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 Скорых Галина Александровна

Москва, 2023

# Содержание

|  |  |
| --- | --- |
|  |  |
| Введение | 3 |
| 1 Аналитическая часть | 5 |
| 1.1 Постановка задачи | 5 |
| 1.2 Описание используемых методов | 6 |
| 1.2.1Линейная регрессия | 6 |
| 1.2.2 Метод опорных векторов | 7 |
| 1.2.3 Метод К-ближайших соседей. | 9 |
| 1.2.4 Дерево принятия решений | 9 |
| 1.2.5 Случайный лес | 11 |
| 1.2.6 Градиентный бустинг | 12 |
| 1.2.7 Стохастический градиентный спуск | 13 |
| 1.3 Разведочный анализ данных | 14 |
| 2 Практическая часть | 19 |
| 2.1 Предобработка данных | 19 |
| 2.2 Разработка и обучение модели | 21 |
| 2.3 Создание нейронной сети. | 23 |
| 2.4 Экспорт модели для создания приложения | 25 |
| 2.5 Создание удаленного репозитория | 25 |
| Заключение | 26 |
| Библиографический список | 27 |
|  |  |

**Введение**

Тема данной работы - прогнозирование конечных свойств новых материалов (композиционных материалов).

Композиционными называются материалы, в которых имеет место сочетание двух (или более) химически разнородных компонентов (фаз) с четкой границей раздела между ними. Это неоднородные по химическому составу и структуре материалы. Структура композиционных материалов представляет собой матрицу (основной компонент), содержащую в своем объеме или армирующие элементы, часто называемые наполнителем. Матрица и наполнитель разделены границей (поверхностью) раздела. Наполнитель равномерно распределен в матрице и имеет заданную пространственную ориентацию.

Внедрение композиционных материалов обусловлено стремлением использовать их преимущества по сравнению с традиционно используемыми металлами и сплавами. Примеры композита – железобетон (сочетание стали арматуры и камня бетона), древесноволокнистая плита ДВП (сочетание древесной основы – щепы и полимерного связующего).

Базальт - магматическая вулканическая порода. Это самая распространённая порода на поверхности Земли и на других планетах Солнечной системы. Базальты образуются путём затвердевания силикатного магматического расплава. Большая часть базальтов образуется на срединно-океанических хребтах и образует океаническую кору. Активно развивается использование композитных материалов на основе базальта.

Базальтопластик - современный композитный материал на основе базальтовых волокон и органического связующего вещества. В настоящее время базальтопластик успешно конкурирует с металлическими изделиями, превосходя их по коррозионной, щелочной, кислотоустойчивости и некоторым другим свойствам. **Целью** данной работы является прогнозирование конечных свойств новых материалов на основе базальтопластика (композиционных материалов).

**Актуальность:** Созданные прогнозные модели помогут сократить количество проводимых испытаний, а также пополнить базу данных материалов возможными новыми характеристиками материалов, и цифровыми двойниками новых композитов.

Расширение разнообразия материалов, используемых при проектировании нового композиционного материала, увеличивает необходимость определения свойств нового композита при минимальных финансовых затратах. Обычно используются два способа: физические тесты образцов материалов или оценка свойств, в том числе на основе физико-математических моделей. Традиционно разработка композитных материалов является долгосрочным процессом, так как из свойств отдельных компонентов невозможно рассчитать конечные свойства композита. Для достижения определенных характеристик требуется большое количество различных комбинированных тестов, что делает насущной задачу прогнозирования успешного решения, снижающего затраты на разработку новых материалов и затрат на рабочую силу.

Суть прогнозирования заключается в моделировании репрезентативного элемента композитного объёма на основе данных о свойствах входящих компонентов (связующего и армирующего компонента). В процессе исследовательской работы были разработаны несколько моделей, способные с высокой вероятностью прогнозировать модули упругости при растяжении и прочности при растяжении, а также были созданы нейронных сети, которые предлагают соотношение «матрицы - наполнитель».

**1 Аналитическая часть**

**1.1 Постановка задачи**

В данной работе исследуется композит с матрицей из базальтопластика и нашивками из углепластика. От специалистов в предметной области был получен датасет, содержащий данные о свойствах матрицы и наполнителя, производственных параметрах и свойствах готового композита. От нас, как специалистов в машинном обучении, требуется разработать модели, прогнозирующие значения некоторых свойств в зависимости от остальных. Так же требуется разработать приложение, делающее удобным использование данных моделей специалистом предметной области. Для исследовательской работы были даны 2 файла: X\_bp (составляющая из базальтопластика) и Х\_nup (составляющая из углепластика).

