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: Шевардаев Максим Александрович

Москва, 2024

# Содержание

[Содержание 2](#_Toc185901604)

[Введение 3](#_Toc185901605)

[1. Аналитическая часть 5](#_Toc185901606)

[1.1. Постановка задачи 5](#_Toc185901607)

[1.2. Описание используемых методов 6](#_Toc185901608)

[1.3. Разведочный анализ данных 20](#_Toc185901609)

[2. Практическая часть 26](#_Toc185901610)

[2.1. Предобработка данных 26](#_Toc185901611)

[2.2. Разработка и обучение модели 26](#_Toc185901612)

[2.3. Тестирование модели 28](#_Toc185901613)

[2.4. Создание архитектуры полносвязной нейронной сети для рекомендации соотношения «матрица – наполнитель». 28](#_Toc185901614)

[2.5. Разработка приложения 30](#_Toc185901615)

[2.6. Создание удалённого репозитория и загрузка 32](#_Toc185901616)

[Заключение 34](#_Toc185901617)

[Список используемой литературы и веб ресурсы. 35](#_Toc185901618)

# Введение

Композиционные материалы представляют собой уникальные вещества, состоящие из нескольких несмешивающихся компонентов, которые сохраняют четкие границы раздела. Основной компонент, называемый матрицей, является основной связующей фазой, в которую внедряются другие элементы, известные как армирующие или наполнители.

Популярность композитов объясняется их преимуществами по сравнению с традиционными металлическими материалами и сплавами. Примеры таких материалов включают железобетон, где стальная арматура и бетон работают вместе, и древесноволокнистую плиту (ДВП), представляющую собой комбинацию древесных волокон и полимерного связующего.

Базальт, магматическая порода, образуется при охлаждении и затвердевании силикатного расплава и широко распространен на поверхности Земли и других планетах Солнечной системы. Он образует океаническую кору, особенно на срединно-океанических хребтах. В последние годы наблюдается рост использования базальтовых волокон в композитных материалах.

Базальтопластик — это современный композитный материал, состоящий из базальтовых волокон и органического связующего вещества. Он успешно конкурирует с традиционными металлическими материалами благодаря своим исключительным свойствам, таким как высокая коррозионная стойкость и устойчивость к химическим воздействиям. Целью данного исследования является прогнозирование свойств новых материалов, созданных на основе базальтопластика.

Увеличение количества материалов для создания новых композитов подчеркивает необходимость точной оценки их свойств при минимальных затратах. Для этого применяются два основных подхода: проведение физического тестирования образцов или использование математических моделей для оценки свойств. Процесс разработки композитов традиционно длителен и сложен, так как точное предсказание свойств композитов на основе свойств их компонентов невозможно без проведения множества тестов. Поэтому актуальной задачей является прогнозирование свойств для снижения затрат на разработку и рабочую силу. В рамках исследования были разработаны модели, способные точно предсказывать модули упругости и прочности при растяжении. Кроме того, были созданы две нейронные сети для оценки сочетаний матрицы и наполнителя. Одна из этих сетей послужила основой для создания удобного веб-приложения на платформе Flask, обеспечивающего высокий уровень пользовательского интерфейса.

Рисунок - скриншот приложения на фреймворке Flask

# Аналитическая часть

## Постановка задачи

В данной выпускной работе использовались 2 файла: X\_bp.xlsx в которых собраны данные о характеристиках базальтопластика, состоящий из 1023 объектов (строк) и 11 признаков (столбцов), а также файл X\_nup.xlsx из 1039 объектов и 4 признаков (столбцов).

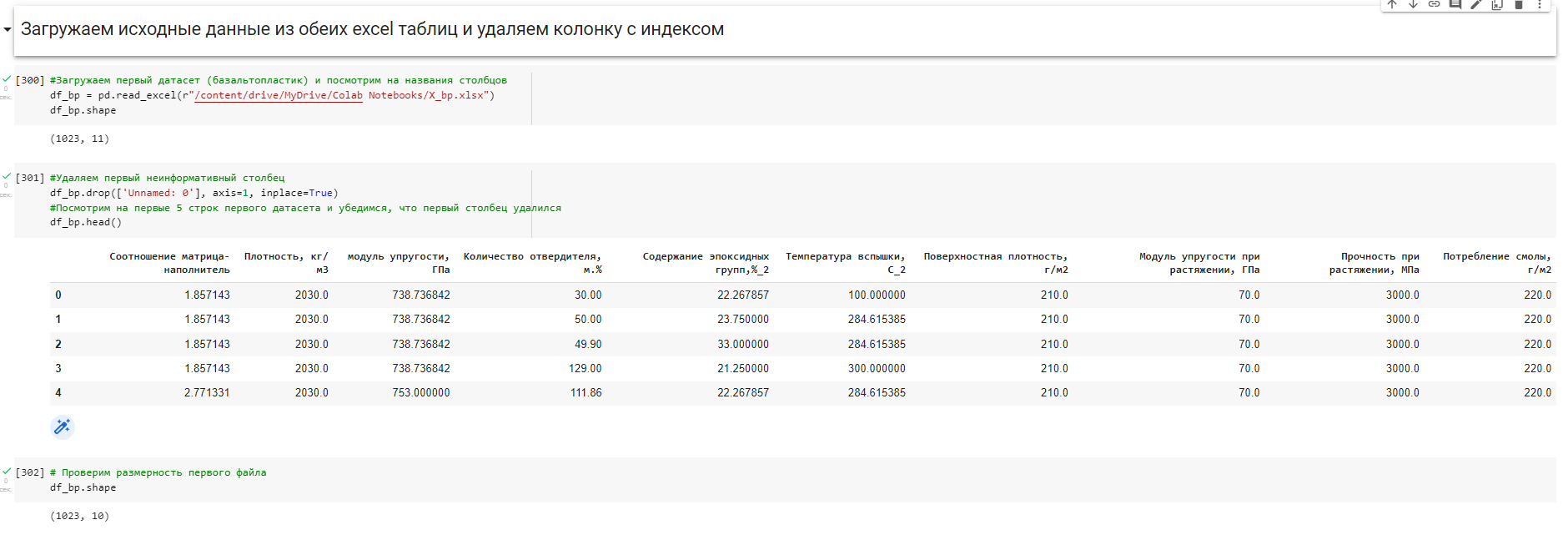
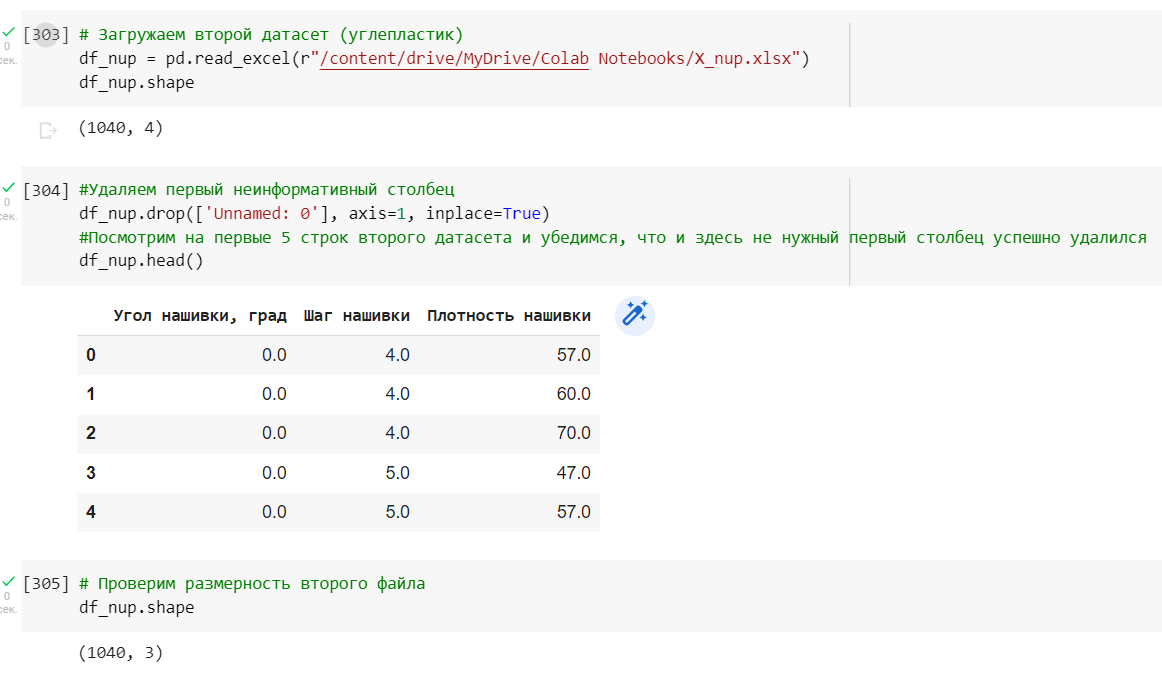


Рисунок – пример начала работы с файлом X\_bp.xlsx

Целью данной выпускной работы является обучение модели для прогноза модуля упругости при растяжении и соотношения «матрица-наполнитель». Объединение производится методом merge по столбцу 'Unnamed: 0', та часть информации, что не имеет аналогичных строк в X\_bp, не входит в итоговую таблицу df.

Рисунок - пример начала работы с файлом X\_nup.xlsx

Далее проводим разведочный анализ данных, строятся различные гистограммы распределения каждой из переменной, диаграммы Boxplot (ящик с усами), попарные графики рассеяния точек, тепловая карта и скрипичная диаграмма.

Для каждого признака получены: среднее, минимальное, максимальное и медианное значения, проведен анализ и исключение выбросов, проверены на наличие пропусков, удалены шумы и выбросы, выполнены нормализация и стандартизация. Произведено обучение ряда моделей для прогноза модуля упругости при растяжении и прочности при растяжении. Создана архитектура полносвязной нейронной сети, веса которой будут оценивать соотношение матрица/наполнитель. Разработаны приложение на Flask с графическим интерфейсом. Оценена точность модели на тренировочном и тестовом датасете. Создан репозиторий в GitHub и размещен код исследования.

## Описание используемых методов

Поскольку целевые показатели: имеют значения из непрерывного диапазона, необходимо использовать модели машинного обучения, предназначенные для регрессии. В данном исследовании использовались следующие методы:

* Полиномиальная регрессия;
* К-ближайших соседей;
* Случайный лес;
* Градиентный бустинг;
* Метод опорных векторов для регрессии;

Перед тем как начать раздел теоретического описания используемых методов необходимо отметить следующие факты:

1. Во всех случаях для выбора гиперпараметров использовался поиск по сетке. Это метод, при котором подбор осуществляется путем создания списка из набора значений гиперпараметров, которые необходимо протестировать на данной модели. Конечный выбор осуществляется перебором всех возможных комбинаций, оценкой производительности, обычно с использованием кросс-валидации, и выбором лучшей комбинации гиперпараметров. Очевидно, что с увеличением числа гиперпараметров рост числа комбинаций экспоненциален, поэтому чаще применяют альтернативы наподобие случайного поиска или байесовской оптимизации. Так как используемый набор данных имеет весьма небольшие размеры, использование поиска по сетке не столь вычислительно затратно и поэтому оправдано.
2. Для оценки производительности использовалась кросс-валидация. Это  
   метод оценки производительности, часто применяемый при недостаточном объеме выборки. Смысл метода в разделении имеющейся выборки   
   на k частей (в данной работе 5) и обучение на k-1 частях и тестирование   
   на одной оставшейся части. По результатам тестирования затем вычисляется средний показатель некоторой оценки. Это позволяет получить более надежную оценку модели, но как и поиск по сетке вычислительно затратен.
3. Так как не имеет смысла каждый раз создавать один и тот же код для обучения с разными целевыми показателями, было решено создать одну функцию, что вобрала бы в себя все модели и используемые гиперпараметры,   
   а также графики и обозначения.

