бустинг позволяет значительно улучшить точность модели путем комбинирования множества слабых прогнозов. В задачах регрессии это достигается через усреднение их прогнозов, а в задачах классификации - через механизм голосования, что в совокупности дает более точный финальный результат.

б) Устойчивость к переобучению - бустинг может снизить риск переобучения за счет присвоения весов неправильно классифицированным входным данным.

в) Улучшенная обработка несбалансированных данных — бустинг позволяет справиться с несбалансированными данными, уделяя больше внимания точкам данных, которые неправильно классифицированы.

г) Лучшая интерпретируемость — бустинг может повысить интерпретируемость модели за счет разбиения процесса принятия решения на несколько процессов.

Недостатки:

а) Чрезмерная чувствительность к выбросам.

б) AdaBoost создаёт сложные модели, которые работают как 'чёрный ящик' - даже если он использует простые деревья, их комбинация становится настолько запутанной, что невозможно понять, как именно модель принимает решения.

в) Методология AdaBoost не позволяет эффективно конструировать ансамбли малого размера (при небольшом T) из алгоритмов высокой сложности, к которым относятся SVM и нейронные сети.

г) Необходима большая обучающая выборка (бэггинг может обходиться более маленькой выборкой).

5. Decision Tree Regressor – модель машинного обучения, основанная на вопросах, требующих ответа типа «да» или «нет» и представленная двоичным деревом. В дереве имеются корневой узел, узлы принятия решений, конечные узлы и ветви.

Глубина дерева решений – это количество уровней под корневым узлом.

Преимущества:

а) Интерпретируемость – деревья решений легко понять и интерпретировать. Графическое представление структуры наглядно визуализирует процессы принятия решений.

б) Работа с нелинейностью – могут фиксировать нелинейные связи в данных, что помогает при решении задач, где границы решений не являются линейными.

в) Работа со смешанными типами данных – без необходимости серьезной предварительной обработки могут принимать как числовые, так и категориальные значения.

г) Устойчивость к выбросам – так как разделение основано на относительных сравнениях и на них не влияют абсолютные значения.

д) Автоматический выбор переменной – снижает необходимость ручного подбора признаков.

Недостатки:

а) Переобучение – если деревья решений глубокие и улавливают шум в обучающих данных.

б) Нестабильность – небольшие изменения в данных могут привести к изменениям в структуре дерева.

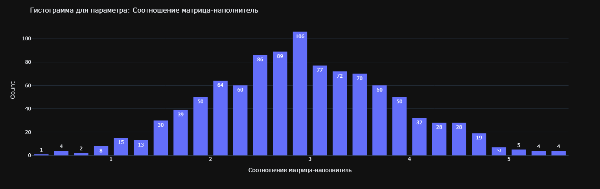
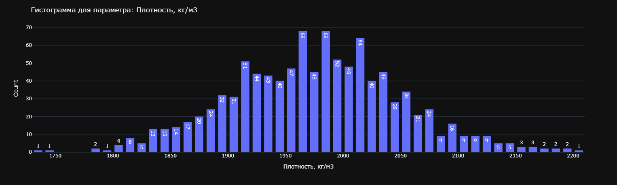
в) Чувствительность к зашумленным данным – выбросы могут привести к определенным разделениям, которые основаны на шуме, а не на значимых закономерностях.

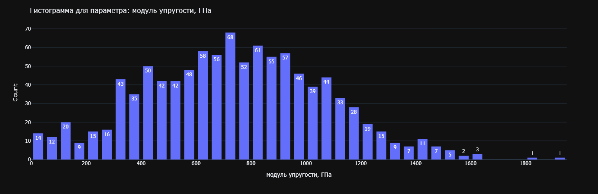
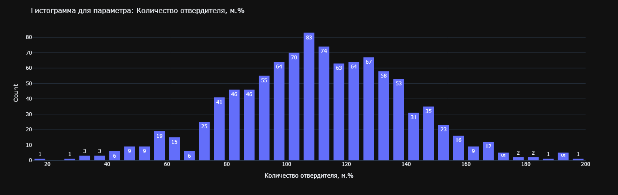
Сводная таблица предпосылок работоспособности каждого метода.

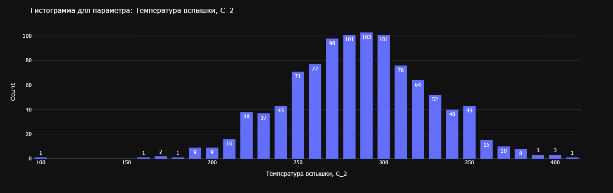
|  |  |  |
| --- | --- | --- |
| Метод | Тип задачи | Предпосылки применения |
| Межквартильный размах | Обнаружение и обработка выбросов | Ненормальное распределение некоторых признаков |
| K-means++ | Кластеризация | Числовые данные, известно k (методом локтя) |
| DBSCAN | Кластеризация | Используется в связке с K-means из-за шумных данных |
| AdaBoost | Регрессия | Нужна высокая точность при несбалансированных данных и слабых моделях |
| Decision Tree regressor | Регрессия | Быстрое прототипирование и нелинейные зависимости |

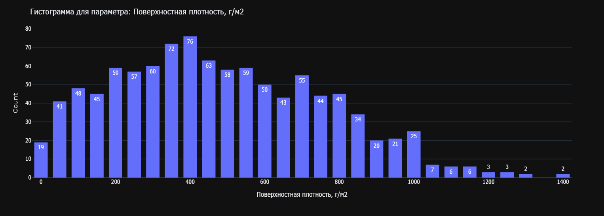
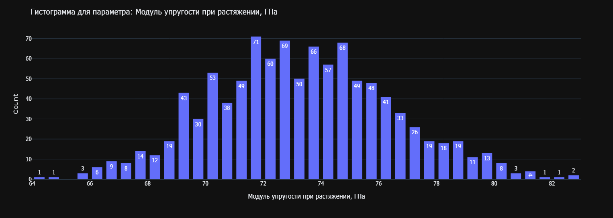
* 1. **Разведочный анализ данных**

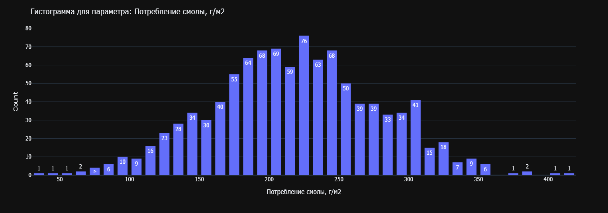
Первая задача на данном этапе – графическое представление распределений с использованием гистограмм для каждого признака. Визуализация позволит понять тренды, паттерны и взаимосвязи в данных. С помощью гистограмм можно определить как часто значение попадает в определенные диапазоны и какие имеются пики и провалы в данных.