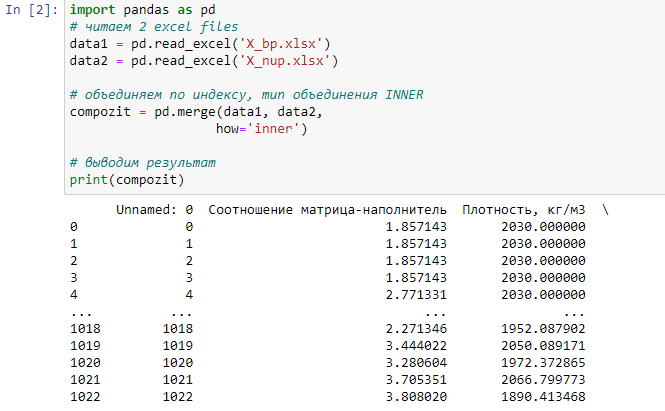
Известно, что файлы требуют объединения с типом INNER по индексу. После объединения часть строк из файла X\_nup была отброшена. И дальнейшие исследования проводим с объединенным датасетом, содержащим 13 признаков и 1023 строк или объектов.

Рисунок 1 - пример начала работы с файлами

**1.2 Описание используемых методов**

Предсказание значений вещественной, непрерывной переменной — это задача регрессии. Данная задача в рамках классификации категорий машинного обучения относится к машинному обучению с учителем. Цель любого алгоритма обучения с учителем — определить функцию потерь и минимизировать её, поэтому для наилучшего решения в процессе исследования были применены следующие методы:

линейная регрессия;

метод опорных векторов;

случайный лес;

градиентный бустинг;

К-ближайших соседей;

дерево решений;

стохастический градиентный спуск.

**1.2.1Линейная регрессия** — Используемая в статистике регрессионная модель зависимости одной переменной y от другой или нескольких других переменных x с линейной функцией зависимости. Это алгоритм машинного обучения, основанный на контролируемом обучении, рассматривающий зависимость между одной входной и выходными переменными. Это один из самых простых и эффективных инструментов статистического моделирования. Она определяет зависимость переменных с помощью линии наилучшего соответствия. Модель регрессии создаёт несколько метрик. R2 , или коэффициент детерминации, позволяет измерить, насколько модель может объяснить дисперсию данных. Если R-квадрат равен 1, это значит, что модель описывает все данные. Если же R-квадрат равен 0,5, модель объясняет лишь 50 процентов дисперсии данных. Оставшиеся отклонения не имеют объяснения. Чем ближе R2 к единице, тем лучше. Коэффициент детерминации не может быть отрицательным, данный вывод исходит из свойств коэффициента детерминации. Однако скорректированный коэффициент детерминации вполне может принимать отрицательные значения.

Достоинства метода: быстр и прост в реализации; легко интерпретируем; имеет меньшую сложность по сравнению с другими алгоритмами;

Недостатки метода: моделирует только прямые линейные зависимости; требует прямую связь между зависимыми и независимыми переменными; выбросы оказывают огромное влияние, а границы линейны.