Полиномиальная регрессия - это один из видов линейной регрессии, используемый в случаях, когда зависимость между независимыми переменными и целевой имеет явно нелинейный характер. Рассмотрим формулу

, (1)

То есть в отличие от линейной регрессии, коэффициенты полинома не ограничиваются нулевой и первой степенью, а имеют более высокие порядки.   
Очевидно, что с ростом степени полинома модель может описывать все более сложные зависимости, но при этом сильно возрастает вероятность переобучения.   
С целью уменьшить вероятность переобучения используют регуляризацию, формулы которых представлены ниже:

L2-регуляризация

(2)

L1-регуляризация

(3)

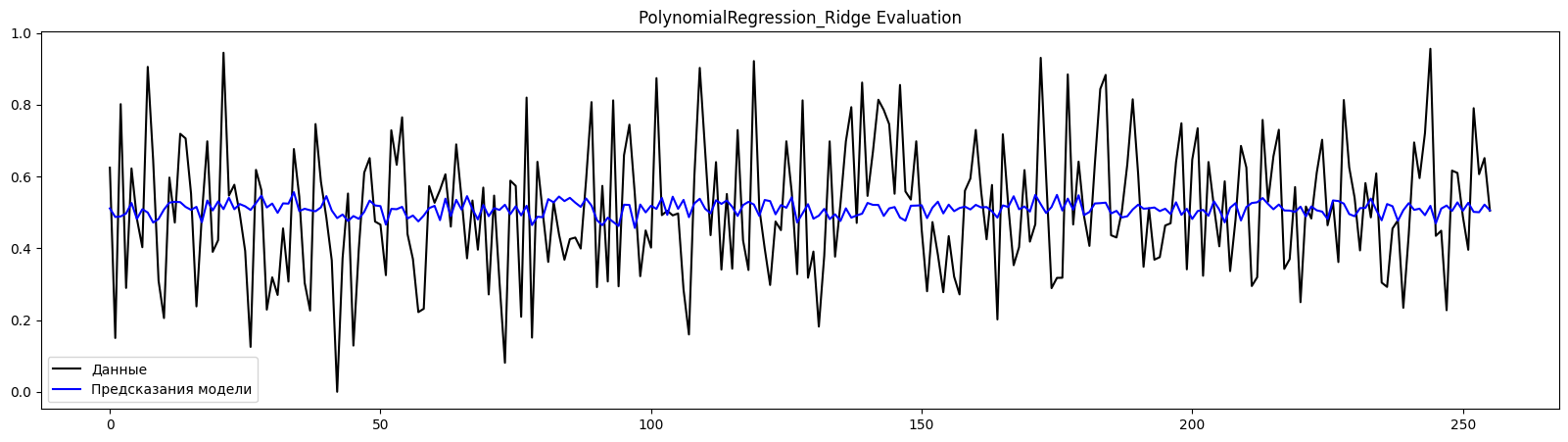
Как можно видеть из представленных выше формул, L2-регуляризация штрафует веса модели на сумму квадратов весов (коэффициентов), что позволяет пресекать чрезмерный их рост. В то же время L1-регуляризация добавляет штраф за   
абсолютное значение весов, что позволяет обнулить те из них, что слабо влияют на целевой признак. Таким образом отбираются признаки и уменьшается сложность модели.

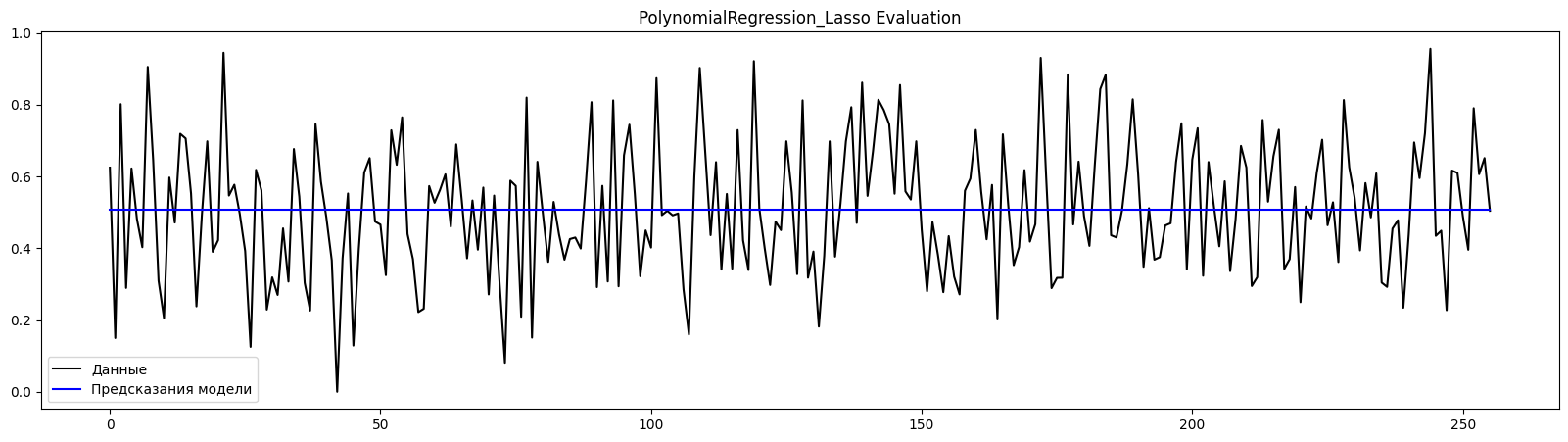
Задача полиномиальной регрессии заключается в подборе коэффициентов β, минимизирующих сумму квадратов отклонений предсказанных значений от истинных. Формулируется это как задача минимизации функции потерь:

, (4)

В данной работе использовались следующие гиперпараметры:

* poly\_\_degree - степени полинома, применяемые для преобразования признаков.
* lasso\_\_alpha - параметр L1-регуляризации. Увеличение значения усиливает регуляризацию, что уменьшает число маловажных для модели деталей. Уменьшение снижает регуляризацию захватывая тем более сложные зависимости, но увеличивая риск переобучения.
* lasso\_\_fit\_intercept - указывает будет ли включен свободный член в модель. Включение позволяет учитывать сдвиг в данных.

Рисунок 4 - график метода полиномиальной регрессии с L2-регуляризацией для прочности при растяжении, МПа

Рисунок 5 - график метода полиномиальной регрессии с L1-регуляризацией для прочности при растяжении, МПа

Метод опорных векторов (Support Vector Regression) - это метод регрессии, основанный на методе опорных векторов (SVM), но с тем отличием, что используется для предсказания непрерывных величин. Основная идея заключается в поиске функции, что могла бы аппроксимировать данные с отклонением от истинных значений на показатель не более установленной величины.

Формута модели опорных векторов выглядит как:

(5)

SVR минимизирует норму вектора весов w, чтобы уменьшить сложность модели,   
с учетом того, что ошибка предсказания f(x) не превышает значение ε:

Ограничения формулируются в виде:

(7)

(8)

(9)

Для работы с нелинейными данными используются ядерные функции:

(10)

Формула линейного ядра представима в виде:

(11)

Rbf-ядро:

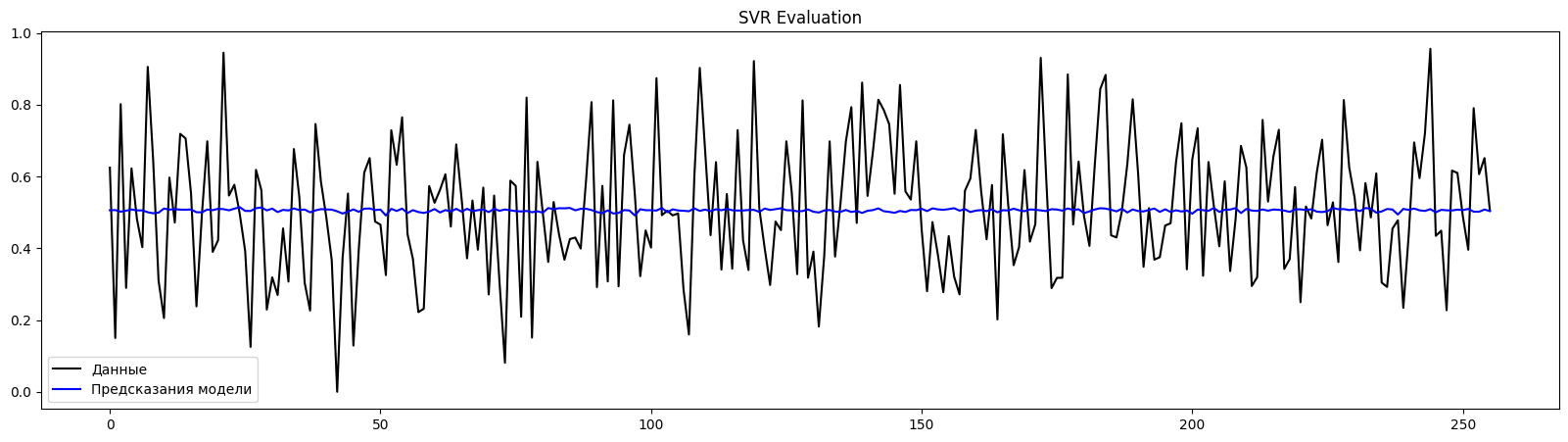
(12)

Результат SVR для входного значения х вычисляется по формуле:

(13)

Используемые гипрепараметры:

* Kernel - тип ядра, используемый для преобразования данных в пространство более высокой размерности, чтобы сделать данные линейно разделимыми. В данной работе использовались следующие ядра: полиномиальное ядро (poly), линейное ядро (linear) и ядро радиальной базисной функции (rbf).
* С - параметр регуляризации, определяющий равновесие между максимально возможной гладкостью модели и минимизации ошибки на тренировочных данных. Меньшие значения приводят к большему числу ошибок на тренировочных данных, но могут улучшить обобщение модели.
* Gamma - параметр, используемый только в некоторых ядрах, таких как RBF и полиномиальные ядра, для определения ширины гауссового колокола или степени полинома.