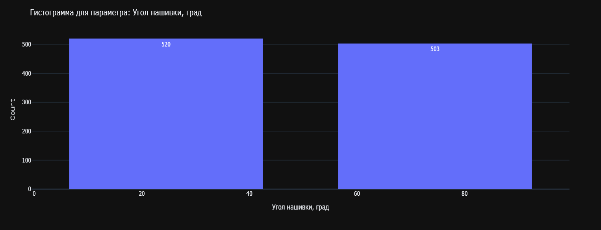
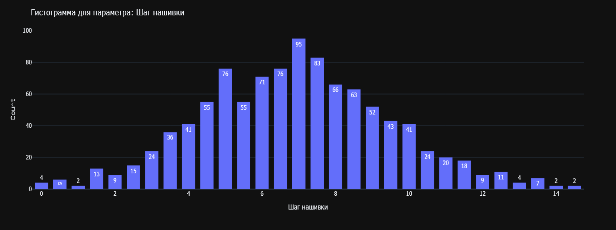
 

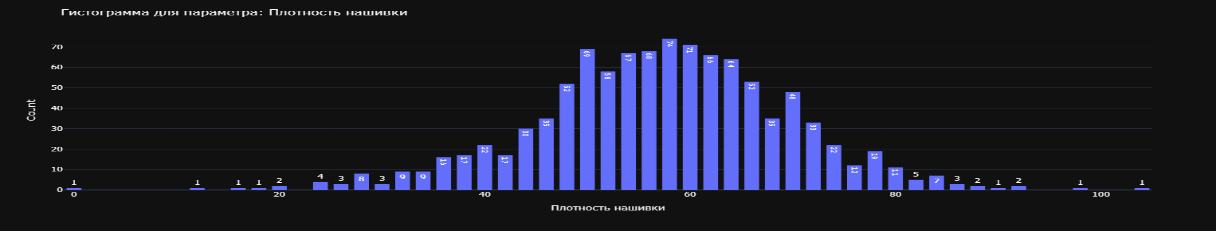
 



Рисунки 6-19 – гистограммы признаков по столбцам.

Попарные графики рассеивания позволяют визуализировать сравнение распределения пар числовых переменных и создают сетку точечных диаграмм. Там, где по горизонтали и вертикали пересекаются разные признаки, строится точечная диаграмма, на пересечении одного и того же признака строится гистограмма. Это позволяет определить, есть ли какая-либо зависимость между признаками.

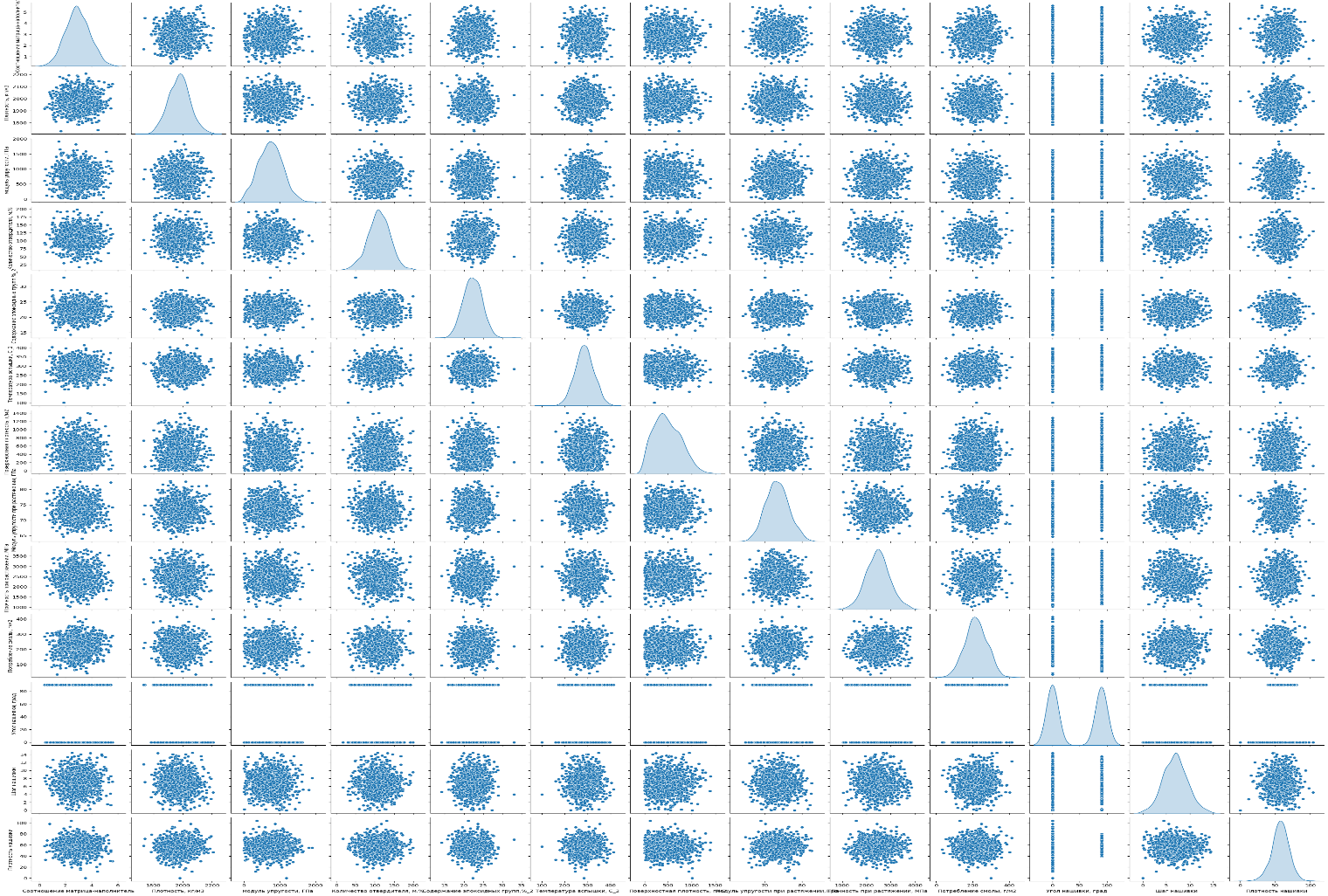


Рисунок 20 – попарная диаграмма рассеивания.

Тепловая карта – графическое представление значений в сетке с цветовой кодировкой. Нанесенные значения являются коэффициентами корреляций и отражают меру линейных отношений. Цветовая шкала показывает степень взаимосвязи между признаками и помогает выявить зависимости в наборах данных. Данная карта показывает крайне слабую корреляцию.

