МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ

РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение

высшего образования

«Московский государственный технический университет имени Н.Э. Баумана

(национальный исследовательский университет)»

**ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА**

**по курсу**

«Data Science Pro»

Слушатель Лосяков Сергей Геннадиевич

Москва, 2025

Содержание:

[**Введение** 3](#_Toc1)

[**1. Аналитическая часть** 4](#_Toc2)

[**1.1. Постановка задачи.** 4](#_Toc3)

[**1.2. Описание используемых методов.** 6](#_Toc4)

[**1.3. Разведочный анализ данных.** 8](#_Toc5)

[**2. Практическая часть** 10](#_Toc6)

[**2.1. Предобработка данных.** 10](#_Toc7)

[**2.2. Разработка и обучение модели.** 26](#_Toc8)

[**2.3. Тестирование модели.** 27](#_Toc9)

[**2.4. Написать нейронную сеть, которая будет рекомендовать соотношение матрицы.**  28](#_Toc10)

[**2.5. Разработка приложения.** 30](#_Toc11)

[**2.6. Создание удаленного репозитория и загрузка результатов работы на него.** 31](#_Toc12)

[**Заключение** 32](#_Toc13)

[**Библиографический список** 34](#_Toc14)

# **Введение**

Современное материаловедение активно развивается в направлении создания новых композитных материалов с заданными свойствами. Композиционные материалы представляют собой многокомпонентные системы, сочетающие в себе преимущества различных веществ — например, прочность волокон и пластичность полимерной матрицы. Благодаря этому композиты находят широкое применение в авиации, строительстве, машиностроении и других высокотехнологичных отраслях.

Однако определение конечных свойств таких материалов по характеристикам исходных компонентов остаётся сложной задачей. Традиционный подход, основанный на физических испытаниях, требует значительных затрат времени, ресурсов и часто ограничен количеством возможных комбинаций. В связи с этим всё более актуальными становятся методы цифрового материаловедения, использующие машинное обучение и анализ данных для построения прогнозных моделей.

Целью данной работы является разработка алгоритмов машинного обучения и нейронных сетей, способных предсказывать ключевые механические характеристики композиционных материалов — такие как модуль упругости и прочность при растяжении — по данным о составе и условиях их производства. Помимо этого, в рамках проекта создаётся рекомендационная система, определяющая оптимальное соотношение компонентов композита, а также программное приложение, реализующее разработанные модели. Работа опирается на реальные производственные данные, предоставленные Центром НТИ «Цифровое материаловедение» при МГТУ им. Н.Э. Баумана.

Использование предложенных решений позволит сократить количество натурных испытаний, ускорить этапы разработки новых материалов и создать цифровых двойников для моделирования их свойств в виртуальной среде.

# **1. Аналитическая часть**

## **1.1. Постановка задачи.**

Современное производство композиционных материалов сталкивается с задачей прогнозирования механических свойств новых образцов без проведения длительных и дорогостоящих испытаний. Несмотря на то, что известны физико-химические параметры компонентов композита (матрицы и наполнителя), взаимосвязь между их свойствами и итоговыми характеристиками материала сложна и нелинейна. В данной работе рассматривается задача построения модели машинного обучения, способной по свойствам компонентов и параметрам производства предсказывать два ключевых механических показателя композита:

1. Модуль упругости при растяжении (ГПа);
2. прочность при растяжении (МПа).

Дополнительно решается задача рекомендации оптимального соотношения "матрица-наполнитель" на основе нейронной сети.

В качестве исходных данных используется таблица, содержащая 1023 записи, каждая из которых представляет собой уникальную комбинацию параметров и соответствующих выходных характеристик. Все признаки числовые. В датасете представлены следующие переменные:

Входные переменные (признаки):

1. Соотношение матрица-наполнитель
2. Плотность, кг/м³
3. Модуль упругости, ГПа
4. Количество отвердителя, м.%
5. Содержание эпоксидных групп, %
6. Температура вспышки, °C
7. Поверхностная плотность, г/м²
8. Потребление смолы, г/м²
9. Угол нашивки, градусы
10. Шаг нашивки
11. Плотность нашивки

Выходные переменные (целевые):

1. Модуль упругости при растяжении, ГПа
2. Прочность при растяжении, МПа

Проверка на пропуски показала, что во всех колонках отсутствуют пропущенные значения. Типы данных корректны и представлены как float64 или int64. Это указывает на хорошее качество данных и возможность перехода к машинному обучению без этапа восстановления пропусков.

Описательная статистика показала, что диапазоны значений по признакам находятся в адекватных пределах. Однако в некоторых переменных, например в "Температура вспышки, °C" и "Поверхностная плотность", наблюдаются существенные различия между минимумами и максимумами, что может указывать на наличие выбросов. Также следует отметить, что большинство признаков имеют широкое распределение, что важно учитывать при масштабировании данных.

Построенная корреляционная матрица показала, что линейная корреляция между входными признаками и целевыми переменными слабо выражена (коэффициенты корреляции находятся преимущественно в диапазоне от 0.01 до 0.07). Это свидетельствует о высокой сложности зависимости между параметрами и необходимости применения нелинейных методов моделирования, таких как ансамблевые модели или нейросети.

Таким образом, задача заключается в построении регрессионной модели, способной учитывать сложные взаимосвязи между характеристиками компонентов композиционного материала и его итоговыми механическими свойствами, с последующей интеграцией модели в прикладное программное обеспечение.

## **1.2. Описание используемых методов.**

Для решения поставленной задачи прогнозирования механических характеристик композиционных материалов были применены два класса методов: ансамблевые модели машинного обучения и искусственные нейронные сети. Ниже приводится краткое описание каждого из них, включая достоинства, недостатки и обоснование их применения.

Ансамблевые методы объединяют предсказания нескольких базовых моделей для получения более стабильного и точного результата. В данной работе использовалась модель StackingRegressor, которая объединяет несколько алгоритмов (ElasticNet, Random Forest, KNN), которые были отобраны после сравнения (таблица 1.1) передавая их выходы в финальную модель (Linear Regression).

Достоинства:

1. Повышенная устойчивость к переобучению за счёт объединения слабых моделей.
2. Возможность учитывать различные типы взаимосвязей за счёт разнообразия базовых моделей.
3. Простота интерпретации при использовании линейного финального регрессора.

Недостатки:

1. Высокая вычислительная сложность и время обучения.
2. Требует тщательной настройки и валидации.

Области применения:

1. Моделирование с ограниченным объёмом данных и низкой корреляцией между признаками и целевой переменной.
2. Ситуации, где важно агрегировать слабые признаки в более устойчивое предсказание.

Таблица 1 – Сравнение методов

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Тип модели | Достоинства | Недостатки | Условия, при которых работает хорошо |
| ElasticNet | Линейная модель | Сбалансированная регуляризация, устойчив к мультиколлинеарности | Требует подбора параметров α и l1\_ratio | Линейные зависимости с шумом, много коррелированных признаков |
| Random Forest | Ансамблевый | Не переобучается, устойчив к выбросам, работает с нечисловыми признаками | Большое потребление памяти, не интерпретируем | Сложные, нелинейные зависимости, небольшие датасеты |
| KNN Regressor | Алгоритм ближайших соседей | Прост в реализации, не требует обучения модели | Зависим от масштаба, чувствителен к выбросам | Когда данные компактны и имеют локальную структуру |
| Ridge | Линейная модель | Хорош при мультиколлинеарности, стабильен | Менее гибок, чем ElasticNet и Lasso | Простые линейные задачи, где важна устойчивость |
| Lasso | Линейная модель | Отбрасывает ненужные признаки, простота | Может нестабильно работать при сильной корреляции признаков | Малое количество признаков, где важно выбрать важные |
| Gradient Boosting Regressor | Ансамблевый | Высокая точность, гибкость | Медленный, требует настройки, может переобучаться | Когда нужен максимум качества, при большом количестве данных |
| Linear Regression | Линейная модель | Интерпретируемый, быстрый | Работает только при наличии линейной зависимости | Как базовая модель, начальный анализ |

Нейросеть использовалась для решения задачи регрессии — прогнозирования значения соотношения матрица-наполнитель на основе других физических характеристик. Архитектура сети состоит из трёх полносвязных слоёв с функцией активации tanh.