Рисунок 6 - график метода опорных векторов для прочности при растяжении, МПа

Случайный лес - это ансамблевый метод машинного обучения, использующий комбинацию большого числа слабых моделей (чаще деревьев решений) с последующим усреднением всех ответов посредством голосования (классификация) или усреднения (регрессия). При этом каждое дерево обучается не на всех данных, а на случайных подвыборках признаков и подвыборке имеющихся данных. На каждом узле дерева принимается решение о разбиении подмножества на подвыборки используя:

Индекс Джини (классификация)

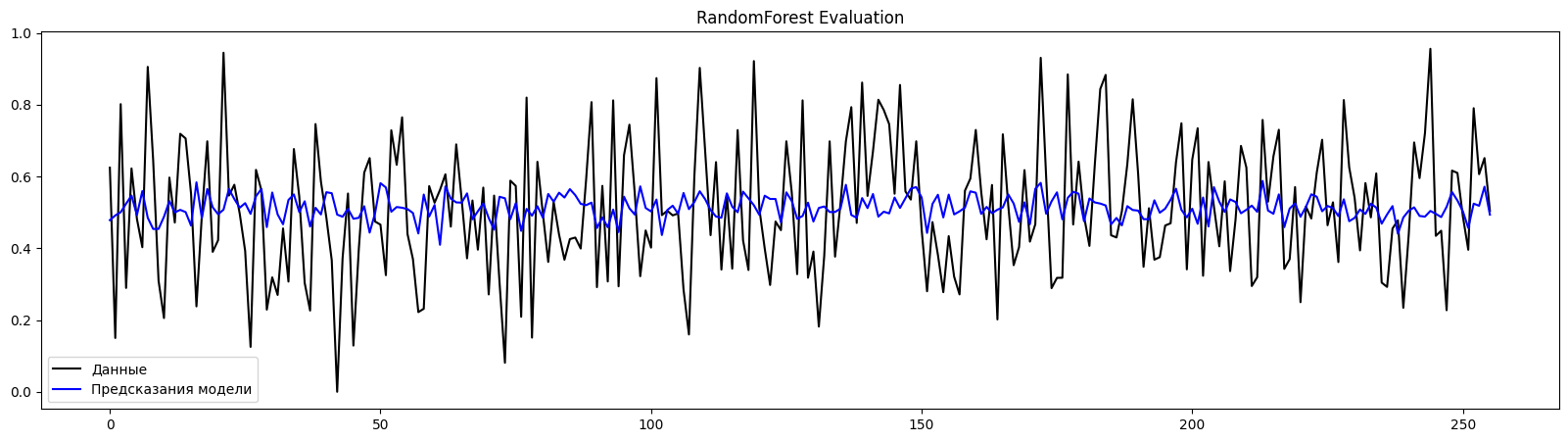
(14)

Среднеквадратичная ошибка (регрессия)

(15)

Используемые гиперпараметры:

* n\_estimators - количество деревьев используемых в ансамбле. Увеличение числа деревьев может повысить точность и устойчивость модели, но также увеличивает вычислительная сложность.
* max\_depth - максимальная глубина дерева. Увеличение глубины может привести к переобучению.
* min\_samples\_split - минимальное число образцов, необходимых для разбиения узла на два дочерних. Более высокие значения предотвращает создание слишком мелких узлов, что может привести к переобучению.

Рисунок 7 - график метода случайного леса для прочности при растяжении, МПа

Градиентный бустинг - это также ансамблевый метод, основанный на   
бустинге, в отличие от бэггинга используемого в случайном лесе. Используются слабые модели, обучающиеся последовательно, причем каждая последующая минимизирует остаточную ошибку предыдущих. Формула для начального значения предсказаний (инициализация модели):

(16)

Вычисление остатка

(17)

Обучение дерева на остатках

(18)

Обновление предсказаний

(19)

Функция потерь для классификации (логистическая регрессия)

(20)

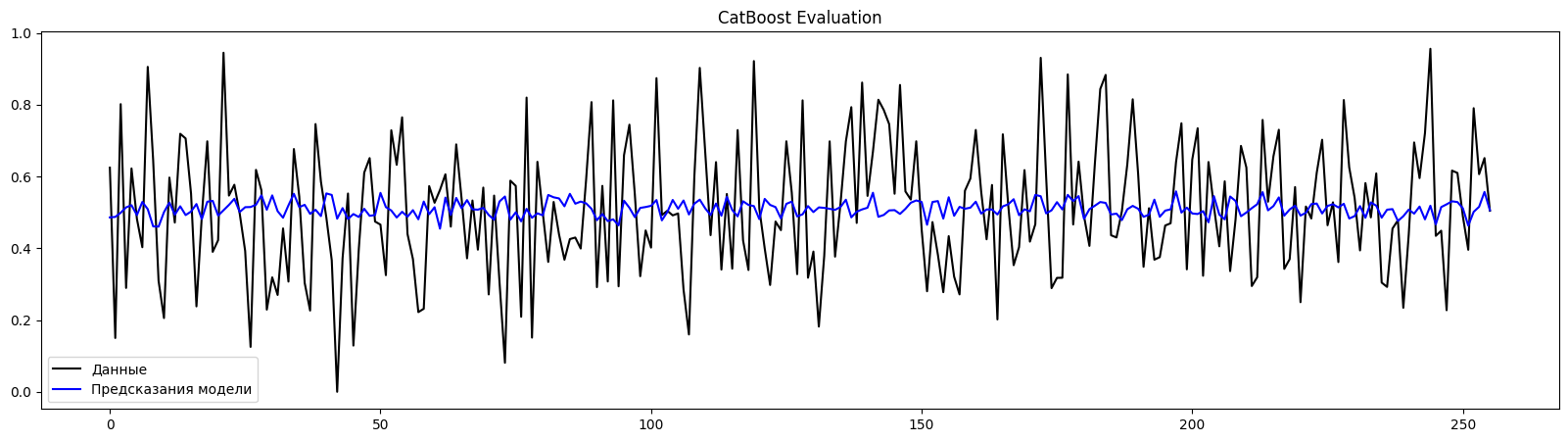
Функция потерь для регрессии (среднеквадратичная ошибка)

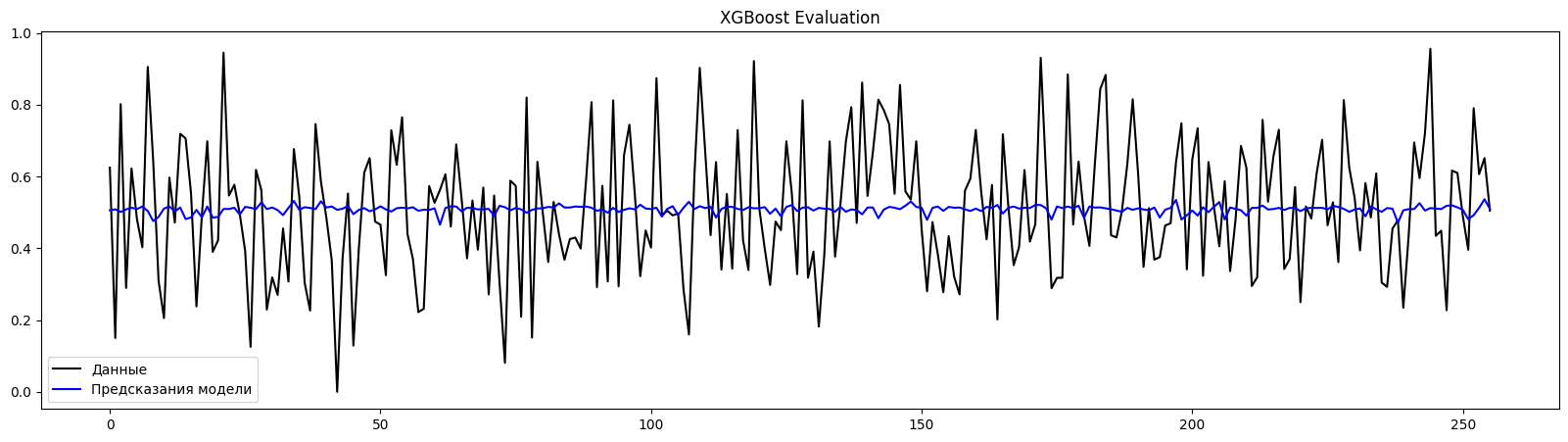
(21)

Суммарное предсказание

(22)

В работе используются две модели: XGBoost и CatBoost.

Рисунок 8 - график метода CatBoost для прочности при растяжении, МПа

Рисунок 9 - график метода XGBoost для прочности при растяжении, МПа

Метод ближайших соседей - относится к метрическим методам. По заданной метрике вычисляется расстояние до объектов с известными целевой переменной и методом большинства голосов выбирает по k-ближайшим методом, обычно, взвешенного голосования класс объекта (классификация) или среднее значение (регрессия).

К примеру, взвешенное расстояние в задаче регрессии представлена формулой:

(23)

В данной работе используются: манхэттенское расстояние

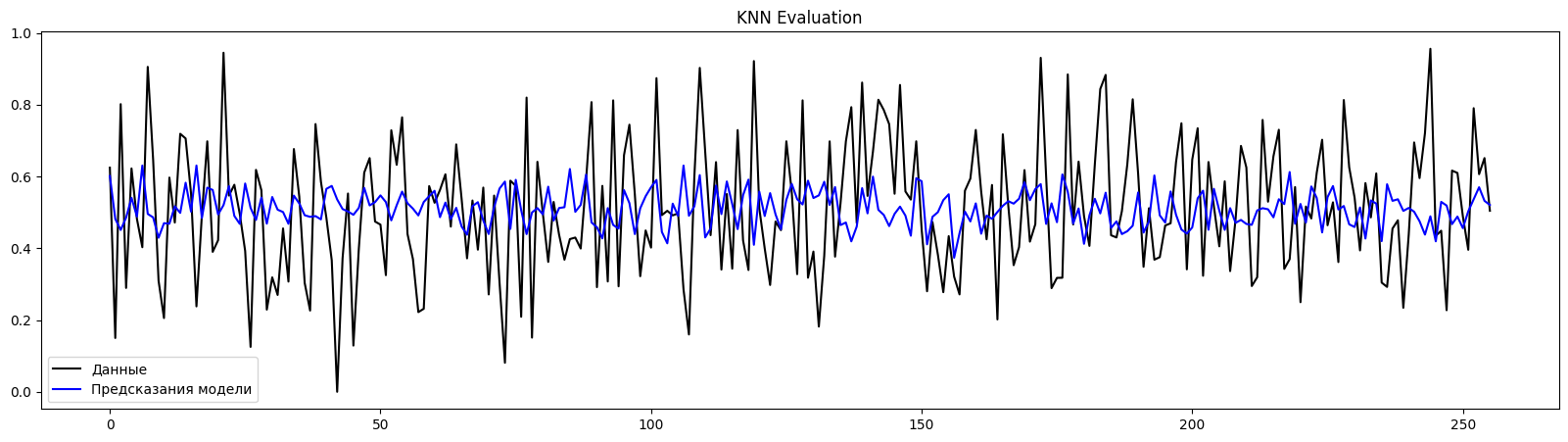
(24)

Евклидово расстояние:

(25)

Используемые гиперпараметры:

* n\_neighbors - количество ближайших соседей, используемых для предсказания. Меньшие значения приводят к более гибкой модели более чувствительной к шуму и склонной к переобучению. Большие значения могут сделать ее более устойчивой, но менее чувствительной к локальным изменениям;
* Weights - определяет будут ли учитываться расстояние расстояния до соседей при предсказании;
* Р - определяет тип расстояния. В данном случае это евклидово и манхеттенское расстояние.