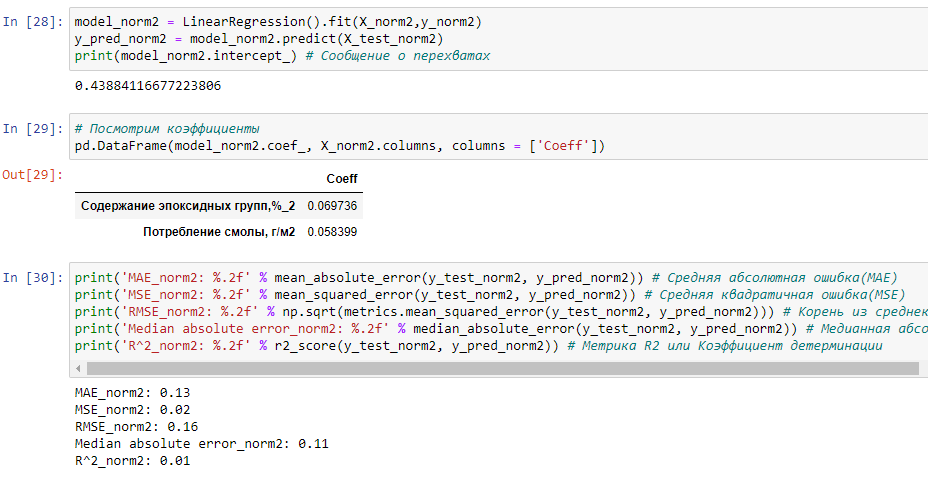


Рисунок 2 – создание модели линейной регрессии

**1.2.2 Метод опорных векторов** – один из наиболее популярных методов машинного обучения. Он создает гиперплоскость или набор гиперплоскостей в многомерном пространстве, которые могут быть использованы для решения задач классификации и регрессии.Чаще всего он применяется в постановке бинарной классификации. Данный алгоритм – это алгоритм обучения с учителем, использующихся для задач классификации и регрессионного анализа, это контролируемое обучение моделей с использование схожих алгоритмов для анализа данных и распознавания шаблонов. Учитывая обучающую выборку, где алгоритм помечает каждый объект, как принадлежащий к одной из двух категорий, строит модель, которая определяет новые наблюдения в одну из категорий. Модель метода опорных векторов – отображение данных точками в пространстве, так что между наблюдениями отдельных категорий имеется разрыв, и он максимален. Каждый объект данных представляется как вектор (точка) в p-мерном пространстве. Он создаёт линию или гиперплоскость, которая разделяет данные на классы.

Достоинства метода: для классификации достаточно небольшого набора данных. При правильной работе модели, построенной на тестовом множестве, вполне возможно применение данного метода на реальных данных. Эффективен при большом количестве гиперпараметров. Способен обрабатывать случаи, когда гиперпараметров больше, чем количество наблюдений. Существует возможность гибко настраивать разделяющую функцию. Алгоритм максимизирует разделяющую полосу, которая, как подушка безопасности, позволяет уменьшить количество ошибок классификации.

Недостатки метода: чувствительность к выбросам, поэтому в работе была проведена тщательнейшая работа с выбросами. Для больших наборов данных требуется долгое время обучения; достаточно сложно подбирать полезные преобразования данных; параметры модели сложно интерпретировать, поэтому были рассмотрены и другие методы.

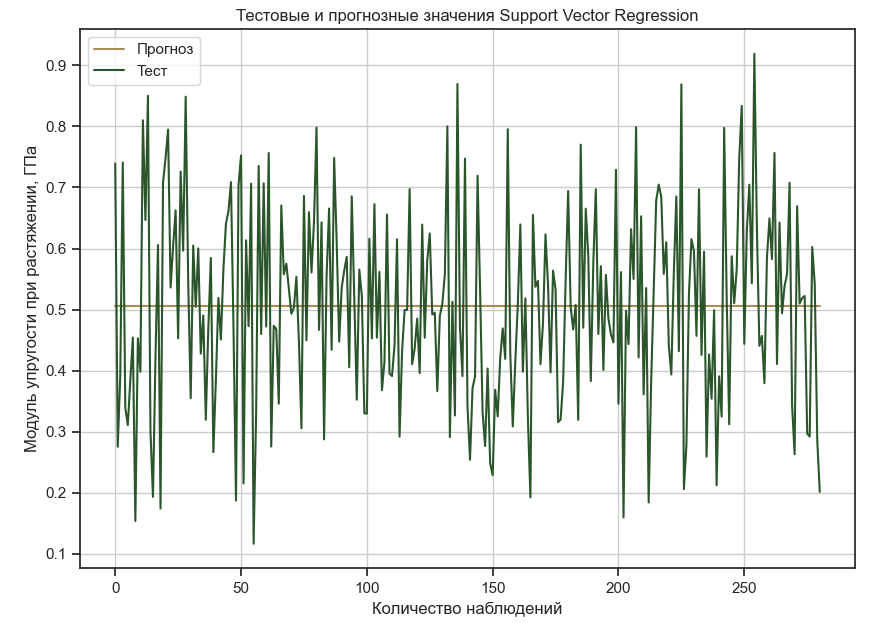


Рисунок 3 – тестовые и прогнозные значения модели Методом опорных векторов для 'Модуль упругости при растяжении, ГПа'

**1.2.3 Метод К-ближайших соседей.**

Еще один метод классификации, который адаптирован для регрессии - метод k-ближайших соседей (k Nearest Neighbors). На интуитивном уровне суть метода проста: посмотри на соседей вокруг, какие из них преобладают, таковым ты и являешься. В случае использования метода для регрессии, объекту присваивается среднее значение по k ближайшим к нему объектам, значения которых уже известны.