Достоинства:

1. Способность обучать сложные нелинейные зависимости даже при низкой линейной корреляции.
2. Высокая гибкость и адаптируемость под структуру данных.

Недостатки:

1. Необходимость нормализации данных и большого числа параметров настройки.
2. Сложность интерпретации модели.
3. Уязвимость к переобучению на малых выборках.

Области применения:

1. Сложные регрессионные задачи, где простые модели не дают удовлетворительного качества.
2. Прогнозирование в условиях слабой линейной зависимости между переменными.

## **1.3. Разведочный анализ данных.**

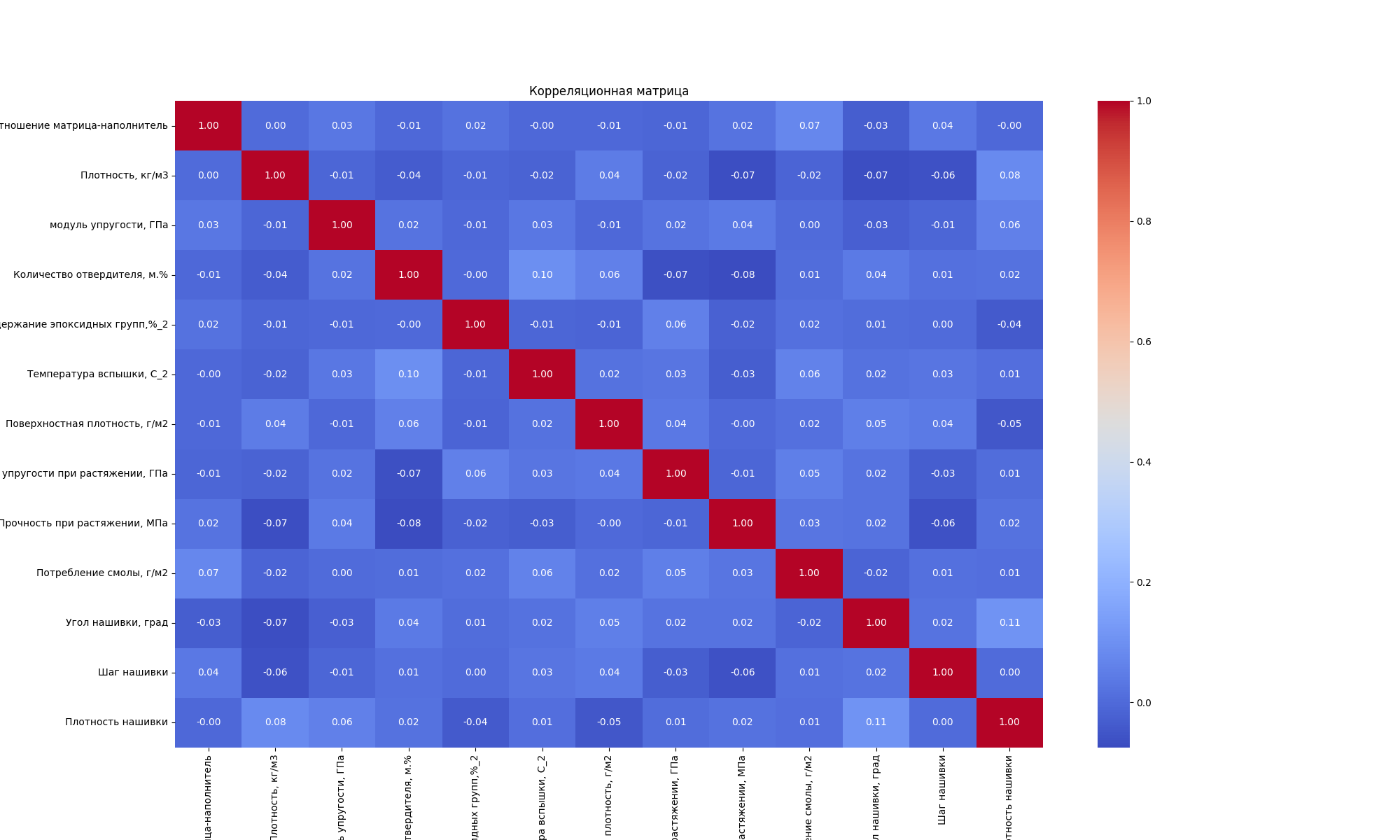
Разведочный анализ данных представляет собой начальный этап анализа, целью которого является выявление закономерностей, структуры данных, а также наличие выбросов, пропущенных значений и особенностей распределений признаков. В рамках данного проекта были применены следующие методы:

Анализ распределения признаков – для количественных признаков использовались гистограммы, boxplot-графики и описательная статистика (среднее, медиана, минимум, максимум, стандартное отклонение). Были выявлены признаки с сильными выбросами, например, плотность и температура, которые искажали общее распределение.

Обработка выбросов – применялся метод межквартильного размаха (IQR): значения за пределами [Q1 – 1.5·IQR, Q3 + 1.5·IQR] были обрезаны или исключены из анализа. Это позволило стабилизировать обучение моделей и снизить влияние экстремальных значений на качество предсказания.

Корреляционный анализ – построена матрица корреляций между признаками и целевыми переменными (механические характеристики). Было выявлено, что некоторые признаки имеют низкую корреляцию с целевыми (|r| < 0.04) и были исключены из дальнейшего анализа, поскольку не вносили полезную информацию в предсказание.

Рисунок 1 – Тепловая матрица корреляции



Визуализация – для оценки взаимосвязей между признаками применялись тепловые карты (heatmap) и диаграммы рассеяния (pairplot). Это позволило оценить структуру данных и убедиться в отсутствии сильной мультиколлинеарности между оставшимися признаками.

Масштабирование данных – для моделей, чувствительных к масштабу признаков (например, KNN и нейронные сети), использовались StandardScaler и MinMaxScaler. Это позволило привести признаки к единому диапазону и обеспечить устойчивость моделей.

# **2. Практическая часть**

## **2.1. Предобработка данных.**

На этапе предобработки данных была проведена нормализация числовых признаков с использованием метода Min-Max масштабирования. Это позволило привести значения всех признаков к единому масштабу в диапазоне от 0 до 1. Такая процедура необходима для корректной работы большинства алгоритмов машинного обучения, особенно чувствительных к масштабу входных данных (например, алгоритмы градиентного спуска, методы на основе расстояний и нейронные сети).

Перед нормализацией для каждого признака были построены гистограммы распределения значений, что позволило выявить особенности распределения и потенциальные выбросы. Ниже представлены примеры таких графиков.