Рисунок 10 - график метода KNN для прочности при растяжении, МПа

1.3. Принцип работы глубокой нейронной сети

Искусственная нейронная сеть представляет собой сложную дифференцируемую функцию, задающую отображение из пространства признаков в пространство ответов. В частном случае нейронная сеть представлена в виде последовательности дифференцируемых параметрических преобразований и представляется в виде вычислительного графа, вершинами которого являются соответствующие преобразования. Параметры сети, такие как веса и смещения, настраиваются одновременно и взаимосвязанно в ходе обучения с целью минимизации ошибки на обучающем наборе данных. Рассмотрим отдельные этапы обучения:

1. Инициализация параметров. Веса (w) и смещения (b) обычно инициализируются случайными значениями, взятыми из нормального распределения с математическим ожиданием 0 и дисперсией.

(26)

2. Прямой проход. На данном этапе обучающая выборка или ее часть (батч) проходит через слои нейронной сети получая на выходе предсказание используемого алгоритма. Для одного слоя формула будет выглядеть следующим образом:

(27)

3. Вычисление функции потерь. Функция потерь измеряет насколько хорошо предсказания сети соответствуют истинным значениям или классам выборки (для настройки гиперпараметров в машинном обучении обычно используется валидационная подвыборка). Для различных задач используются разные функции потерь. Рассмотрим наиболее часто используемые кросс-энтропию (классификация) и MSE (регрессия).

(28)

(29)

4. Обратный проход. На данном этапе вычисляются градиенты функции потерь по параметрам сети с использованием алгоритма обратного распространения ошибки. Для слоя *l* этот этап можно представить следующим образом:

(30)

(31)

(32)

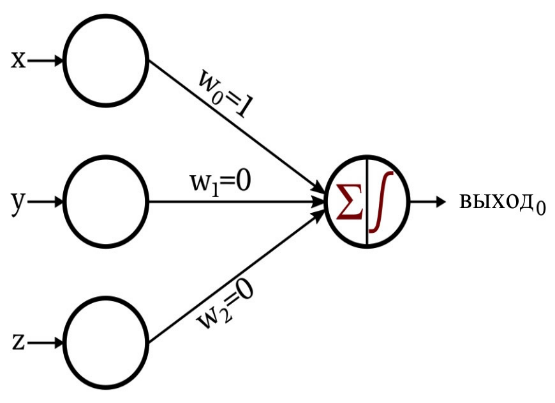
5. Обновление параметров. Параметры обновляются с использованием метода градиентного спуска или одного из его вариаций.

В качестве примера рассмотрим принцип работы перцептрона и полносвязной нейронной сети.

Перцептрон - это базовая модель нейронной сети, состоящая из одного нейрона или одного слоя нейронов и являющийся элементарным строительным блоком более сложных нейронных сетей. Как можно увидеть на рисунке 11, на вход перцептрону поступают данные из входного слоя, умноженные на веса (w). Линейная комбинация этих входных данных проходит через функцию активации. Оптимизация параметров происходит с использованием алгоритма коррекции веса.

(33)

(34)

  
Рис. 11 — Принцип работы перцептрона

Функция активации. Рассмотрим в качестве примера сигмоиду.

(35)

(36)

Выходной сигнал перцептрона.

(37)

Полносвязная нейронная сеть представляет собой архитектуру,   
имеющей несколько слоев: один входной слой, один выходной и произвольное число скрытых слоев, как показано на рисунке 12. Принцип работы схож   
с таковым для перцептрона, но с тем исключением, что вместо метода   
коррекции весов используется алгоритм обратного распространения ошибки.

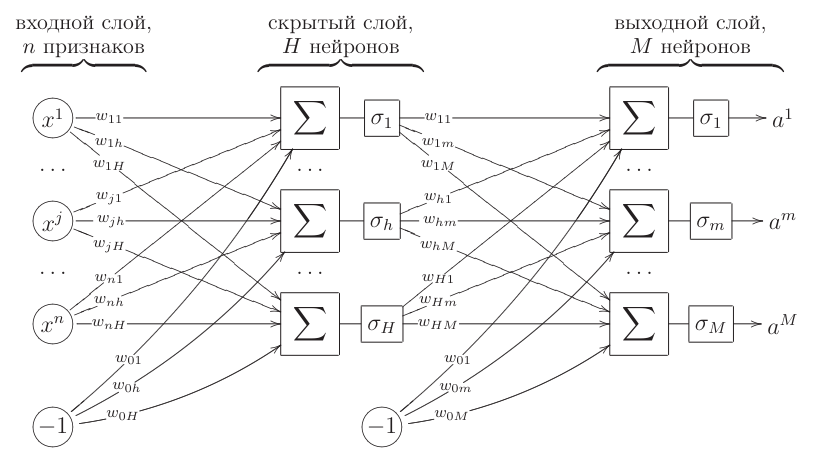


Рис. 12 — Принцип работы многослойного перцептрона

Рассмотрим отдельно используемые метрики качества моделей: R2 или коэффициент детерминации измеряет долю дисперсии, объяснённую моделью, в общей дисперсии целевой переменной.

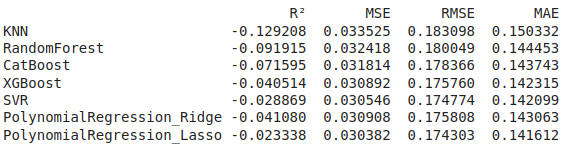


Рисунок 13 – часть кода для результата метрики для всех методов

Если он близок к единице, то модель хорошо объясняет данные, если же он близок к нулю, то качество прогноза идентично средней величине целевой переменной (т.е. очень низкое). Отрицательные значение коэффициента детерминации означают плохую объясняющую способность модели.

MSE (Mean Squared Error) или средняя квадратичная ошибка принимает значениях в тех же единицах, что и целевая переменная. Чем ближе к нулю MSE, тем лучше работают предсказательные качества модели.

RMSE (Корень из среднеквадратичной ошибки) является квадратным корнем из MSE. Эта метрика также измеряет среднее отклонение предсказанных значений от фактических, но в тех же единицах, что и сами данные. RMSE позволяет лучше интерпретировать ошибку в контексте реальных значений и легче воспринимается, поскольку не возводится в квадрат.

MAE (Средняя абсолютная ошибка) вычисляет среднее значение абсолютных разностей между предсказанными и фактическими значениями. MAE измеряет среднюю величину ошибок предсказания, игнорируя направление ошибки (положительное или отрицательное). Она предоставляет ясное представление о том, насколько близки предсказания модели к реальным значениям, и менее чувствительна к выбросам, чем MSE.

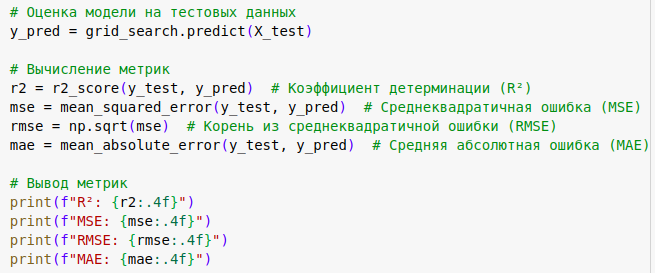
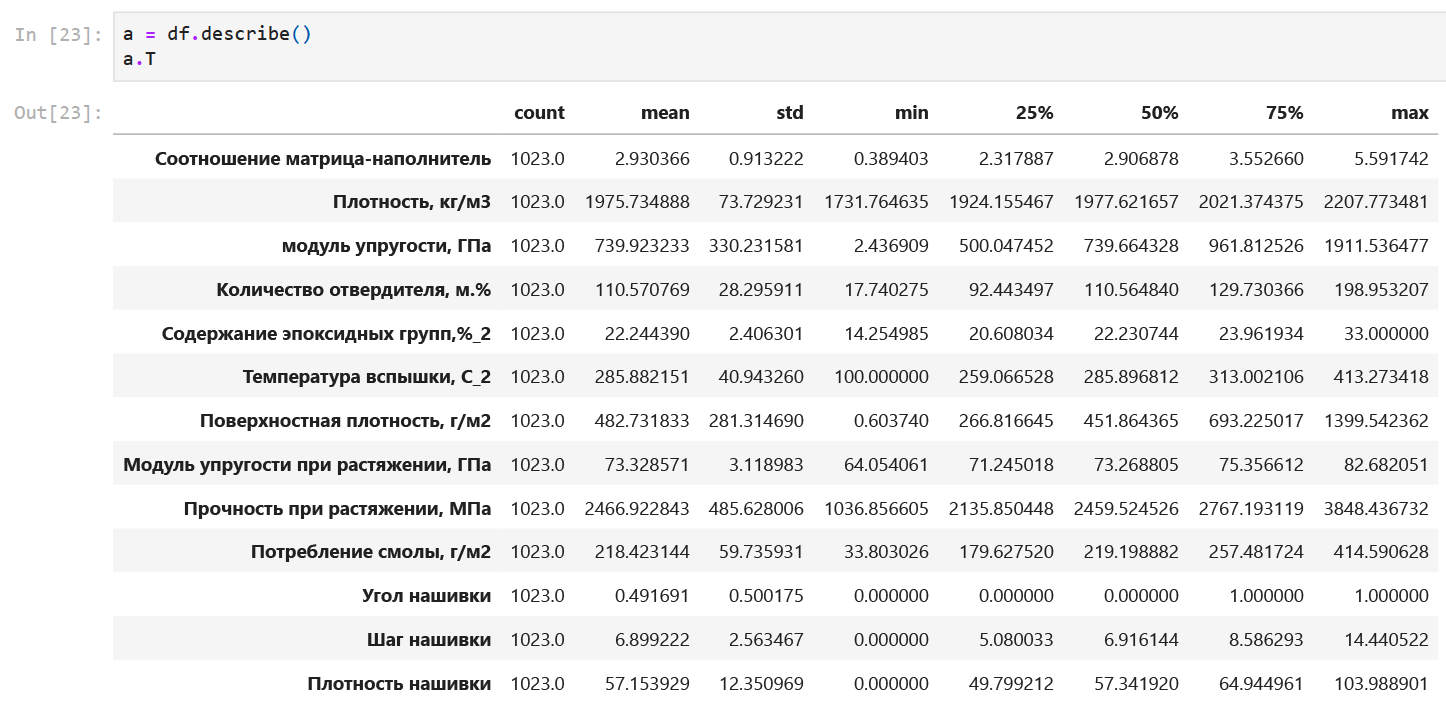


Рисунок 14 - код для вывода различных метрик для метода опорных векторов

## Разведочный анализ данных

Для передачи данных для обучения модели прежде всего необходимо очистить данные от выбросов, заполнить или удалить отсутствующие значения (некоторые модели, к примеру, случайный лес, нечувствительны к отсутствующим значениям).