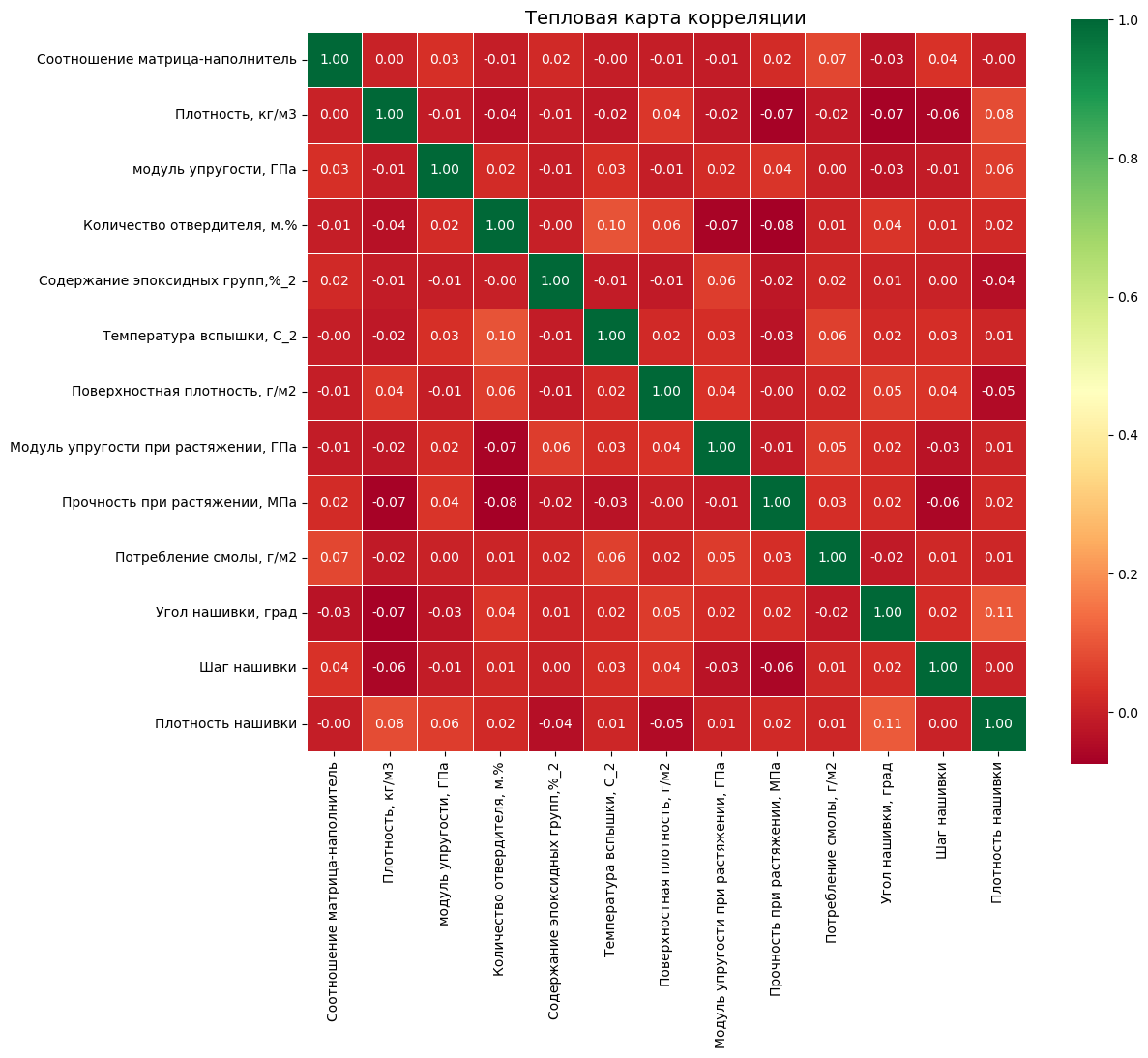


Рисунок 21 – тепловая карта корреляции.

Также расчет корреляции Пирсона и Кендалла не выявил очевидных взаимосвязей.

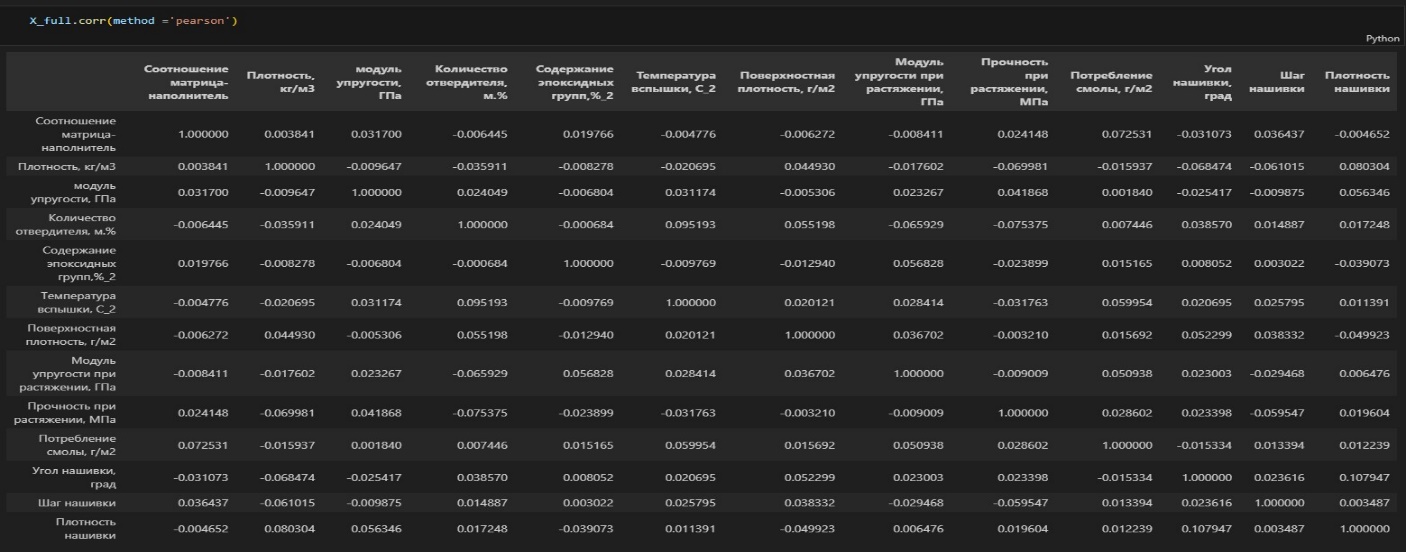


Рисунок 22 – расчет корреляции Пирсона

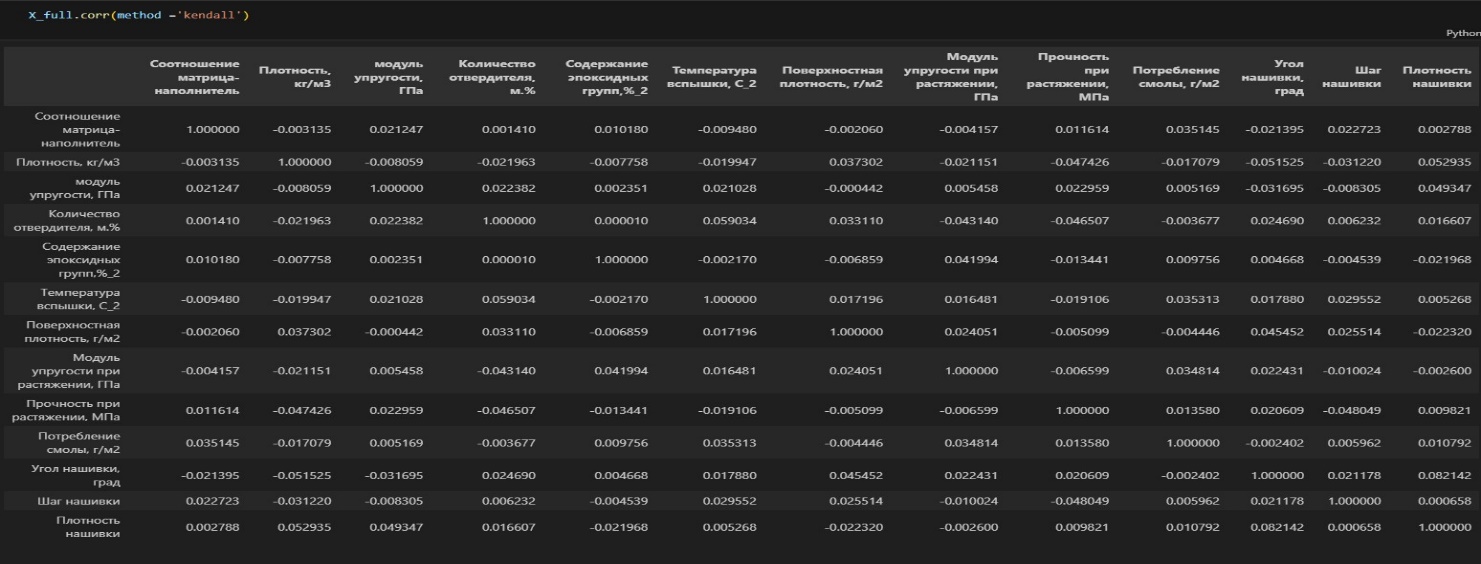


Рисунок 23 – расчет корреляции Кендалла.