Рисунок 2 – Распределение признака до нормализации

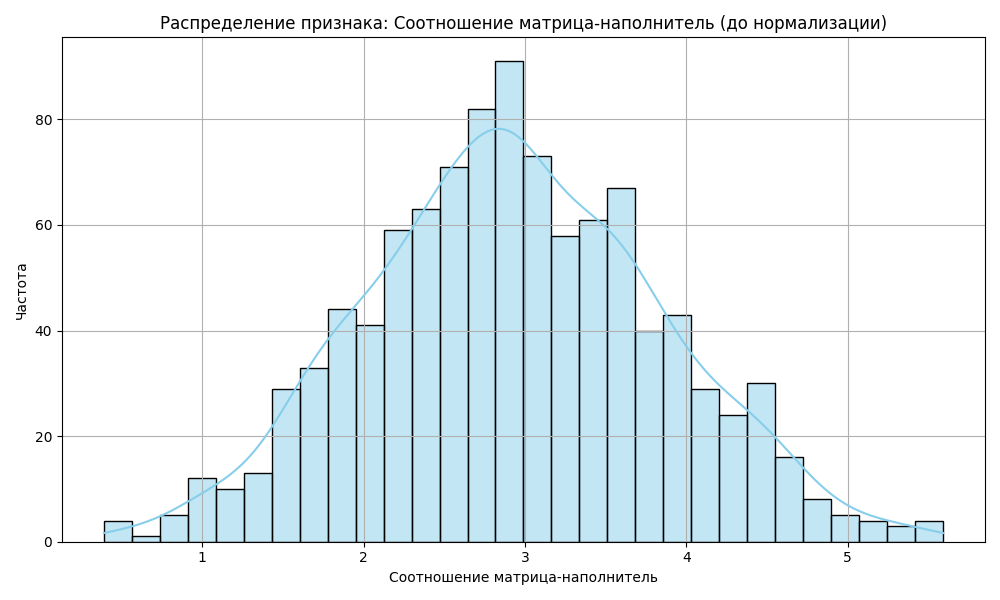


Рисунок 3 – Распределение признака до нормализации

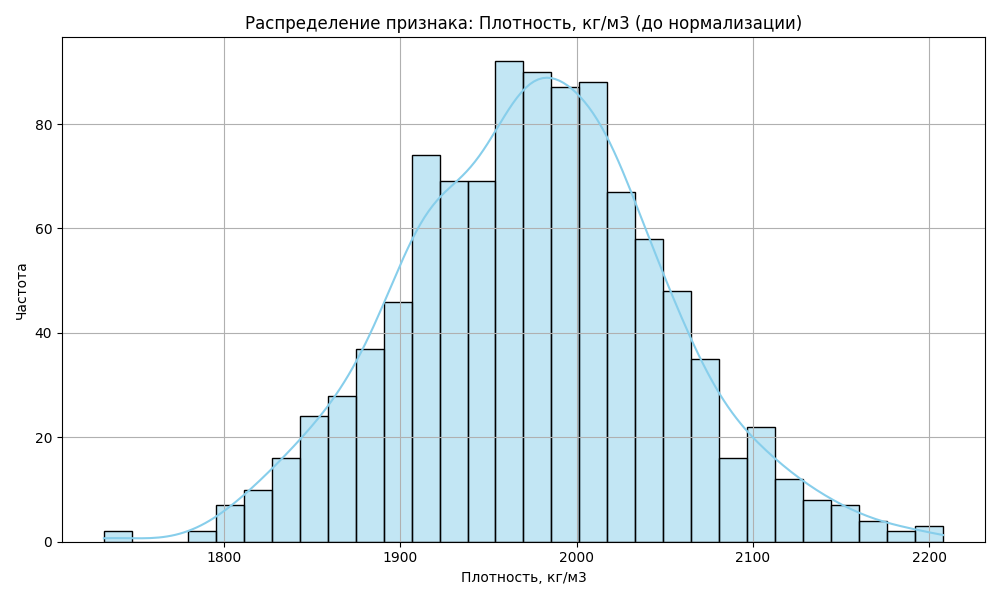


Рисунок 4 – Распределение признака до нормализации



Рисунок 5 – Распределение признака до нормализации



Рисунок 6 – Распределение признака до нормализации



Рисунок 7 – Распределение признака до нормализации

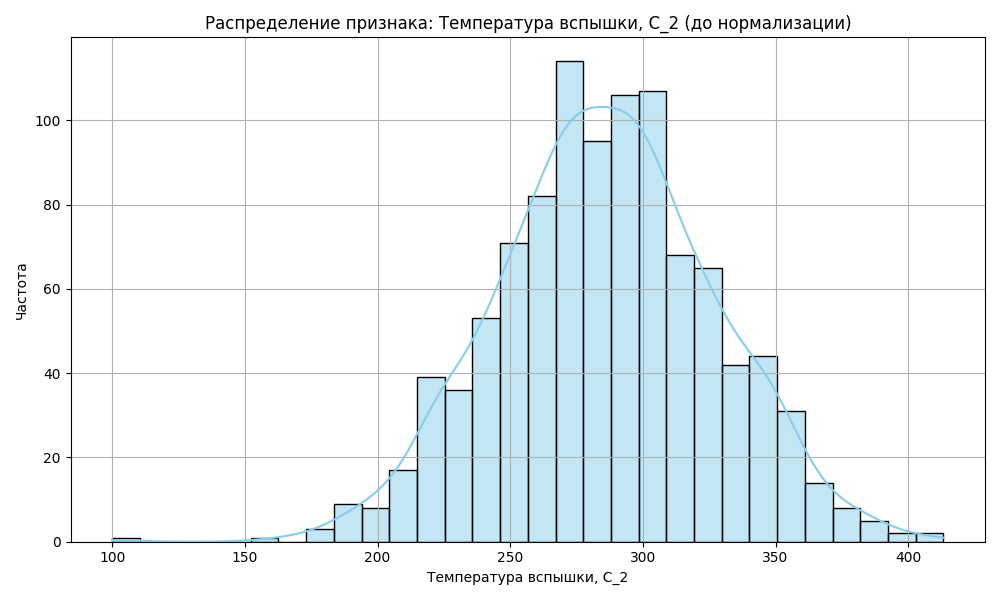


Рисунок 8 – Распределение признака до нормализации

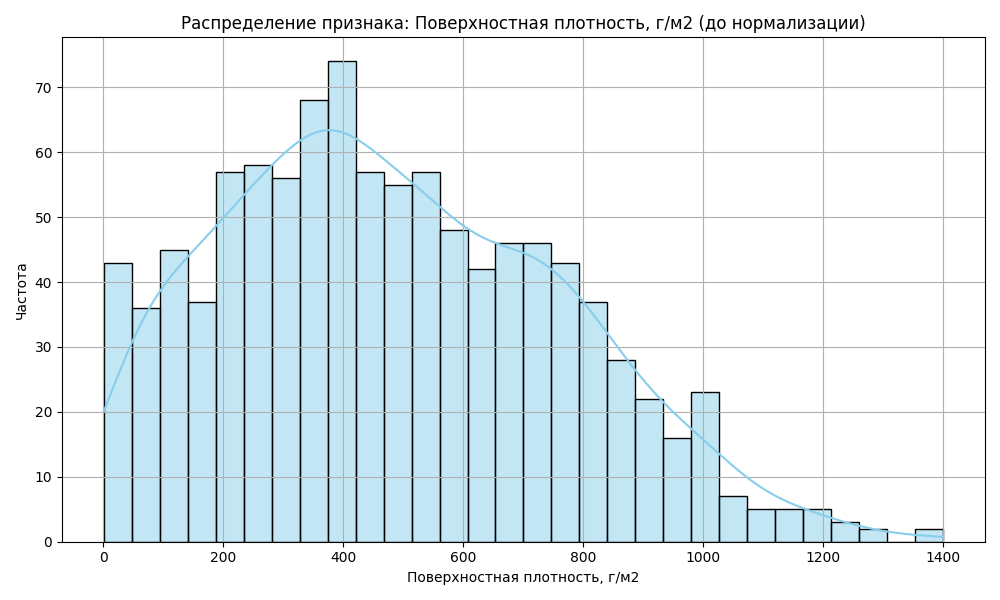


Рисунок 9 – Распределение признака до нормализации