Рисунок 15 - описательная статистика датасета

Целями разведочного анализа являются: получение представления о характере распределения данных, оценка качества этих данных (наличие выбросов и пропусков), выявление взаимосвязи между признаками (в том числе и проверка на мультиколлинеарность). Анализ также позволяет представить какие из моделей машинного обучения подходят более всего под имеющиеся данные.

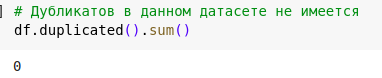


Рисунок 16 - проверка датасета на наличие дубликатов

Из инструментов разведочного анализа использовались: оценка статистических характеристик датасета; гистограммы распределения каждой из переменной (несколько различных вариантов); диаграммы ящика с усами (несколько интерактивных вариантов); попарные графики рассеяния точек (несколько вариантов); график «квантиль-квантиль»; тепловая карта; описательная статистика для каждой переменной; анализ и полное исключение выбросов (5 повторных итераций); проверка наличия пропусков и дубликатов; ранговая корреляция Кендалла и Пирсона.

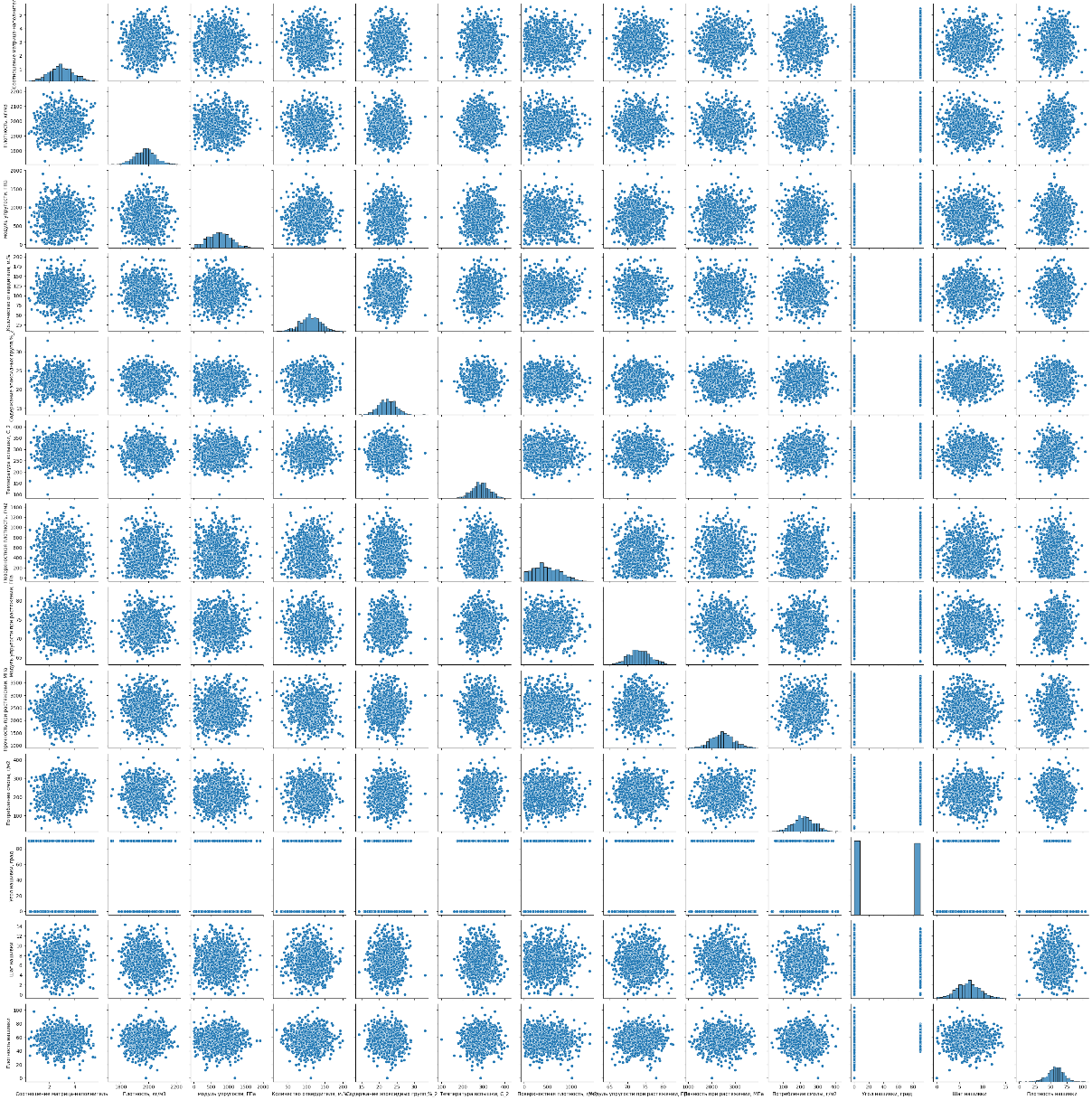


Рисунок 15 - попарные график рассеяния

Полученный графику рассеяния указывает на очень незначительный показатель корреляции между всеми признаками.

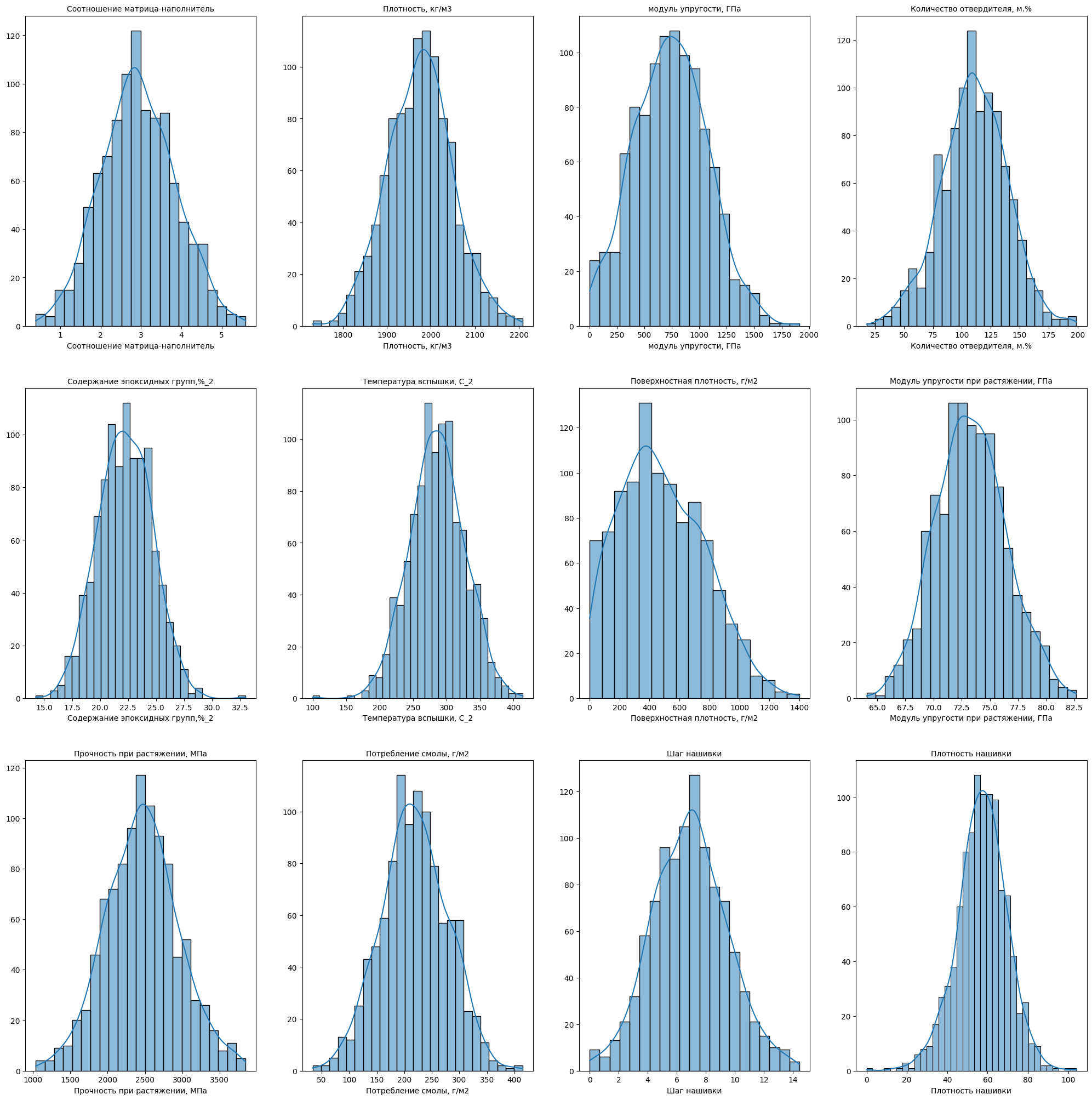


Рисунок 16 - гистограммы распределения

Гистограммы рассеяния указывают на нормальное распределение всех признаков, отсутствие отрицательных значений (что очевидно, учитывая природу данных показателей).