Выбросы были обработаны методом межквартильного размаха, так как применение методов Z-score и 3sigma показало худшие результаты, что продемонстрировано на рисунках 24-26.

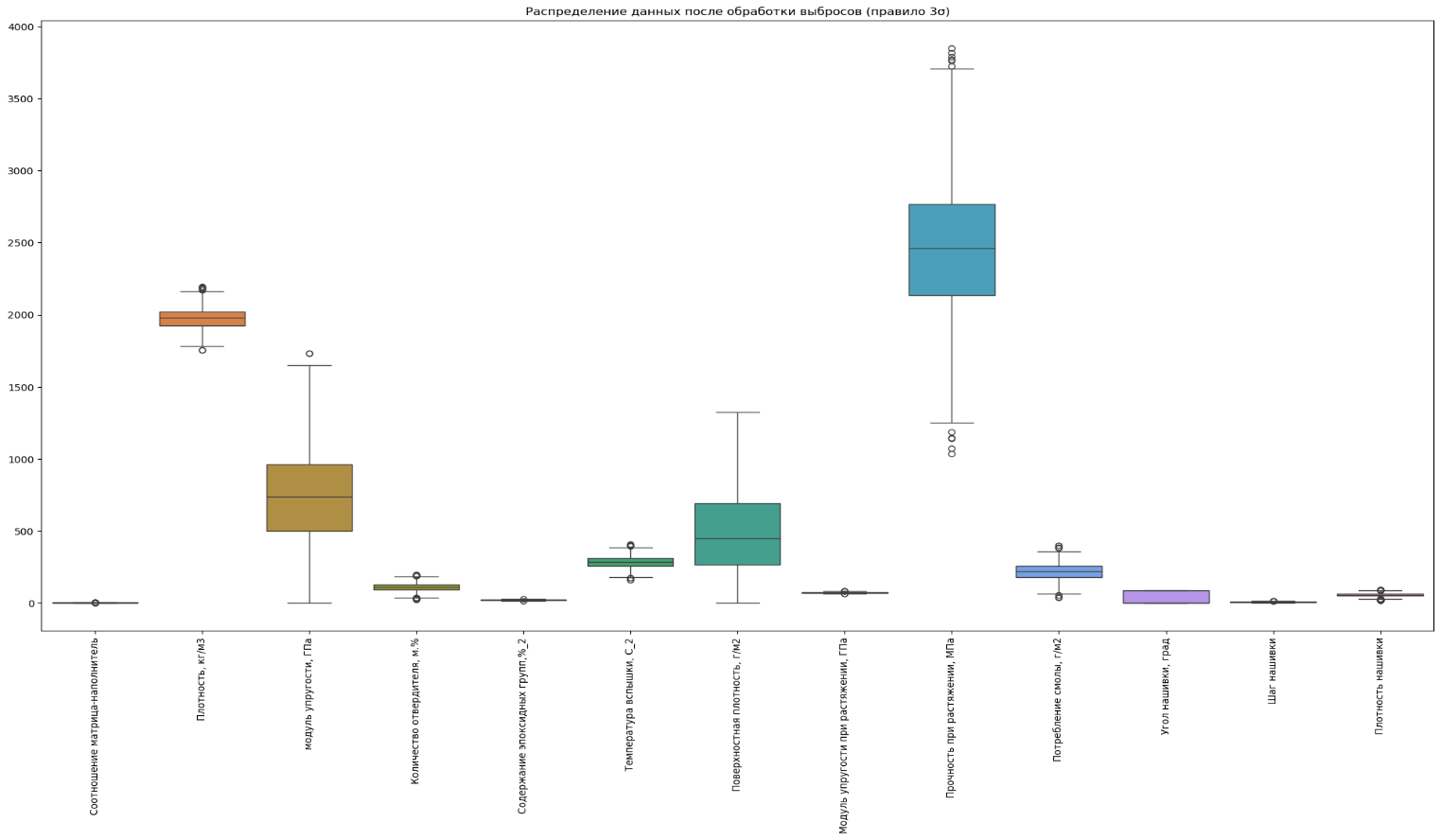


Рисунок 24 – обработка выбросов методом 3sigma.

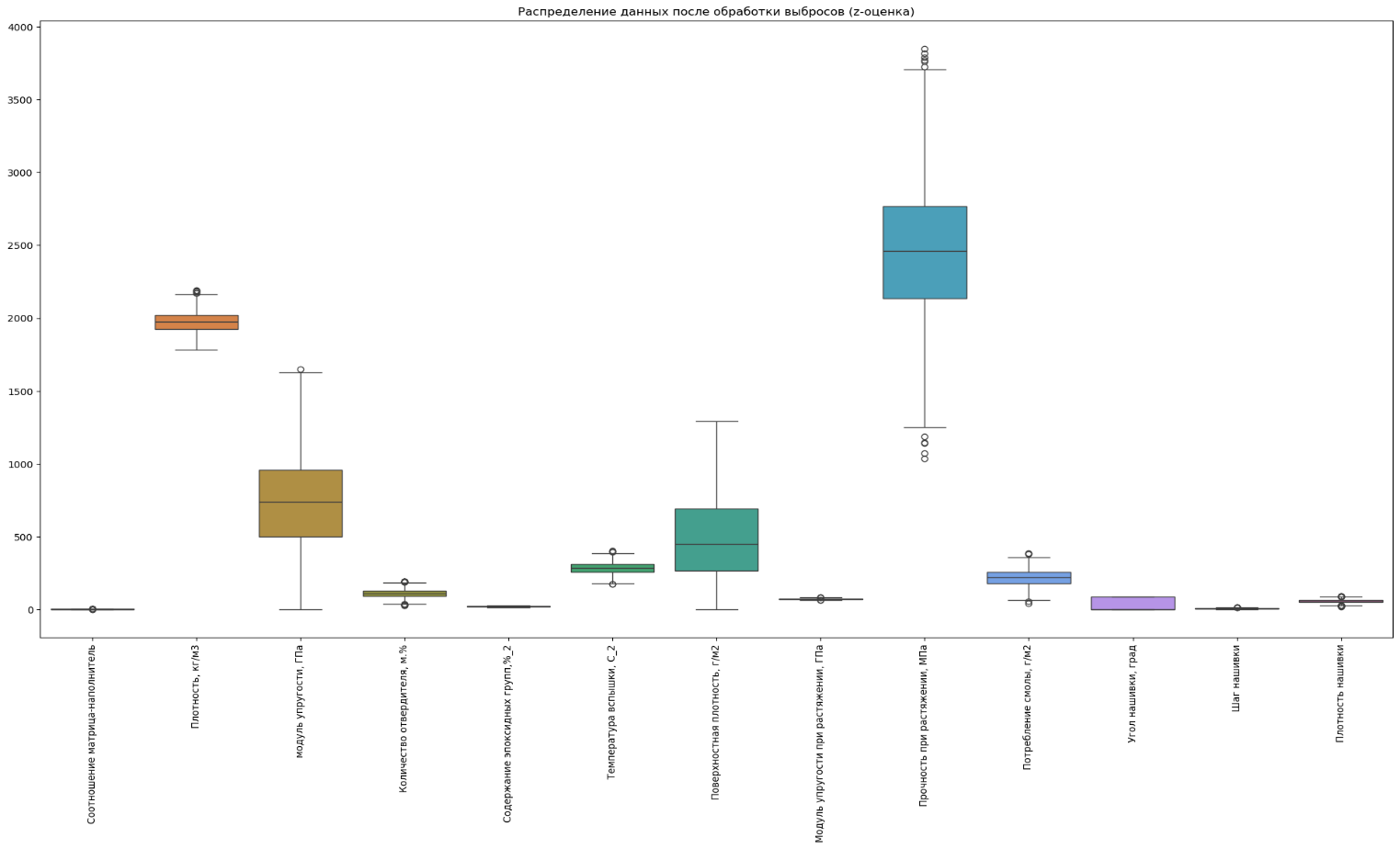


Рисунок 25 – обработка выбросов методом Z-score.