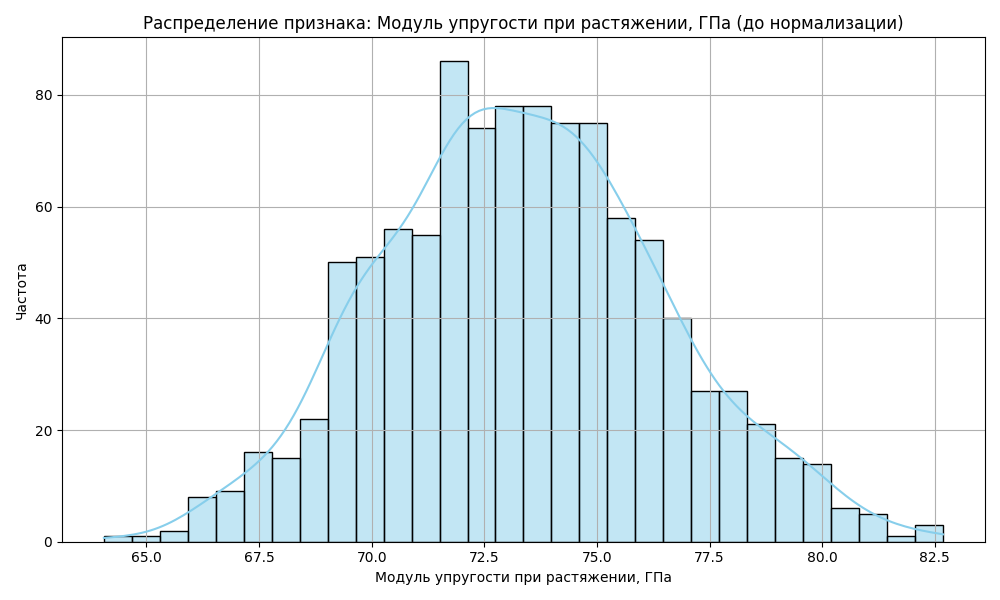


Рисунок 10 – Распределение признака до нормализации



Рисунок 11 – Распределение признака до нормализации



Рисунок 12 – Распределение признака до нормализации

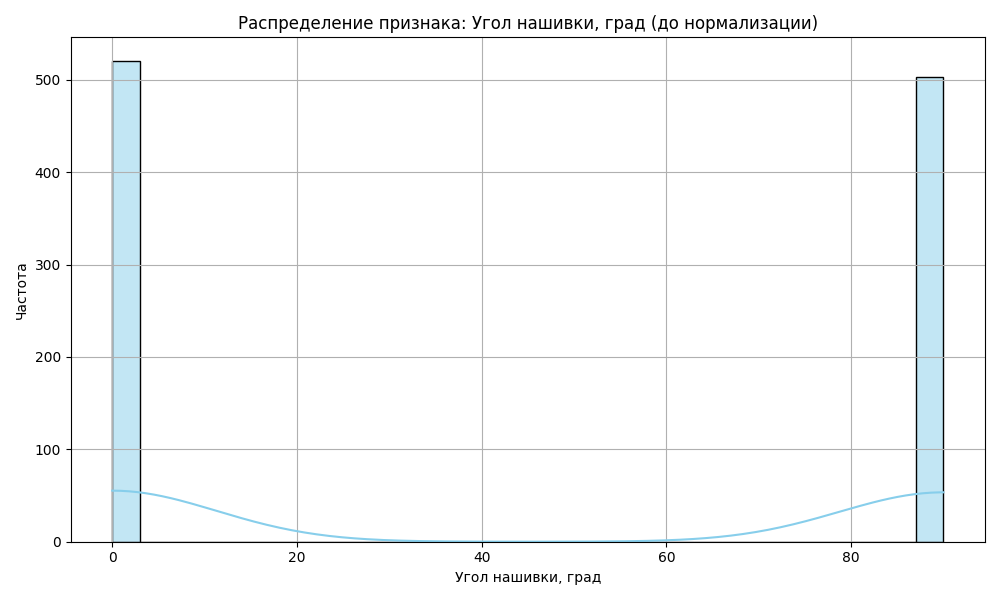


Рисунок 13 – Распределение признака до нормализации



Рисунок 14 – Распределение признака до нормализации



После применения StandardScaler (для нейронной сети использовался MinMaxScaler) нормализации были построены аналогичные гистограммы для оценки изменения формы распределения (результаты приведены для StandardScaler).