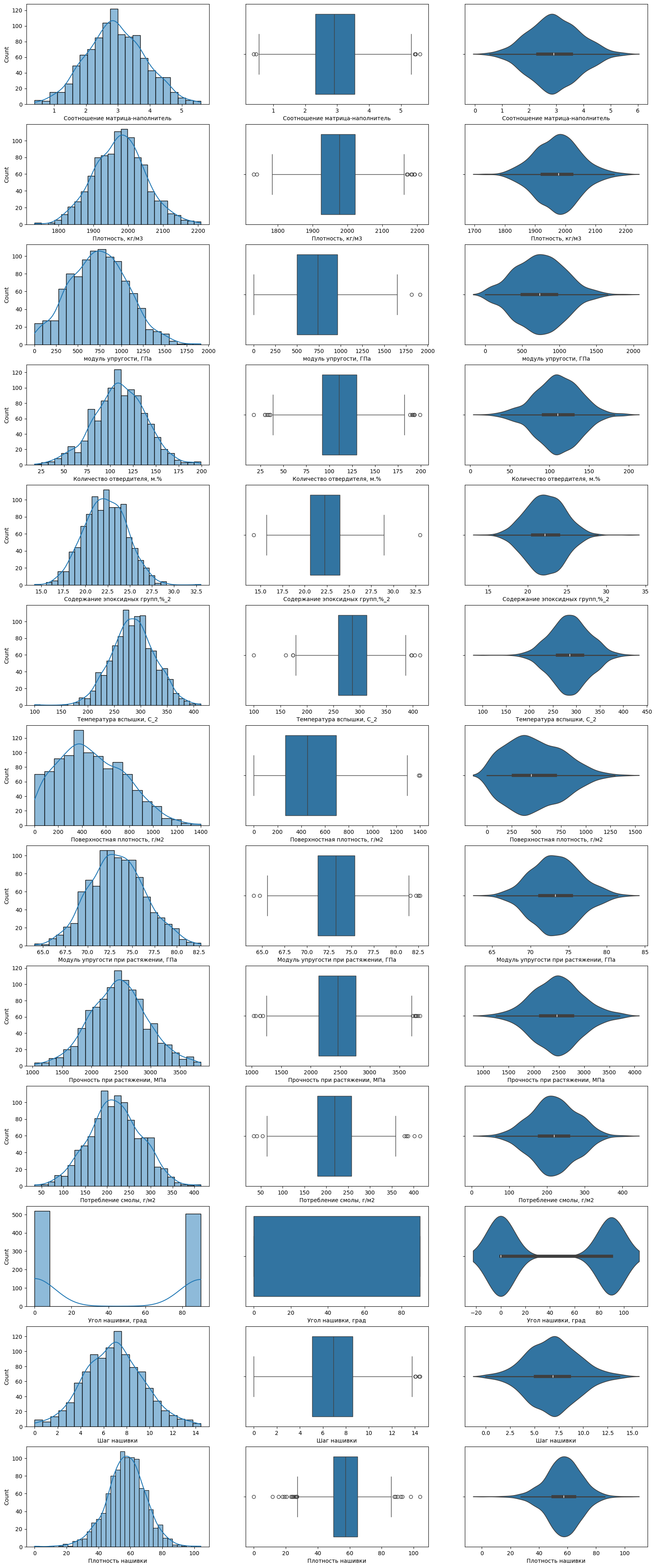


Рисунок 17 - объединенный график гистограмм, боксплотов   
и скрипичных диаграмм

Графики боксплотов указывают на наличие небольшого числа выбросов. Учитывая достаточно небольшое число имеющихся объектов и 13 признаков необходимо решить стоит ли удалять объекты содержащие выбросы или заменить их на медианные или средние значения. Для обнаружения выбросов использовался метод трех сигм и межквартильного расстояния. Так как число выбросов оказалось невелико было принято решения по удалению объектов, в которых данные показатели были обнаружены.

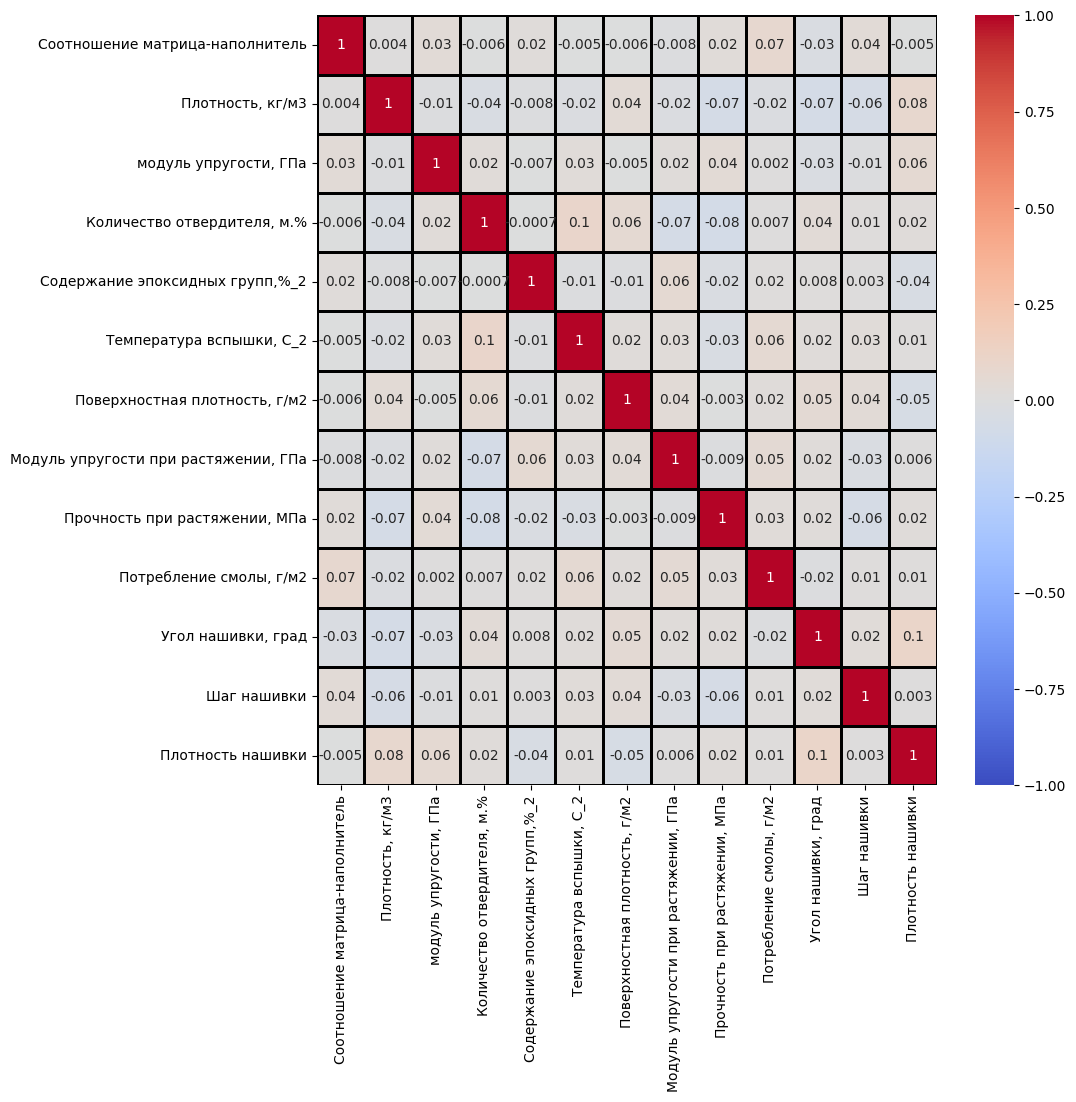


Рисунок 18 - тепловая карта с корреляцией данных

Максимальная корреляция обнаружилась между плотностью нашивки и углом нашивки 0.11, что указывает на очень незначительную зависимость между этими признаками. Корреляция между всеми остальными параметрами стремится к нулю.

# Практическая часть

## Предобработка данных

Необходимо нормализовать значения, для чего применяется MinMaxScaler(), затем применим Normalizer(). Второе даёт нам больше выбросов.

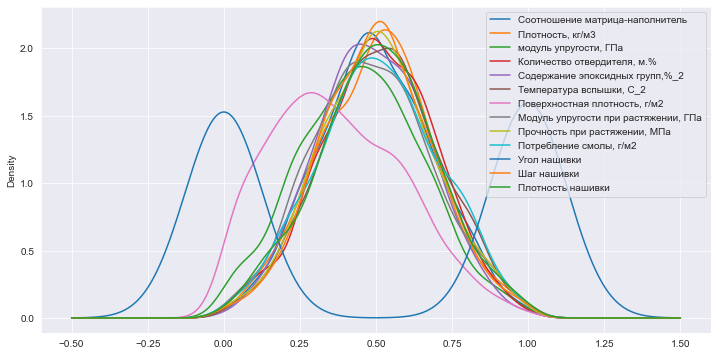
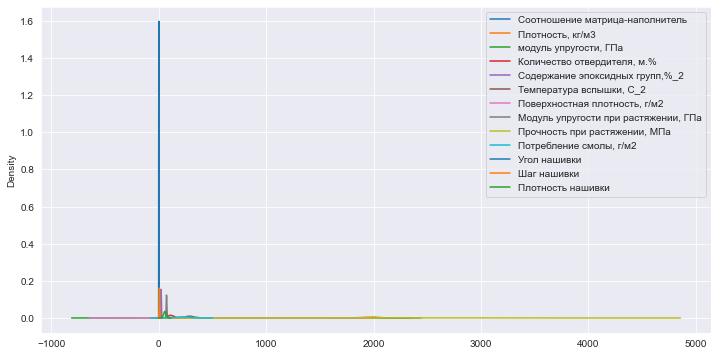


Рисунок 19 - визуализированные данные до и после нормализации

## Разработка и обучение модели

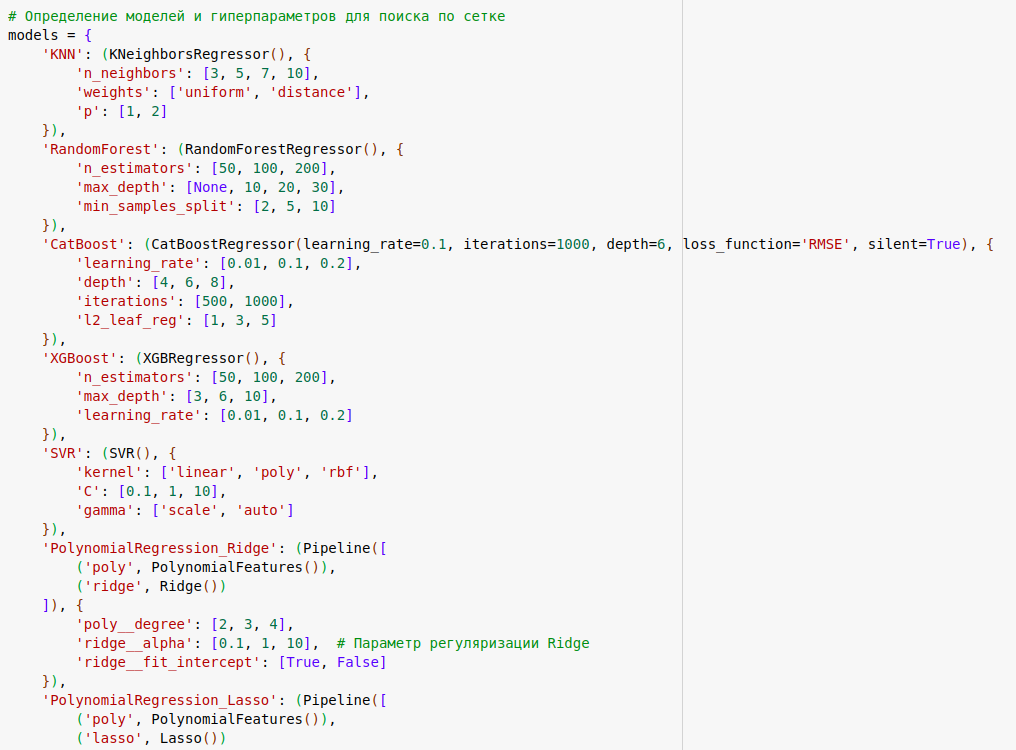
Для моделей машинного обучения использовали две целевые переменные «Прочность при растяжении» и «Модуль упругости при растяжении» по отдельности.

Рисунок 20 - поиск гиперпараметров

Порядок разработки модели для каждого параметра и для каждого выбранного метода можно разделить на следующие этапы: разделение нормализованных данных на обучающую и тестовую выборки (в соотношении 70 на 30%); проверка моделей при стандартных значениях; сравнение с результатами модели, выдающей среднее значение; создание графика; сравнение моделей по метрике МАЕ; поиск сетки гиперпараметров, по которым будет происходить оптимизация модели. В качестве параметра оценки выбран коэффициент детерминации (R2); оптимизация подбора гиперпараметров модели с помощью выбора по сетке и перекрёстной проверки с количеством фолодов равным 5; подстановка оптимальных гиперпараметров в модель и обучение модели на тренировочных данных; оценка полученных данных; сравнение со стандартными значениями.