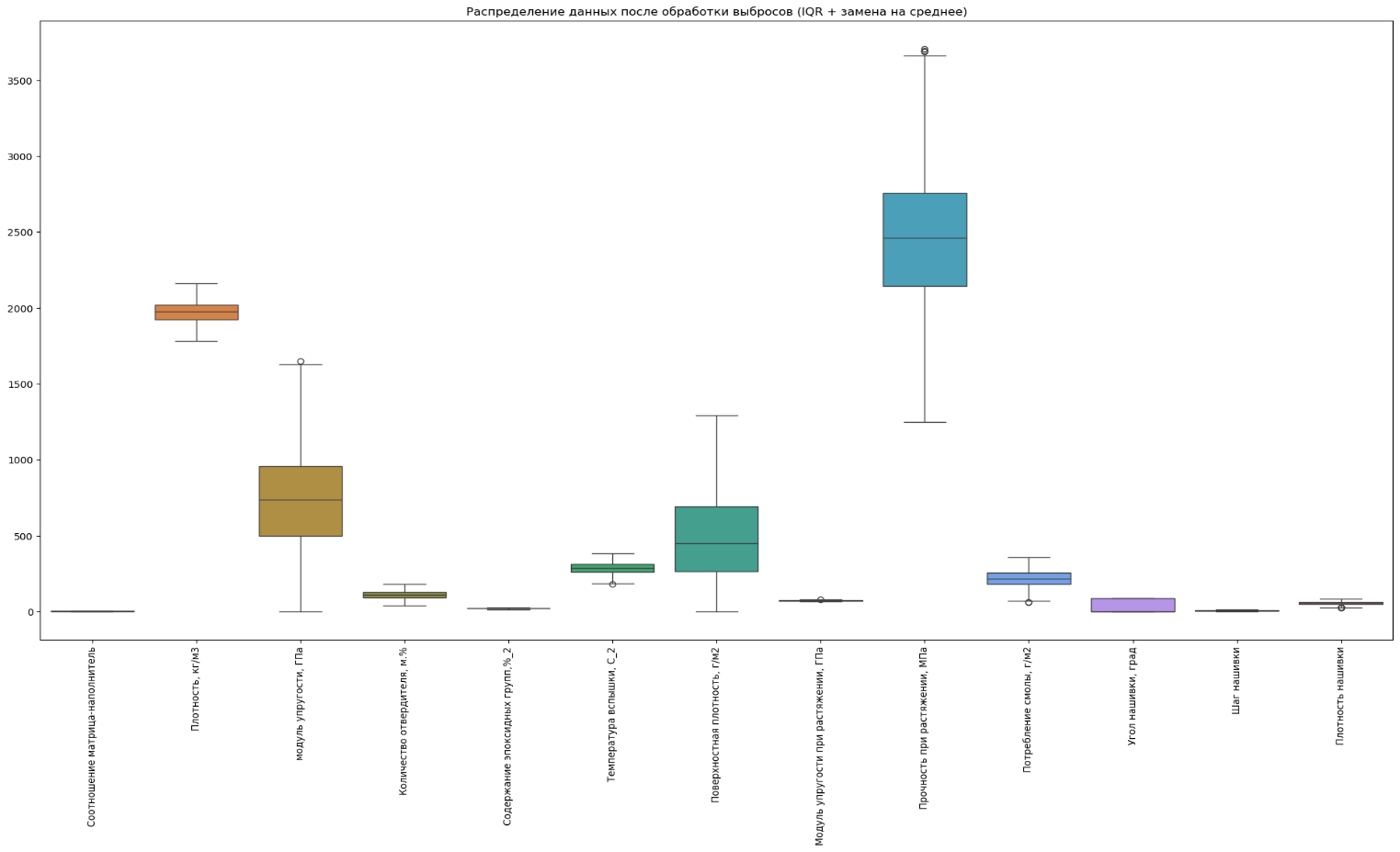


Рисунок 26 – обработка выбросов методом IQR.

Так как выборка небольшая, было принято решение во избежание недообученности моделей выбросы заменить средним значением вместо удаления.

1. **Практическая часть**
   1. **Предобработка данных**

На этапе предобработки данных нормализация не проводилась в связи с тем, что DBSCAN работает на разномасштабных данных и, в случае нормализации,

невозможно провести обратную трансформацию данных по причине наличия в датафрейме столбцов кластеров.

**2.2 Разработка и обучение модели**

При разработке и обучении моделей были использованы: AdaBoost Regressor и Decision Tree Regressor. В результате подбора

гиперпараметров и чередования было определено, что в первой модели лучший результат показал - AdaBoost Regressor, для второй - Decision Tree Regressor.

Данные были разделены на обучающую, валидационную и тестовую выборки. Размер тестовой выборки 30%.

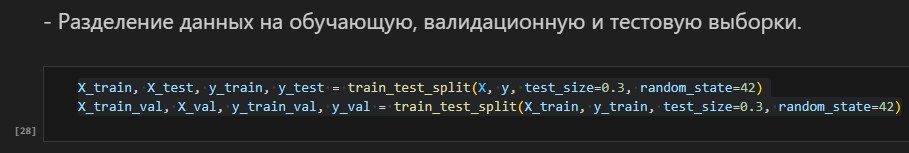


Рисунок 27 – разделение выборки.

Данные были подготовлены для обучения моделей: выделены признаки и целевая переменная.

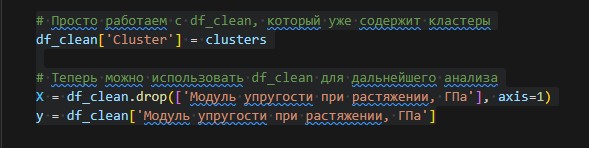


Рисунок 28 - подготовка данных для обучения моделей: выделение признаков и целевой переменной

На основе имеющегося набора данных создан слабый обучающий алгоритм (дерево решений глубиной 3), а также проведена оптимизация параметров с помощью GridSearchCV.