Рисунок 15 – Распределение признака после нормализации

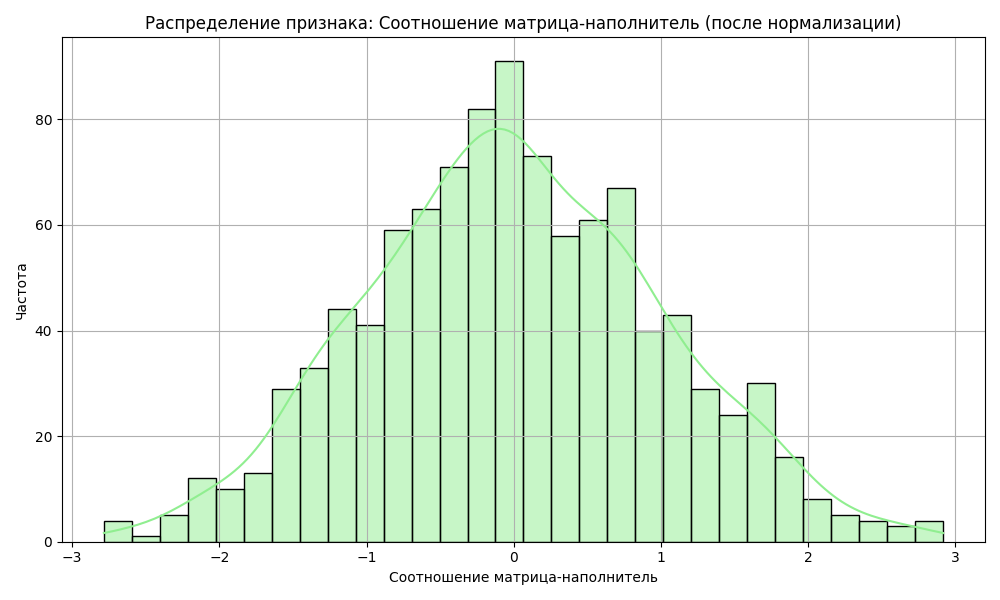


Рисунок 16 – Распределение признака после нормализации

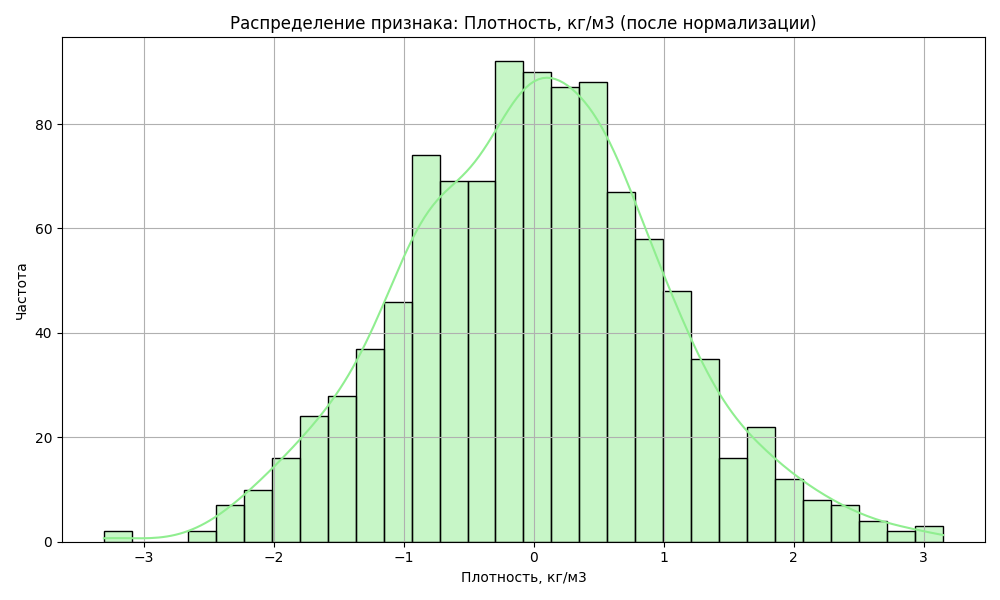


Рисунок 17 – Распределение признака после нормализации

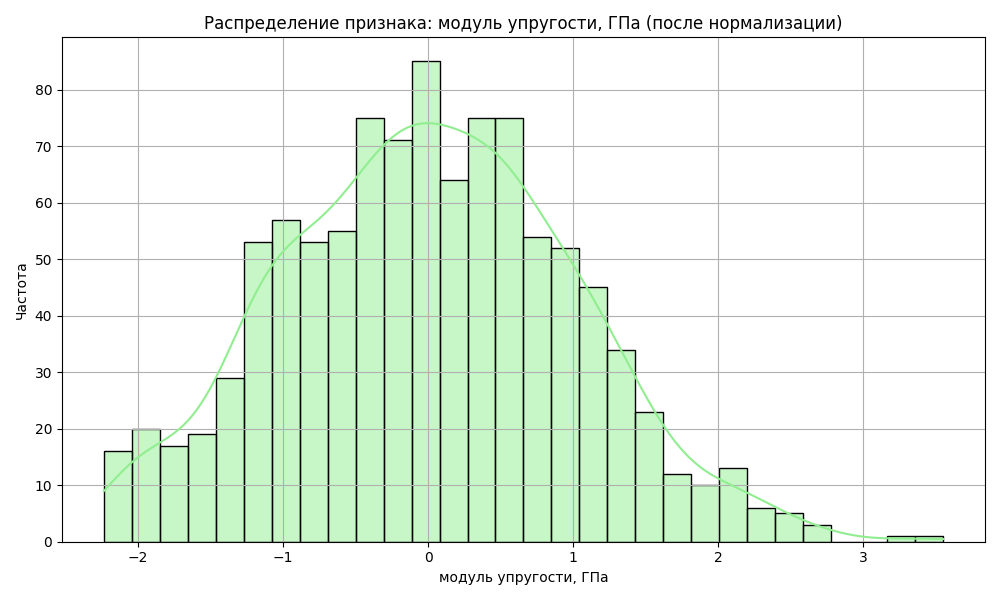


Рисунок 18 – Распределение признака после нормализации



Рисунок 19 – Распределение признака после нормализации



Рисунок 20 – Распределение признака после нормализации

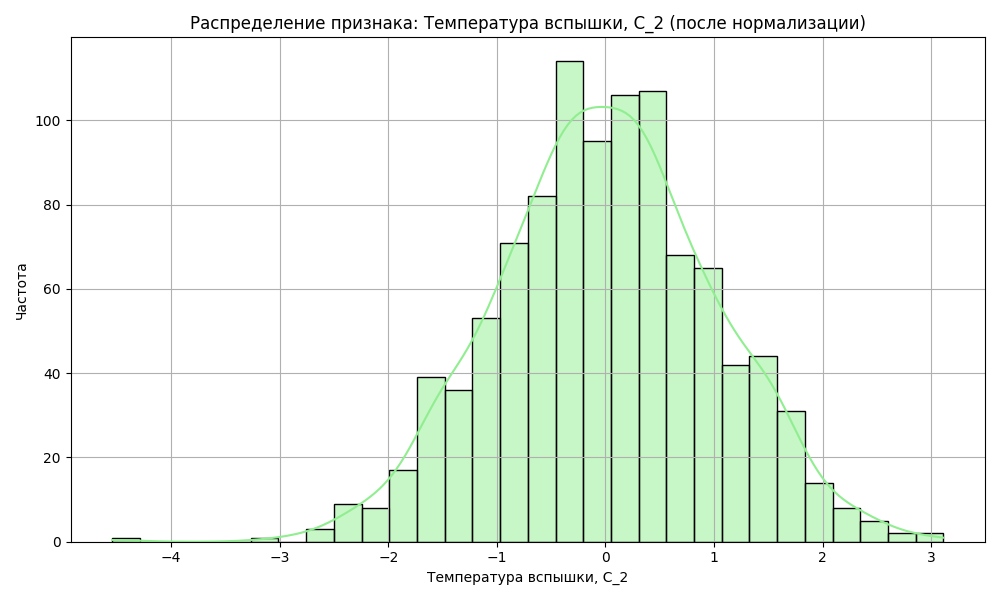


Рисунок 21 – Распределение признака после нормализации



Рисунок 22 – Распределение признака после нормализации

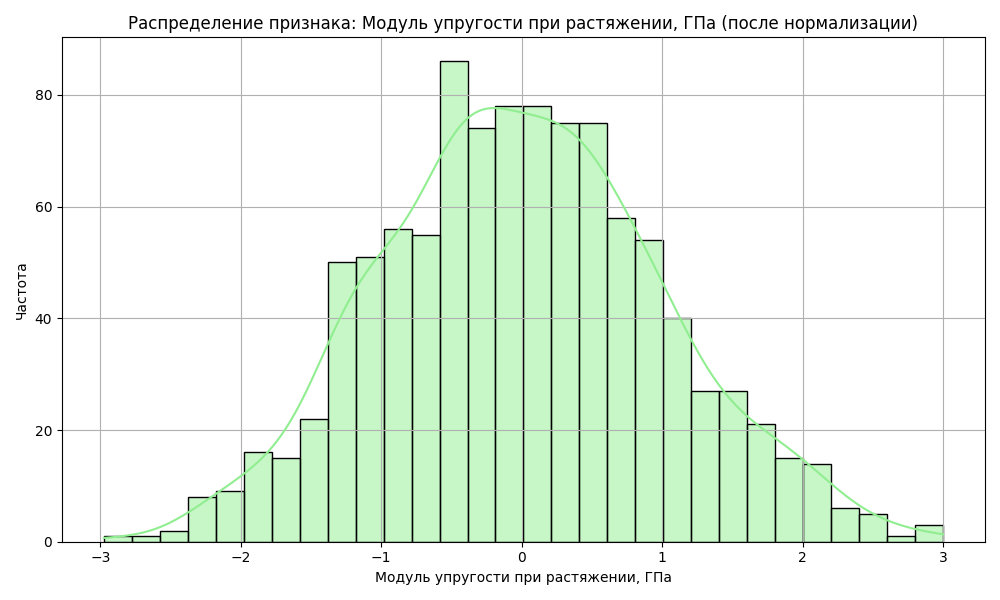


Рисунок 23 – Распределение признака после нормализации



Рисунок 24 – Распределение признака после нормализации



Рисунок 25 – Распределение признака после нормализации

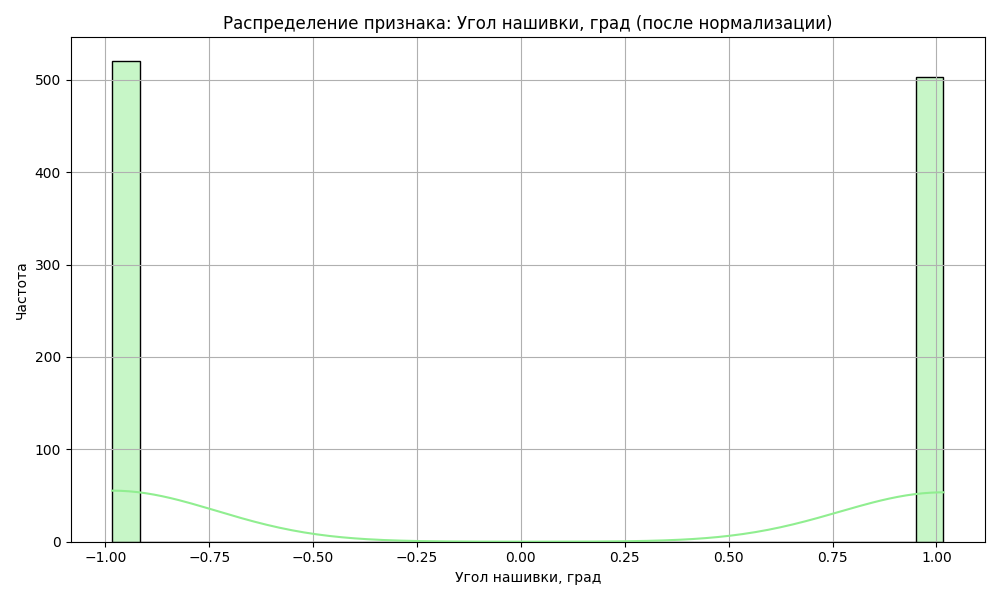


Рисунок 26 – Распределение признака после нормализации

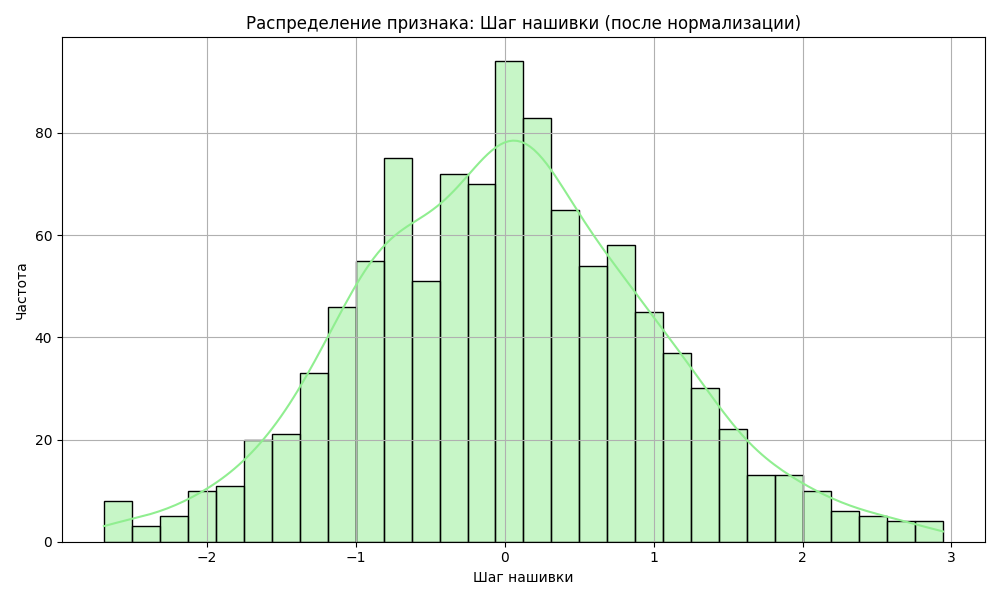


Рисунок 28 – Распределение признака после нормализации



Также была сформирована таблица, содержащая минимальные и максимальные значения признаков до и после нормализации. Эта таблица представлена ниже.

Таблица 2 – Сравнение распределения признаков до и после нормализации

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Минимум (до) | Максимум (до) | Минимум (после) | Максимум (после) |
| Соотношение матрица-наполнитель | 0.389403 | 5.591742 | -2.783776 | 2.915695 |
| Плотность, кг/м3 | 1731.764635 | 2207.773481 | -3.310622 | 3.148712 |
| модуль упругости, ГПа | 2.436909 | 1911.536477 | -2.234332 | 3.549589 |
| Количество отвердителя, м.% | 17.740275 | 198.953207 | -3.282308 | 3.125033 |
| Содержание эпоксидных групп,%\_2 | 14.254985 | 33.000000 | -3.321825 | 4.471955 |
| Температура вспышки, С\_2 | 100.000000 | 413.273418 | -4.542215 | 3.112932 |