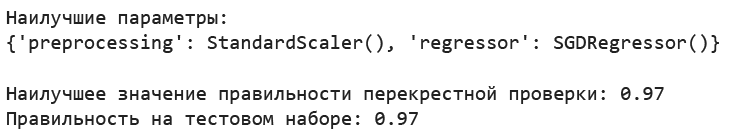


Рисунок 21 - наилучшие гиперпараметры

Модель после настройки гиперпараметров показала результат немного лучше. Однако, ниже, чем базовая модель. Прочность при растяжении и модуль упругости не имеет линейной зависимости. Все использованные модели не справились с задачей. Результат неудовлетворительный. Свойства композитных материалов в первую очередь зависят от используемых материалов.

Таблица 1. Результаты построения и обучения моделей

|  | **Model** | **MAE** | **R2 score** |
| --- | --- | --- | --- |
| **Модуль упругости при растяжении** | KNeighborsRegressor | 2.517884 | -0.001000 |
| **Прочность при растяжении** | KNeighborsRegressor | 368.126111 | -0.009000 |
| **Модуль упругости при растяжении** | PolynomialRegression\_Ridge | 2.546419 | -0.021000 |
| **Прочность при растяжении** | PolynomialRegression\_Ridge | 370.542618 | -0.021000 |
| **Модуль упругости при растяжении** | RandomForestRegressor | 2.597992 | -0.061000 |
| **Прочность при растяжении** | RandomForestRegressor | 370.967356 | -0.013000 |
| **Модуль упругости при растяжении** | RandomForestRegressor | 2.514920 | -0.000084 |
| **Прочность при растяжении** | RandomForestRegressor | 367.611428 | -0.009000 |

## Тестирование модели

После обучения моделей была проведена оценка точности этих моделей на обучающей и тестовых выборках при использовании метрик указанных выше. Результат оказался неудовлетворительный, поскольку результат коэффициента детерминации колебался в диапазоне 0. Наиболее вероятной причиной кажется используемый датасет.

## Создание архитектуры полносвязной нейронной сети для рекомендации соотношения «матрица – наполнитель».

В процессе обучения нейронной сети происходит подбор оптимальных параметров (весов) сети при помощи обратного распространения ошибки. Для построения собственной архитектуры полносвязной нейронной сети используем библиотеку tensorflow.



Рисунок 24 - создание нейронной сети

Определим параметры, поищем оптимальные параметры, посмотрим на результаты. С помощью KerasClassifier выйдем на наилучшие параметры для нашей нейронной сети и построим окончательную нейросеть.

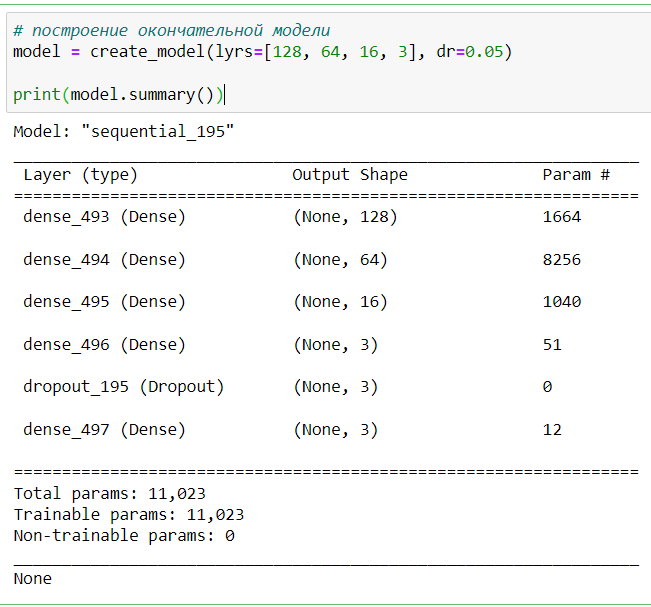


Рисунок 25 - архитектура нейросети

Обучим и оценим модель, посмотрим на потери, зададим функцию для визуализации факт/прогноз для результатов моделей.

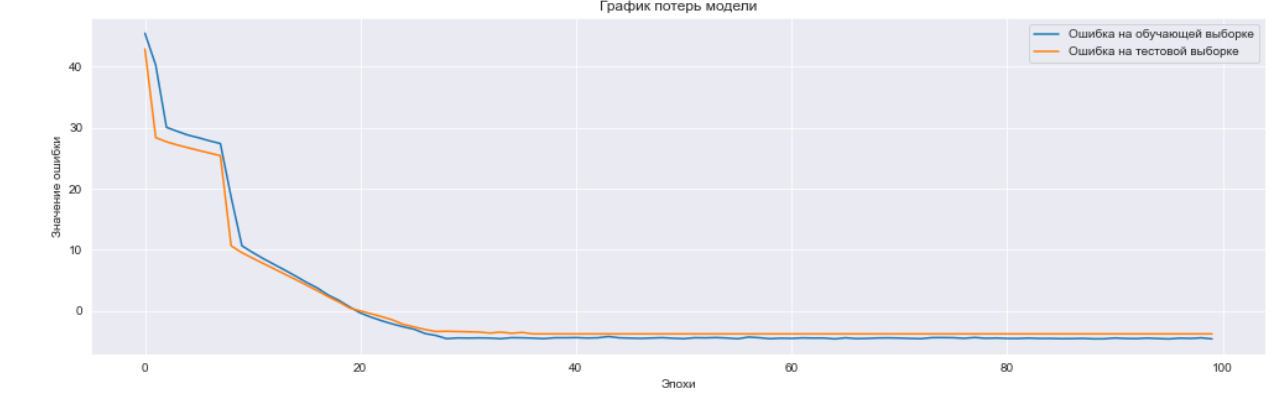
Рисунок 26 - график потерь

Рисунок 27 - тестовые и прогнозные значения модели 1

Не удовлетворившись таким результатом, создадим другую простую модель глубокого обучения с другой архитектурой. Обучим её, посмотрим на потери, оценим MSE, построим график.

## Разработка приложения

Приложение выдает прогноз для соотношения «матрица – наполнитель».   
Однако, поскольку обученная модель на тестовой выборке показывает низкий результат, доверять подобному результату не следует.



Рисунок 28 - результат работы приложения

Приложение представляет собой основной файл Flask, шаблон html - страницы и сохраненная модель с весами полносвязной нейронной сети.

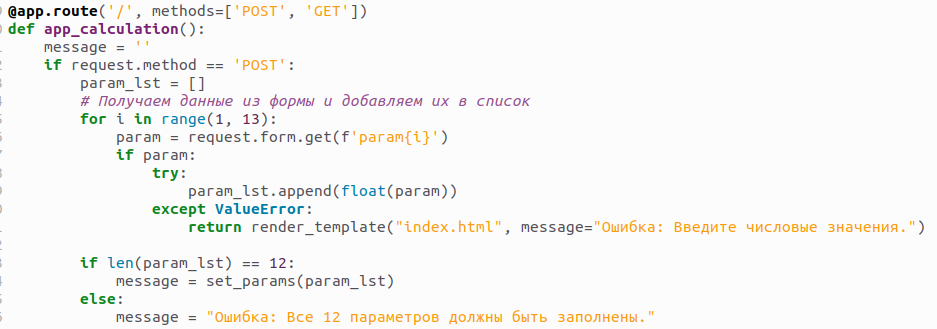


Рисунок 29 - часть кода приложения

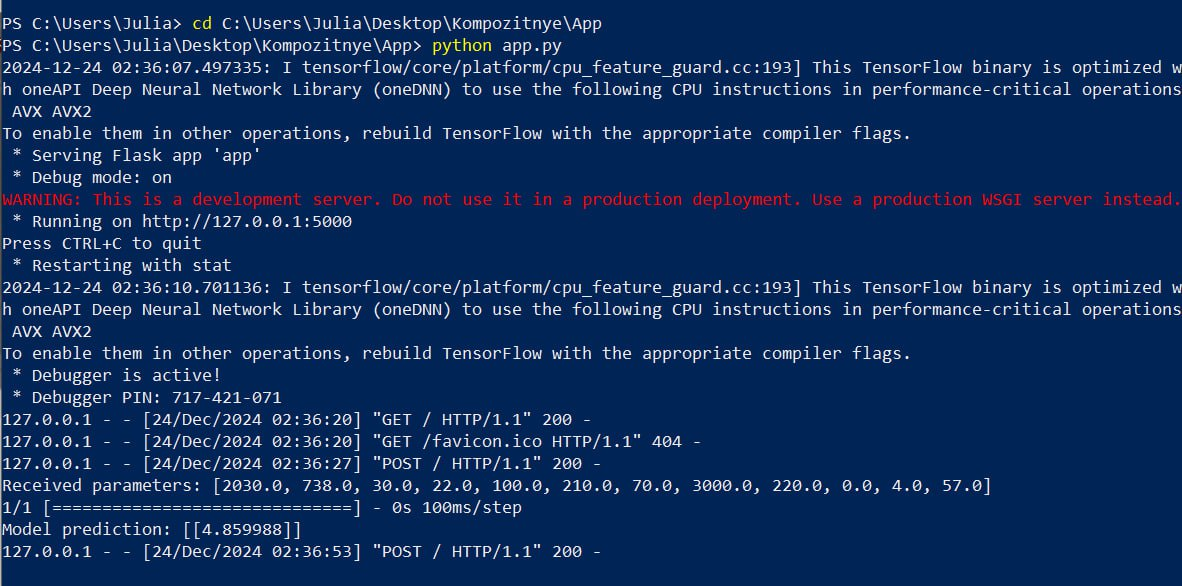
При запуске приложения, пользователь переходит на: <http://127.0.0.1:5000/>.

Рисунок 30 - запуск приложения

В открывшемся окне пользователю необходимо ввести требуемые значения в соответствующие ячейки и нажать «Готово».