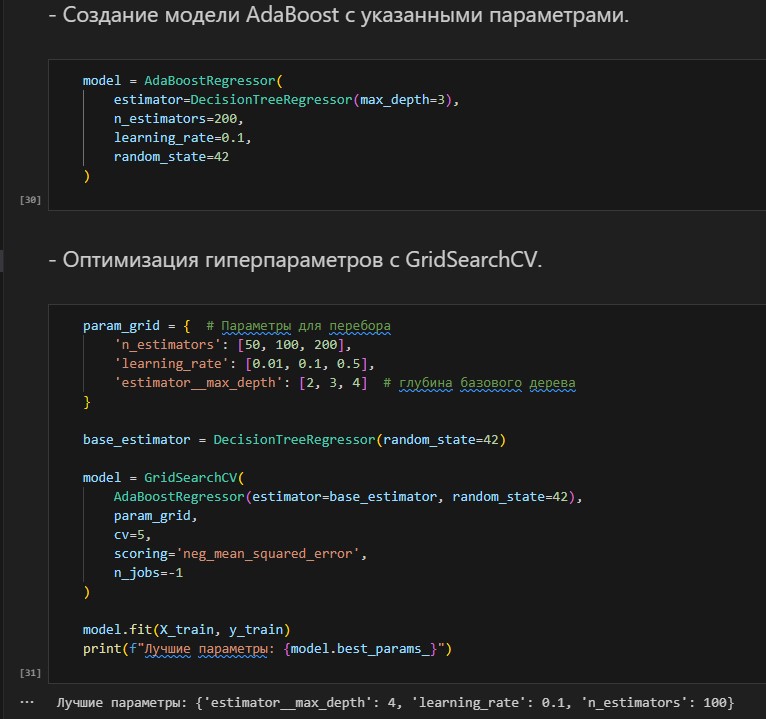


Рисунок 29 – создание модели 1 и оптимизация параметров.

GridSearchCV — это метод подбора гиперпараметров в машинном обучении, который:

1. Автоматически проверяет все возможные комбинации заданных гиперпараметров модели.
2. Оценивает качество каждой комбинации с помощью кросс-валидации (CV).
3. Выбирает наилучший набор параметров, максимизирующий заданную метрику.

Следующим этапом осуществлялось обучение модели 1 на валидационной и полной обучающей выборке, затем произведено прогнозирование на тестовой выборке и вывод метрик оценки модели.

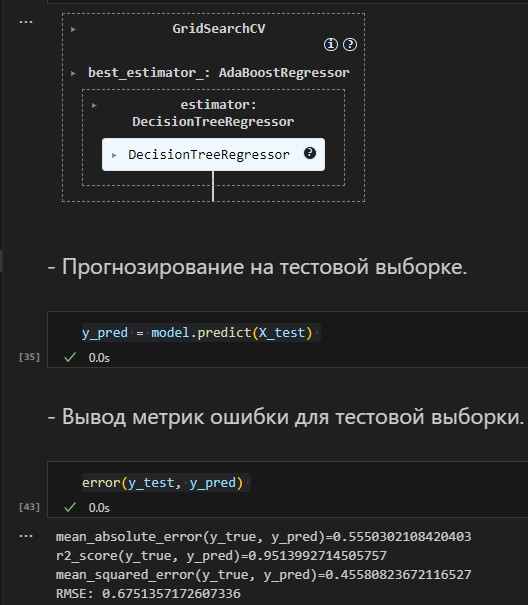


Рисунок 30 – обучение модели 1 и метрики.

Для второй модели действия были аналогичными, используя Decision Tree вместо AdaBoost.

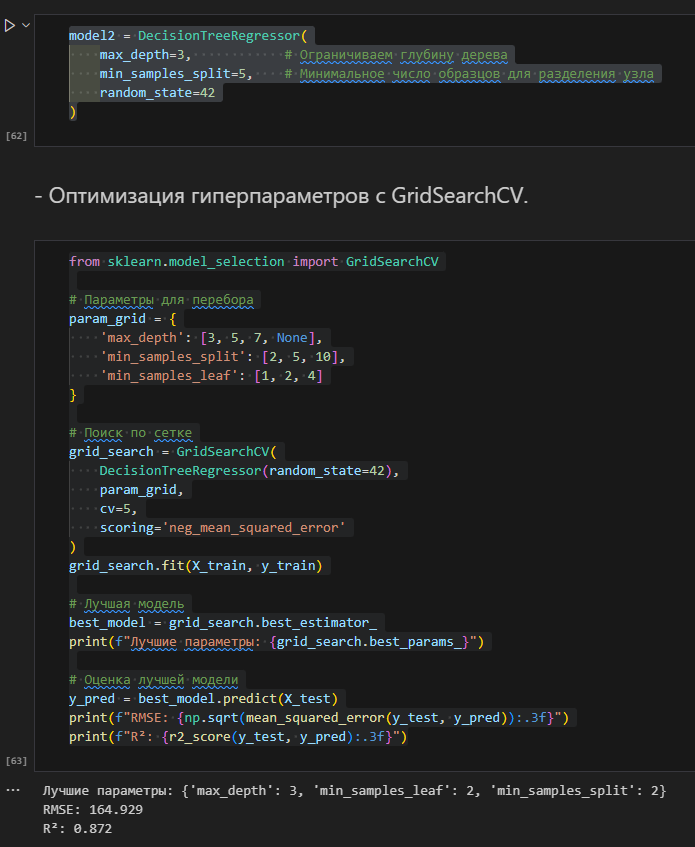


Рисунок 31 - создание модели 2 и оптимизация параметров.

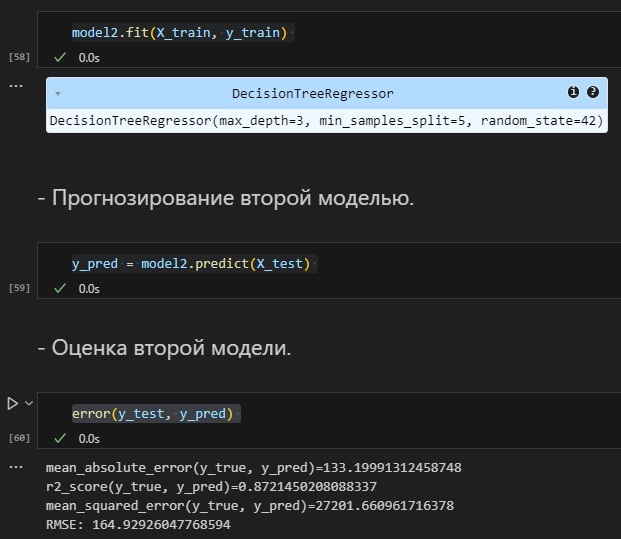


Рисунок 32 – обучение модели 2 и метрики.

**2.3 Тестирование модели**

В связи с регрессионной природой задачи, метрики будут представлять собой среднюю оценку разности между действительными и спрогнозированными значениями.

MAE (Mean Absolute Error) – оценивает насколько в среднем прогнозы модели отклоняются от средних значений по модулю. MAE измеряется в тех же единицах, что и исходные данные. Чем выше значение MAE, тем больше ошибок в прогнозах модели. Модель 1 в среднем ошибается на ±0.555 единицы измерения целевой переменной. Модель 2 в среднем ошибается на ±133.2 единицы измерения целевой переменной.

R2 – коэффициент детерминации. Показывает, какую долю дисперсии целевой переменной объясняет модель. Аналог точности в процентах для задач регрессии. Модель 1 объясняет 95.1% изменчивости данных, модель 2 объясняет 87.2% изменчивости данных. Эти значения можно интерпретировать как то, что модель 1 имеет очень высокое значение R2 и объясняет почти всю вариацию зависимой переменной. Модель 2 имеет высокое значение R2 и объясняет большую часть вариации зависимой переменной. Данные значения согласуются с предметной областью, т.к. в физических экспериментах ожидается высокая точность.

MSE (Mean Squared Error) – показывает насколько в среднем прогнозы модели отклоняются от реальных значений в квадрате. То есть среднее значение квадратов разностей между фактическими и предсказанными значениями. MSE используется в регрессионных задачах основной целью которых является предсказание числовых значений. Чем меньше значение MSE, тем лучше. В связи с тем, что MSE в отличие от MAE не сохраняет единицы измерения исходных данных - корректная интерпретация значений MSE сложна. Для того, чтобы показатель эффективности MSE имел размерность исходных данных, из него извлекают квадратный корень и получают показатель эффективности RMSE.