| Поверхностная плотность, г/м2 | 0.603740 | 1399.542362 | -1.714677 | 3.260615 |
| Модуль упругости при растяжении, ГПа | 64.054061 | 82.682051 | -2.975023 | 3.000355 |
| Прочность при растяжении, МПа | 1036.856605 | 3848.436732 | -2.946217 | 2.846190 |
| Потребление смолы, г/м2 | 33.803026 | 414.590628 | -3.092116 | 3.285517 |
| Угол нашивки, град | 0.000000 | 90.000000 | -0.983518 | 1.016758 |
| Шаг нашивки | 0.000000 | 14.440522 | -2.692680 | 2.943275 |
| Плотность нашивки | 0.000000 | 103.988901 | -4.629749 | 3.793863 |

Дополнительно, на этапе предобработки были удалены выбросы с использованием метода межквартильного размаха (IQR). Это позволило исключить аномальные значения, которые могли бы исказить обучение моделей.

Таким образом, предобработка обеспечила корректное масштабирование признаков и очистку данных от выбросов, что является важным этапом перед построением моделей машинного обучения.

## **2.2. Разработка и обучение модели.**

Для решения задачи прогнозирования модуля упругости при растяжении и прочности при растяжении были выбраны и обучены несколько моделей регрессии. Каждая из них была протестирована как отдельно, так и в составе ансамбля. В качестве целевых переменных выступали два параметра:

1. Модуль упругости при растяжении, ГПа
2. Прочность при растяжении, МПа

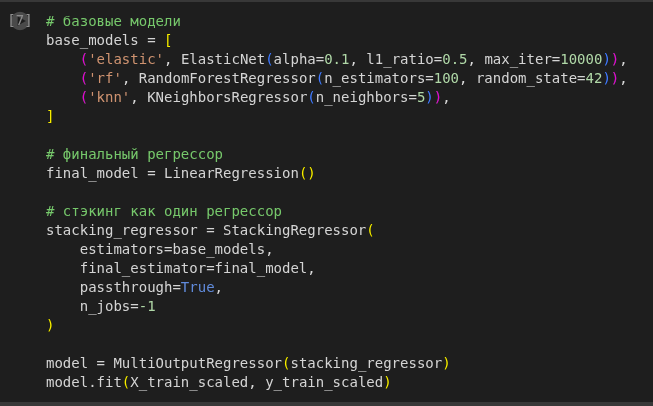
В ходе работы были протестированы следующие модели:

1. ElasticNet — линейная модель с L1 и L2-регуляризацией.
2. Random Forest Regressor — ансамблевый метод на основе решающих деревьев.
3. K-Nearest Neighbors Regressor (KNN) — метод на основе ближайших соседей.
4. Linear Regression — базовая линейная модель, использовавшаяся как мета-регрессор в стекинге.
5. Stacking Regressor — ансамблевая модель, объединяющая предсказания нескольких регрессоров с последующим обучением финального регрессора.