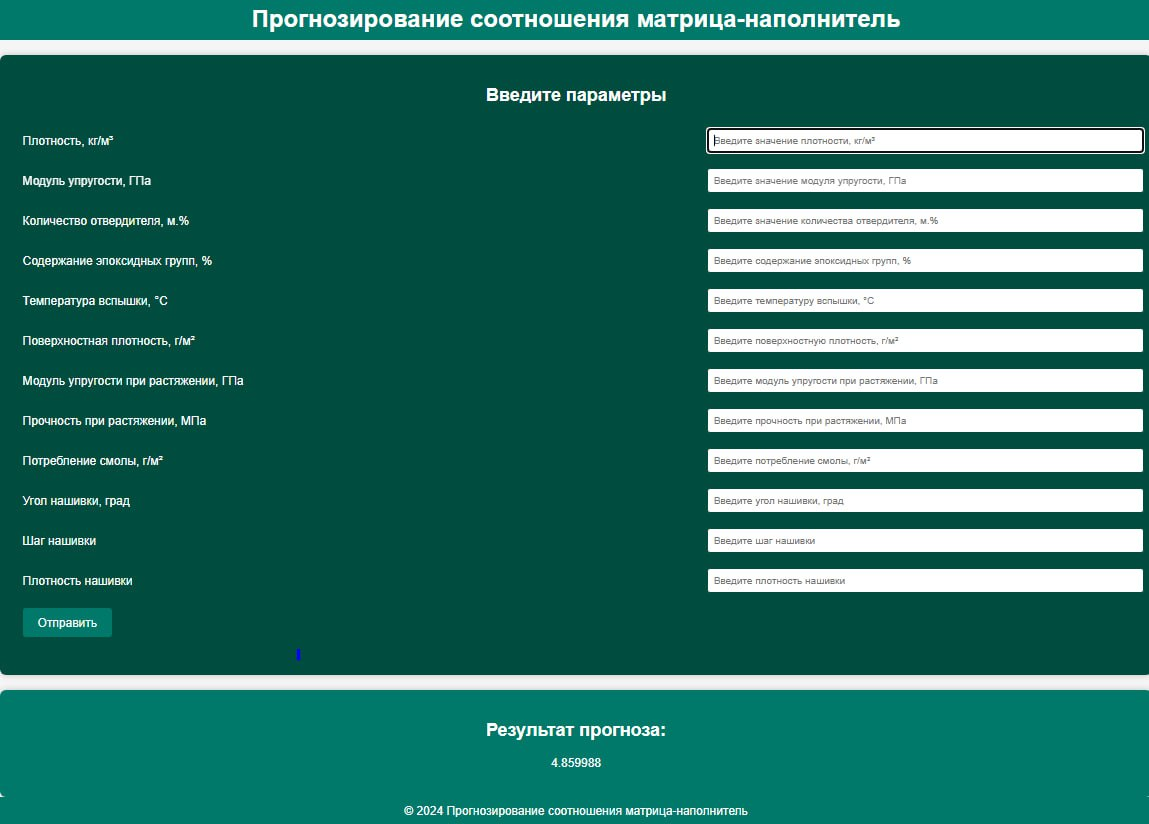


Рисунок 31- скриншот пользовательского приложения

На выходе получаем прогноза для значения параметра «Соотношение «матрица – наполнитель»».

## Создание удалённого репозитория и загрузка

Репозиторий создан на github.com по адресу: <https://github.com/maxshevik/Kompozitnye>

Рисунок 32 - часть страницы на github.com

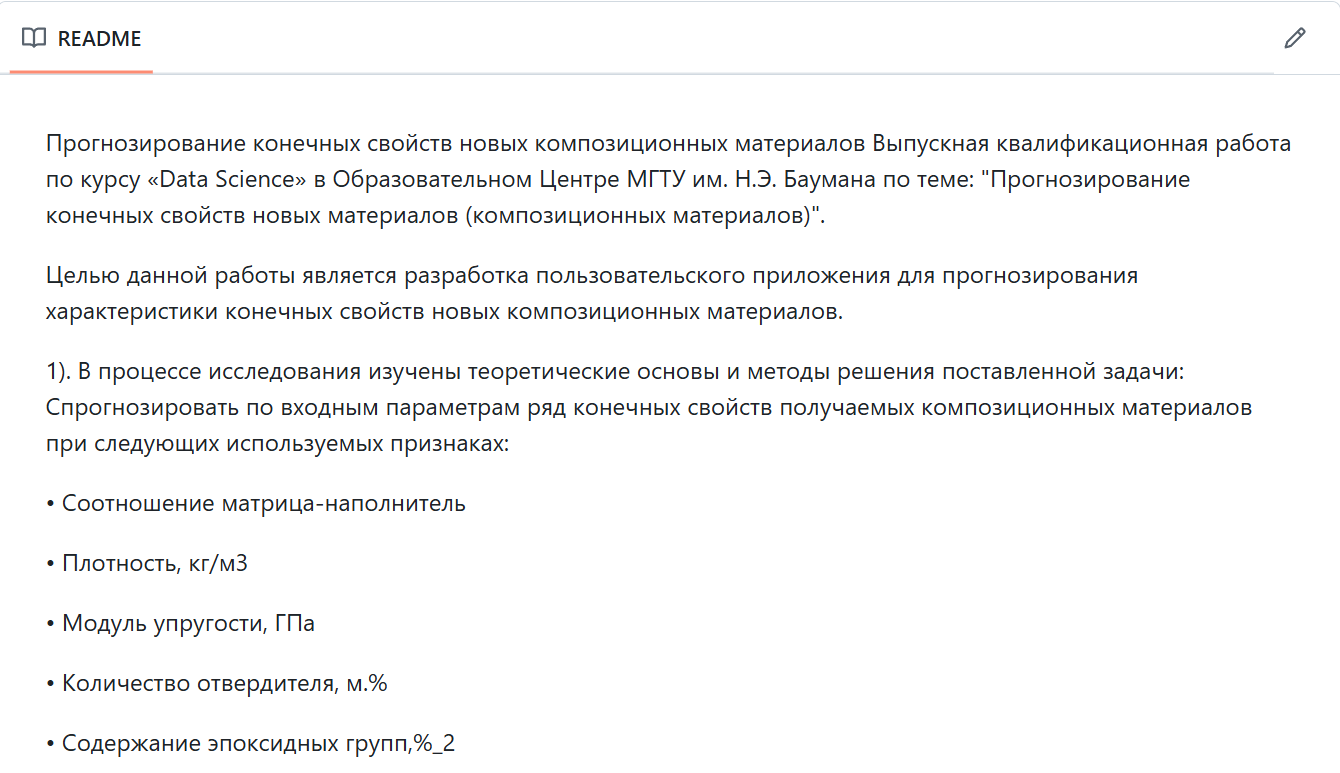


Рисунок 33 - часть созданного файла README

# Заключение

В данной работе выполнены следующие выводы: распределение признаков объединенного датасета имеют нормальное распределение за исключением   
признака “угол нашивки”, имеющей бинарное распределение. Коэффициент корреляции между всеми признаками близок к нулю. Примененные модели, включая   
полносвязную нейронную сеть не позволяют получить сколь-либо достоверных прогнозов.

Был сделан вывод, что невозможно определить из свойств материалов   
соотношение «матрица – наполнитель». Данный факт не указывает на то, что прогнозирование характеристик композитных материалов на основании предоставленного набора данных невозможно, но может указывать на недостатки базы данных, подходов, использованных при прогнозе, необходимости пересмотра инструментов для прогнозирования.

Прогнозирование признаков материалов без изучения соответствующей   
литературы не всегда реализуемо. Ни одна из использованных моделей не показала удовлетворительного результата. Для улучшения прогноза модели необходимо   
получить дополнительные данные, желательно в кооперации с химиками, получить новые данные или скорректировать существующие.

# Список используемой литературы и веб ресурсы.

1. Alex Maszański. Метод k-ближайших соседей (k-nearest neighbour): –  
   Режим доступа: <https://proglib.io/p/metod-k-blizhayshih-sosedey-k-nearest-neighbour-2021-07-19>. (дата обращения: 15.09.2024)
2. Гафаров, Ф.М., Галимянов А.Ф. Искусственные нейронные сети и приложения: учеб. пособие /Ф.М. Гафаров, А.Ф. Галимянов. – Казань: Издательство Казанского университета, 2018. – 121 с.
3. Грас Д. Data Science. Наука о данных с нуля: Пер. с англ. - 2-е изд.,   
   перераб. и доп. - СПб.: БХВ-Петербурr, 2021. - 416 с.: ил.
4. Документация по библиотеке keras: – Режим доступа:   
   https://keras.io/api/.(дата обращения: 17.09.2024).
5. Документация по библиотеке numpy: – Режим доступа:   
   https://numpy.org/doc/1.22/user/index.html#user. (дата обращения: 17.09.2024).
6. Документация по библиотеке pandas: – Режим доступа:   
   https://pandas.pydata.org/docs/user\_guide/index.html#user-guide. (дата обращения: 17.09.2024).
7. Документация по библиотеке scikit-learn: – Режим доступа:   
   https://scikit-learn.org/stable/user\_guide.html. (дата обращения: 15.09.2024).
8. Документация по библиотеке seaborn: – Режим доступа:   
   https://seaborn.pydata.org/tutorial.html. (дата обращения: 18.09.2024).
9. Документация по библиотеке Tensorflow: – Режим доступа: <https://www.tensorflow.org/overview> (дата обращения: 18.09.2024).
10. Документация по языку программирования python: – Режим доступа:    
    https://docs.python.org/3.8/index.html. (дата обращения: 15.09.2024).
11. Иванов Д.А., Ситников А.И., Шляпин С.Д – Композиционные материалы: учебное пособие для вузов, 2019. 13 с.
12. Краткий обзор алгоритма машинного обучения Метод Опорных Векторов (SVM) – Режим доступа: <https://habr.com/ru/post/428503/> (дата обращения 17.09.2024)
13. Ларин А. А., Способы оценки работоспособности изделий из композиционных материалов методом компьютерной томографии, Москва, 2013, 148 с.
14. Материалы конференции: V Всероссийская научно-техническая конференция «Полимерные композиционные материалы и производственные технологии нового поколения», 19 ноября 2021 г.
15. Миронов А.А. Машинное обучение часть I ст.9 – Режим доступа: <http://is.ifmo.ru/verification/machine-learning-mironov.pdf>. (дата обращения 17.09.2024)
16. Плас Дж. Вандер, Python для сложных задач: наука о данных и машинное обучение. Санкт-Петербург: Питер, 2018, 576 с.
17. Реутов Ю.А.: Прогнозирование свойств полимерных композиционных материалов и оценка надёжности изделий из них, Диссертация на соискание учёной степени кандидата физико-математических наук, Томск 2016.
18. Роббинс, Дженнифер. HTML5: карманный справочник, 5-е издание.: Пер. с англ. - М.: ООО «И.Д. Вильямс»: 2015. - 192 с.: ил.
19. Руководство по быстрому старту в flask: – Режим доступа: <https://flask-russian-docs.readthedocs.io/ru/latest/quickstart.html>. (дата обращения: 20.09.2024)
20. Силен Дэви, Мейсман Арно, Али Мохамед. Основы Data Science и Big Data. Python и наука о данных. – СПб.: Питер, 2017. – 336 с.: ил.
21. Скиена, Стивен С. С42 Наука о данных: учебный курс.: Пер. с англ. - СПб.: ООО "Диалектика", 2020. - 544 с. : ил.
22. Справочник по композиционным материалам: в 2 - х кн. Кн. 2 / Под ред. Дж. Любина; Пер. с англ. Ф. Б. Геллера, M. М. Гельмонта; Под ред. Б. Э. Геллера - М.: Машиностроение, 1988. - 488 с. : ил;
23. Траск Эндрю. Грокаем глубокое обучение. – СПб.: Питер, 2019. – 352 с.: ил.
24. Чун-Те Чен и Грейс Х. Гу. Машинное обучение для композитных материалов (март 2019г.) – Режим доступа: [https://www.cambridge.org/core/journals/mrs-communications/article/machine- learning-for-composite-materials/F54F60AC0048291BA47E0B671733ED15](https://www.cambridge.org/core/journals/mrs-communications/article/machine-%20learning-for-composite-materials/F54F60AC0048291BA47E0B671733ED15). (дата обращения 15.09.2024)