Таким образом для анализа вариации ошибок лучше использовать MAE и RMSE.

На этапе тестирования модели показали следующие результаты:

Модель 1 (AdaBoost Regressor) –

MAE: 0.5550302108420403

RMSE: 0.6751357172607336

R2: 0.9513992714505757

MSE: 0.45580823672116527

Модель 2 (Decision Tree Regressor) –

MAE: 133.19991312458748

RMSE: 164.92926047768594

R2: 0.8721450208088337

MSE: 27201.660961716378.

Отношение MAE/RMSE для модели 1 составляет 0.12.

Отношение MAE/RMSE для модели 2 – 31.73.

Такое значительное расхождение между RMSE и MAE в модели 2 может указывать на наличие выбросов или асимметрию в распределении ошибок.

**2.4 Написать нейронную сеть, которая будет рекомендовать соотношение матрица-наполнитель.**

Представленная нейронная сеть представляет собой многослойный перцептрон (MLP) для решения задачи регрессии. Ключевые характеристики:

- Тип сети: Полносвязная нейронная сеть (Dense Neural Network)

- Задача: Предсказание соотношения матрица-наполнитель для композитных материалов

- Входные данные: 12 числовых признаков, описывающих свойства материалов

- Выход: Одно непрерывное значение (регрессия).

Архитектура имеет несколько ключевых особенностей:

1. Настраиваемая структура:

- Количество слоев: 1-3 (настраивается через Hyperband)

- Количество нейронов: от 32 до 256 с шагом 32

- Функции активации: ReLU или tanh (выбирается автоматически).

2. Регуляризация:

- L2-регуляризация ядер (коэффициент 0.01)

- Dropout слои (вероятность от 0.0 до 0.5)

- Ранняя остановка (patience=10)

3. Оптимизация:

- Используется Adam optimizer

- Настраиваемый learning rate (от 1e-4 до 1e-2)

- Функция потерь: MSE (Mean Squared Error)

Преимущества архитектуры:

1. Гибкость и адаптивность:

- Возможность автоматического подбора гиперпараметров через Hyperband

- Динамическое изменение структуры сети в зависимости от данных

2. Механизмы предотвращения переобучения:

- Комбинация L2-регуляризации и Dropout

- Ранняя остановка обучения

- Кросс-валидация (5 фолдов)

3. Качественная подготовка данных:

- Стандартизация признаков (StandardScaler)

- Разделение на train/validation/test наборы

4. Комплексная оценка модели:

- Использование нескольких метрик (MAE, MSE, MAPE, R²)

- Визуализация предсказаний и важности признаков

Особенности реализации:

1. Процесс обучения:

- Поддержка кросс-валидации

- Поиск лучших гиперпараметров через Hyperband

- Сохранение лучшей модели

2. Инструменты мониторинга:

- Отслеживание метрик на валидационном наборе

- Визуализация процесса обучения

- Анализ важности признаков

3. Производственные особенности:

- Автоматическое создание директорий для хранения моделей

- Сохранение и загрузка моделей

- Полный пайплайн от данных до предсказаний



Рисунок 33 – метрики нейронной сети.

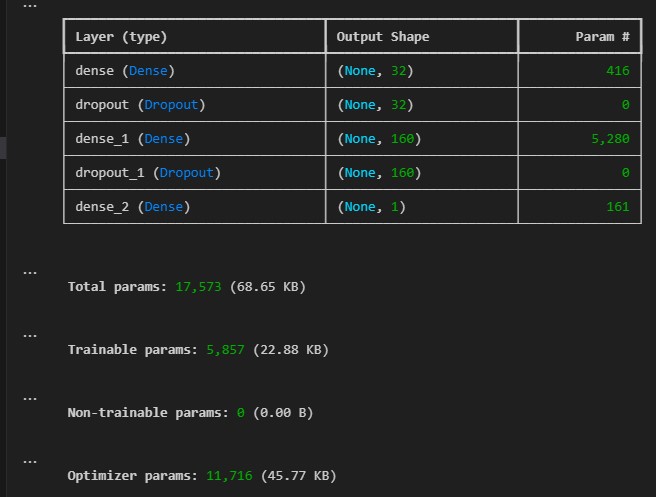


Рисунок 34 – summary.

**2.5 Разработка приложения**

Данное приложение — это основной файл Flask, папка templates, с шаблоном html - страницы, папка model c сохранённой моделью для данных.