Подробное сравнение качества работы всех моделей приведено в [таблице сравнения регрессионных методов](file:///snap/onlyoffice-desktopeditors/477/opt/onlyoffice/desktopeditors/editors/web-apps/apps/documenteditor/main/index.html?_dc=0&lang=ru-RU&type=desktop&frameEditorId=placeholder&isForm=false&parentOrigin=file://&fileType=docx#%D1%82%D0%B0%D0%B1%D0%BB%D0%B8%D1%86%D0%B0-%D1%80%D0%B5%D0%B3%D1%80%D0%B5%D1%81%D1%81%D0%B8%D0%BE%D0%BD%D0%BD%D1%8B%D1%85-%D0%BC%D0%BE%D0%B4%D0%B5%D0%BB%D0%B5%D0%B9) в разделе 1.2. По результатам анализа наилучшие показатели по метрике R² и RMSE были достигнуты с использованием ансамблевой стекинг-модели, в которую входили Random Forest, KNN и ElasticNet.

Все модели обучались на обучающей выборке (70% от общего объема данных), с последующим тестированием на отложенной выборке (30%). Использовалась кросс-валидация и масштабирование признаков с помощью StandardScaler.

Рисунок 29 – Фрагмент кода



## **2.3. Тестирование модели.**

Для оценки качества обученной модели был проведён анализ ошибок на тренировочной и тестовой выборке. Основными метриками, использованными для оценки точности прогноза, стали:

1. Mean Squared Error (MSE) — среднеквадратичная ошибка;
2. R² (коэффициент детерминации) — отражает долю объяснённой дисперсии;
3. MAE – средняя абсолютная ошибка;
4. MAPE — средняя абсолютная процентная ошибка (дополнительно для лучшей интерпретируемости).

Таблица 3 – Оценки точности прогнозирования показателей

|  |  |  |
| --- | --- | --- |
| Показатель | Модуль упругости | Прочность |
| MSE | 9.13 | 229322.27 |
| MAE | 2.44 | 371.10 |
| MAPE | 3.32% | 16.95% |
| R² | 0.061 | 0.045 |

Полученные значения R² показывают, что модель слабо объясняет дисперсию в целевых переменных. Это может быть связано с высокой сложностью данных, наличием скрытых факторов или ограниченным размером выборки. Тем не менее, лучшую производительность продемонстрировал ансамбль на базе стекинга, что подтверждает выбор данной модели в качестве финальной.

## **2.4. Написать нейронную сеть, которая будет рекомендовать соотношение матрицы.**

Для задачи прогнозирования соотношения матрица-наполнитель была разработана и обучена нейронная сеть. Цель — предоставить инструмент, способный рекомендовать оптимальное значение данного параметра на основе других технологических и физико-химических характеристик материала.

Выбор архитектуры:

Модель представляет собой полносвязную нейросеть (Sequential) с несколькими скрытыми слоями. В качестве функции активации для скрытых слоёв использовалась tanh, что хорошо подходит для нормализованных данных. Выходной слой — линейный (один нейрон), так как решается задача регрессии.

Структура сети:

1. Входной слой: количество признаков после отбора;
2. Скрытые слои: 128 → 64 → 32 нейрона с активацией tanh;
3. Выходной слой: 1 нейрон (предсказание значения соотношения);
4. Оптимизатор: RMSprop;
5. Функция потерь: среднеквадратичная ошибка (MSE);
6. Регуляризация переобучения: EarlyStopping по валидационным данным с параметром patience = 50.

Предобработка данных:

1. Удаление выбросов с использованием межквартильного размаха (IQR);
2. Масштабирование признаков с помощью MinMaxScaler (в интервал [0, 1]);
3. Исключение целевого признака и некоторых коррелированных признаков (Плотность, Плотность нашивки) из набора входных данных.

Таблица 4 – Оценки точности прогнозирования целевого значения

|  |  |
| --- | --- |
| Метрика | Значение |
| MSE | 0.887 |
| MAE | 0.779 |
| MAPE | 31.97% |
| R² | 0.009 |

R² оказался очень низким, что указывает на слабую способность модели объяснять вариацию целевого признака. Это может быть связано с высокой шумностью данных, наличием латентных переменных, влияющих на результат, или сложной зависимостью, не поддающейся аппроксимации текущей архитектурой.

Тем не менее, модель продемонстрировала разумную интерполяцию для новых данных.

Таблица 5 – Сравнения реальных и прогнозированных значений целевой переменной

|  |  |
| --- | --- |
| Истинное значение | Предсказание |
| 1.857 | 2.540 |
| 4.194 | 3.146 |
| 3.306 | 3.152 |
| 2.710 | 2.969 |

Несмотря на низкое значение R², нейросеть обучена и может использоваться в качестве вспомогательного инструмента для предварительной оценки или ранжирования возможных значений соотношения матрицы, особенно в случае отсутствия лабораторных данных. Однако для производственного применения требуется доработка модели, возможно с использованием дополнительных признаков или другой архитектуры.

## **2.5. Разработка приложения.**

Для практического использования обученных моделей было разработано простое веб-приложение на основе Flask — лёгкого фреймворка для Python. Приложение реализует два ключевых функционала:

Прогнозирование прочностных характеристик (модуль упругости и прочность при растяжении) на основе входных параметров с использованием модели стекинга (stacking\_pipeline.pkl);

Рекомендация соотношения матрица-наполнитель на основе физических и технологических параметров материала с помощью нейросети (neiro\_model.h5).

Запустить приложение с помощью Docker:

1. перейти в директорию с веб-приложением «cd flask»;
2. создать докер образ «docker build -t flask-ml-app .»;
3. запустить докер контейнер «docker run -p 5000:5000 flask-ml-app»;
4. перейти в браузере по адресу: [http://localhost:5000](http://localhost:5000/).
5. выбрать нужную форму: «Перейти к предсказанию с помощью машинного обучения» — ввод данных для получения модуля упругости и прочности и «Перейти к предсказанию с помощью нейросети» — ввод данных для рекомендации соотношения матрица-наполнитель;
6. ввести значения параметров;
7. нажать кнопку "Предсказать" и получить результат.

## **2.6. Создание удаленного репозитория и загрузка результатов работы на него.**

Для хранения исходного кода проекта, обученных моделей и документации был создан удалённый репозиторий на GitHub. Это позволяет обеспечить надёжное хранение, версионность и возможность совместной работы или демонстрации результатов.

Страница слушателя:

https://github.com/Sergeev-IT

Репозиторий проекта:

https://github.com/Sergeev-IT/diplom

# **Заключение**

В рамках выпускной квалификационной работы была разработана интеллектуальная система прогнозирования и рекомендаций, направленная на повышение точности оценки механических свойств композиционных материалов. Основной целью исследования стало построение моделей, способных предсказывать такие характеристики, как модуль упругости и прочность при растяжении, а также рекомендовать оптимальное соотношение матрица-наполнитель.

В процессе работы были решены следующие задачи:

1. Проведена предварительная обработка и анализ данных, включая удаление выбросов и отбор признаков;
2. Построена ансамблевая модель стекинга, включающая несколько регрессионных алгоритмов, для предсказания механических свойств;
3. Разработана нейронная сеть для индивидуального прогнозирования рекомендуемого соотношения компонентов;
4. Разработано веб-приложение на Flask, интегрирующее обе модели и предоставляющее простой интерфейс для пользователя;
5. Организовано развёртывание приложения с помощью Docker, что обеспечивает переносимость и удобство использования;
6. Создан удалённый репозиторий для хранения исходного кода, моделей и документации.

Несмотря на то, что достигнутый уровень точности (в частности, по метрике R²) остаётся ограниченным, предложенный подход демонстрирует перспективность применения методов машинного обучения в задачах материаловедения. Полученные результаты подтверждают возможность автоматизации аналитических задач в инженерной практике и служат отправной точкой для дальнейших улучшений, таких как расширение обучающей выборки, использование более сложных моделей или привлечение дополнительных признаков.

Таким образом, поставленные цели были достигнуты, а практическая реализация системы делает возможным её дальнейшее применение и развитие в рамках промышленных или научно-исследовательских задач.

# **Библиографический список**

1. Российская Федерация. Законы. Федеральный закон от 27 июля 2006 г. № 149-ФЗ «Об информации, информационных технологиях и о защите информации» // Собрание законодательства РФ. – 2006. – № 31 (ч. 1). – Ст. 3448.
2. Российская Федерация. Указы Президента. Указ Президента РФ от 10 октября 2019 г. № 490 «О развитии искусственного интеллекта в Российской Федерации» // Собрание законодательства РФ. – 2019. – № 42. – Ст. 5997.
3. Российская Федерация. Постановления Правительства. Постановление Правительства РФ от 3 марта 2017 г. № 223 «О государственной программе Российской Федерации "Цифровая экономика Российской Федерации"» // Собрание законодательства РФ. – 2017. – № 10. – Ст. 1496.
4. Герон, А. Практическое машинное обучение с помощью Scikit-Learn, Keras и TensorFlow / А. Герон. – М. : ДМК Пресс, 2020. – 800 c.
5. Гудфеллоу, И., Бенджио, Й., Курвилль, А. Глубокое обучение / И. Гудфеллоу, Й. Бенджио, А. Курвилль. – М. : ДМК Пресс, 2018. – 800 с.
6. Маккинни, У. Python для анализа данных / У. Маккинни. – М. : ДМК Пресс, 2018. – 544 с.
7. ВандерПлас, Дж. Python Data Science Handbook / Дж. ВандерПлас. – М. : ДМК Пресс, 2017. – 548 с.
8. Брюс, П., Брюс, А. Практическая статистика для специалистов по данным / П. Брюс, А. Брюс. – М. : ДМК Пресс, 2018. – 320 с.
9. Грус, Дж. Data Science с нуля / Дж. Грус. – М. : ДМК Пресс, 2017. – 330 с.
10. Шалев-Шварц, Ш., Бен-Давид, Ш. Понимание машинного обучения / Ш. Шалев-Шварц, Ш. Бен-Давид. – М. : ДМК Пресс, 2016. – 410 с.
11. Спигельхалтер, Д. Искусство статистики: Как извлекать смысл из данных / Д. Спигельхалтер. – М. : Альпина Паблишер, 2020. – 352 с.
12. Провост, Ф., Фосетт, Т. Data Science для бизнеса / Ф. Провост, Т. Фосетт. – М. : ДМК Пресс, 2016. – 432 с.
13. Чолле, Ф. Глубокое обучение с использованием Python / Ф. Чолле. – М. : ДМК Пресс, 2018. – 384 с.
14. Герон, А. Машинное обучение на практике / А. Герон. – М. : ДМК Пресс, 2019. – 800 с.
15. Вилке, К. Основы визуализации данных / К. Вилке. – М. : ДМК Пресс, 2020. – 320 с.
16. Бурков, А. Машинное обучение за 100 страниц / А. Бурков. – М. : ДМК Пресс, 2019. – 160 с.
17. Робинсон, Э., Нолис, Ж. Построение карьеры в Data Science / Э. Робинсон, Ж. Нолис. – М. : ДМК Пресс, 2020. – 350 с.
18. Пэнг, Р., Мацуи, Э. Искусство Data Science / Р. Пэнг, Э. Мацуи. – М. : ДМК Пресс, 2017. – 150 с.
19. Хернан, М. А., Хсу, Дж., Хили, Б. Data Science как вторая попытка науки в причинном выводе / М. А. Хернан, Дж. Хсу, Б. Хили // arXiv preprint arXiv:1804.10846. – 2018. – 20 с.
20. Карпатне, А. и др. Теория-ориентированная Data Science: новый подход к научным открытиям из данных / А. Карпатне и др. // arXiv preprint arXiv:1612.08544. – 2016. – 15 с.
21. Цао, Л. Data Science: всесторонний обзор / Л. Цао // arXiv preprint arXiv:2007.03606. – 2020. – 30 с.
22. Цао, Л. Data Science: вызовы и направления / Л. Цао // arXiv preprint arXiv:2006.16966. – 2020. – 25 с.
23. Википедия. Data Science [Электронный ресурс]. – Режим доступа: <https://ru.wikipedia.org/wiki/Data_science> (дата обращения: 01.05.2025).
24. Scikit-learn: машинное обучение в Python [Электронный ресурс]. – Режим доступа: <https://scikit-learn.org/stable/> (дата обращения: 01.05.2025).
25. Kaggle. Платформа для соревнований и обучения по Data Science [Электронный ресурс]. – Режим доступа: [https://www.kaggle.com](https://www.kaggle.com/) (дата обращения: 02.05.2025).
26. TensorFlow. Библиотека машинного обучения с открытым исходным кодом [Электронный ресурс]. – Режим доступа: [https://www.tensorflow.org](https://www.tensorflow.org/) (дата обращения: 02.05.2025).
27. PyTorch. Глубокое обучение для всех [Электронный ресурс]. – Режим доступа: [https://pytorch.org](https://pytorch.org/) (дата обращения: 02.05.2025).
28. Towards Data Science – Публикации по Data Science, ML и AI [Электронный ресурс]. – Режим доступа: [https://towardsdatascience.com](https://towardsdatascience.com/) (дата обращения: 02.05.2025).
29. GitHub – Репозитории и проекты по Data Science [Электронный ресурс]. – Режим доступа: <https://github.com/topics/data-science> (дата обращения: 03.05.2025).
30. Google AI Blog – Исследования в области искусственного интеллекта [Электронный ресурс]. – Режим доступа: https://ai.googleblog.com (дата обращения: 04.05.2025).
31. Medium – Раздел Data Science [Электронный ресурс]. – Режим доступа: https://medium.com/topic/data-science (дата обращения: 04.05.2025).
32. Data Science Central – Сообщество и статьи по аналитике и данным [Электронный ресурс]. – Режим доступа: [https://www.datasciencecentral.com](https://www.datasciencecentral.com/) (дата обращения: 05.05.2025).
33. Analytics Vidhya – Платформа по обучению и новостям Data Science [Электронный ресурс]. – Режим доступа: [https://www.analyticsvidhya.com](https://www.analyticsvidhya.com/) (дата обращения: 05.05.2025).
34. Machine Learning Mastery – Практические руководства по ML [Электронный ресурс] / Дж. Браунли. – Режим доступа: [https://machinelearningmastery.com](https://machinelearningmastery.com/) (дата обращения: 06.05.2025).
35. OpenAI. Исследования и инструменты в области ИИ [Электронный ресурс]. – Режим доступа: <https://openai.com/research> (дата обращения: 07.05.2025).