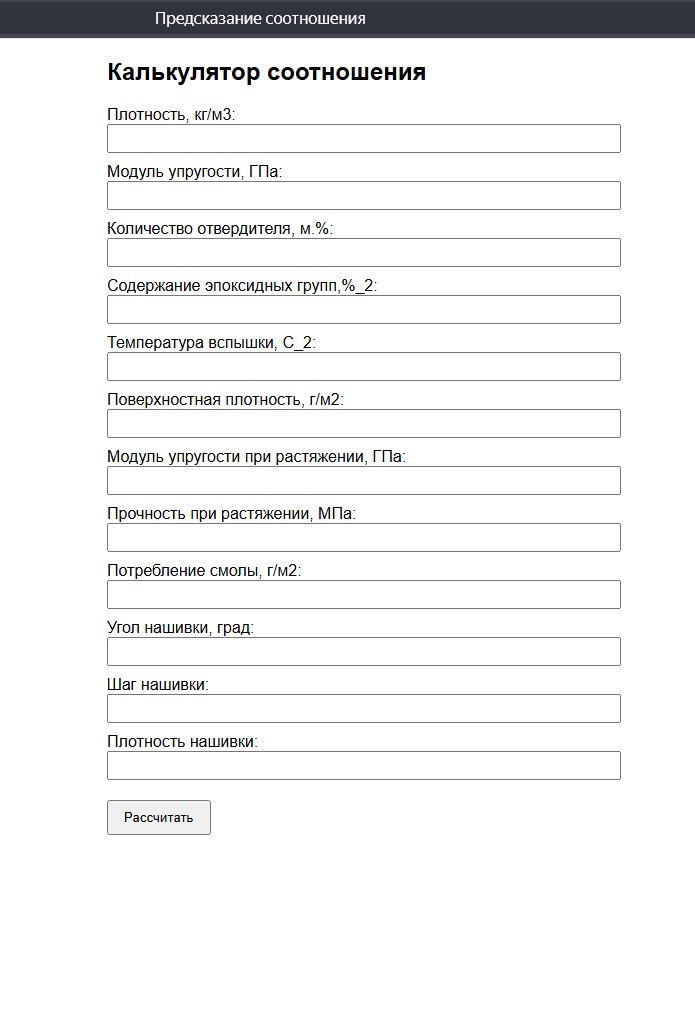


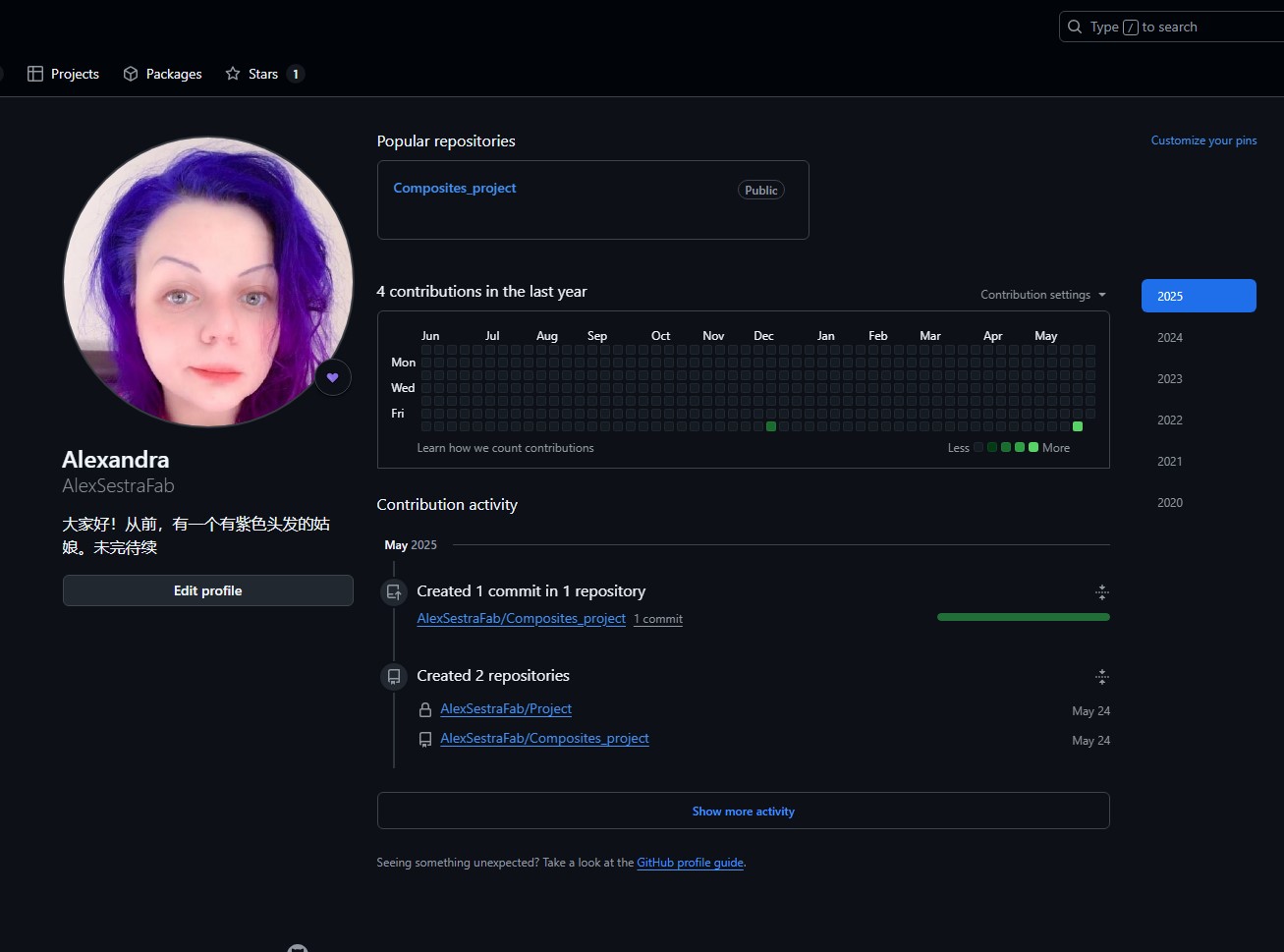
Рисунок 35 – приложение.



Рисунок 36 – код приложения.

**2.5 Создание удаленного репозитория и загрузка**

Репозиторий был создан на github.com по адресу: <https://github.com/AlexSestraFab/Composites_project/tree/master>



**Заключение**

Основываясь на выявленных ограничениях, были предприняты меры по улучшению качества данных: устранение выбросов и генерация новых производных признаков. Однако вследствие невозможности консультации с экспертами в предметной области - эти эксперименты не принесли ощутимого результата. Несмотря на вышеизложенное, метрики, полученные нейронной сетью и AdaBoost Regressor, свидетельствуют о принципиальной возможности построения точных прогнозных моделей в рамках поставленной задачи. Таким образом, исходя из комплексного подхода к проблеме необходима верификация результатов проведенного исследования с экспертами-материаловедами.

**2.7 Список используемой литературы**

1. Шарапова В. А.: Композиционные материалы специального назначения. Учебное пособие. Уральский федеральный университет. 2020
2. В. В. Кудинов, И. К. Крылов: Влияние компонентов на свойства полимерных материалов. Монография-справочник. Изд. Наука, 2021.
3. В. И. Костиков: Технология композиционных материалов. Учебное пособие. Инфра-Инженерия, 2021.
4. И. Ф. Кобылкин, В. В. Селиванов: Материалы и структуры легкой бронезащиты. Изд. МГТУ им Н. Э. Баумана, 2014.
5. Т. А. Гузева, Г. Е. Нехороших, А. И. Долгих: Определение физико-механических характеристик полимерных композиционных материалов. Учебно-методическое пособие. Изд. МГТУ им Н. Э. Баумана, 2020.
6. Луис Серрано: Грокаем машинное обучение. Изд. Питер, 2025.
7. Peter Bruce, Andrew Bruce: Practical statistics for Data Scientists. O’Reilly Media, 2017.
8. Wes McKinney: Python for Data Analysis. O’Reilly Media, 2022.
9. Эндрю Траск: Грокаем глубокое обучение, Изд. Питер, 2025.
10. Франсуа Шолле: Глубокое обучение на Python. Изд. Питер, 2025.
11. Sebastian Raschka, Vahid Mirjalili: Python Machine Learning. Packt publishing, 2019.
12. Stefanie Molin: Hands-on Data Analysis with Pandas. Packt publishing, 2019.
13. Pratap Dangeti: Statistics for Machine Learning. Packt publishing, 2017.
14. Charles Kufs: Stats with Cats. Wheatmark, 2011.
15. Thomas Nield: Essential Math for Data Science, O’Reilly Media, 2022.
16. Lee Vaughan: Python tools for Scientists. No scratch press, 2023.
17. Араки Масахиро: Занимательное машинное обучение. ДМК Пресс, 2020.
18. Машинное обучение и Data Science: Совершенствуем AI-модели с помощью AdaBoost: - Режим доступа: <https://www.mql5.com/ru/articles/14034> (дата обращения 21.05.2025)
19. Библиотека Matplotlib: - Режим доступа <https://devpractice.ru/files/books/python/Matplotlib.book.pdf?ysclid=mb75lqbf3e129025026> (дата обращения 02.05.2025)
20. Decision Tree in Machine Learning: - Режим доступа https://www.geeksforgeeks.org/decision-tree-introduction-example (дата обращения 10.05.2025)
21. Multi-Layer Perceptrons: - Режим доступа <https://towardsdatascience.com/multi-layer-perceptrons-8d76972afa2b/> (дата обращения 16.05.2025)
22. Simple NN with Python: Multi-Layer Perceptron: - Режим доступа

<https://www.kaggle.com/code/androbomb/simple-nn-with-python-multi-layer-perceptron> (дата обращения 17.05.2025)

1. Introduction to the Keras Tuner: - Режим доступа <https://www.tensorflow.org/tutorials/keras/keras_tuner> (дата обращения 21.05.2025)
2. 25+ Most Used Keras Snippets in 2025: - Режим доступа <https://www.geeksforgeeks.org/25-most-used-keras-snippets-in-2025/> (дата обращения 21.05.2025)
3. Enhancing Multi-Layer Perceptron Performance: Demystifying Optimizers: - Режим доступа <https://towardsai.net/p/l/enhancing-multi-layer-perceptron-performance-demystifying-optimizers> (дата обращения 21.05.2025)