Достоинства метода: прост в реализации и понимании полученных результатов; имеет низкую чувствительность к выбросам; не требует построения модели; допускает настройку нескольких параметров; позволяет делать дополнительные допущения; универсален; находит лучшее решение из возможных; решает задачи небольшой размерности.

Недостатки метода: замедляется с ростом объёма данных; не создаёт правил; не обобщает предыдущий опыт; основывается на всем массиве доступных исторических данных; невозможно сказать, на каком основании строятся ответы; сложно выбрать близость метрики; имеет высокую зависимость результатов классификации от выбранной метрики; полностью перебирает всю обучающую выборку при распознавании; имеет вычислительную трудоёмкость.

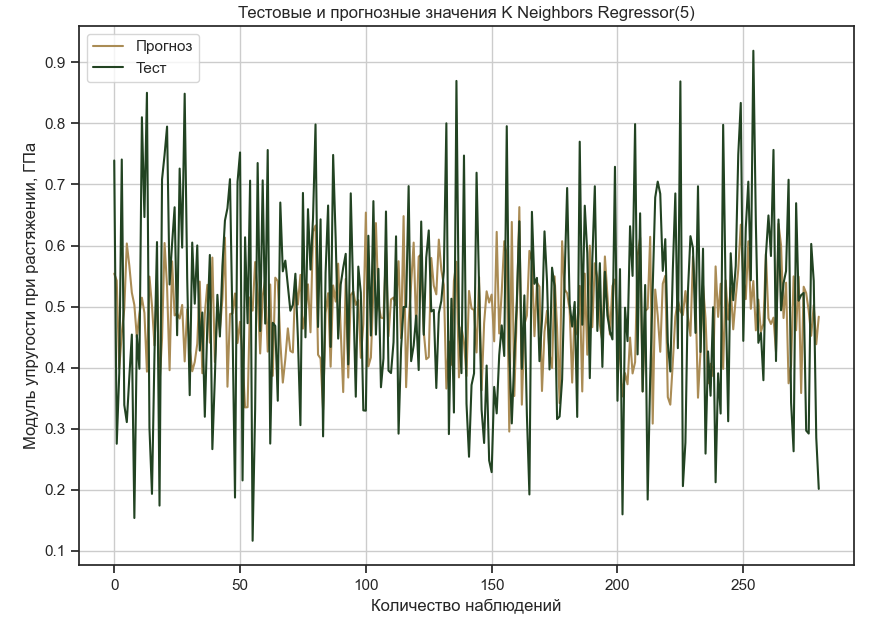


Рисунок 4 – тестовые и прогнозные значения модели Методом К-ближайших соседей для ‘Модуль упругости при растяжении, ГПа'

**1.2.4 Дерево принятия решений** – метод автоматического анализа больших массивов данных. Это инструмент принятия решений, в котором используется древовидная структура, подобная блок-схеме, или модель решений и всех их возможных результатов, включая результаты, затраты и полезность. Дерево принятия решений - эффективный инструмент интеллектуального анализа данных и предсказательной аналитики. Алгоритм дерева решений подпадает под категорию контролируемых алгоритмов обучения. Он работает как для непрерывных, так и для категориальных выходных переменных. Правила генерируются за счёт обобщения множества отдельных наблюдений (обучающих примеров), описывающих предметную область. Регрессия дерева решений отслеживает особенности объекта и обучает модель в структуре дерева прогнозированию данных в будущем для получения значимого непрерывного вывода. Дерево решений один из вариантов решения регрессионной задачи, в случае если зависимость в данных не имеет очевидной корреляции.

Достоинства метода: помогают визуализировать процесс принятия решения и сделать правильный выбор в ситуациях, когда результаты одного решения влияют на результаты следующих решений; создаются по понятным правилам; просты в применении и интерпретации; заполняют пропуски в данных наиболее вероятным решением; работают с разными переменными; выделяют наиболее важные поля для прогнозирования;

Недостатки метода: ошибается при классификации с большим количеством классов и небольшой обучающей выборкой; имеют нестабильный процесс (изменение в одном узле может привести к построению совсем другого дерева); имеет затратные вычисления; необходимо обращать внимание на размер; ограниченное число вариантов решения проблемы.

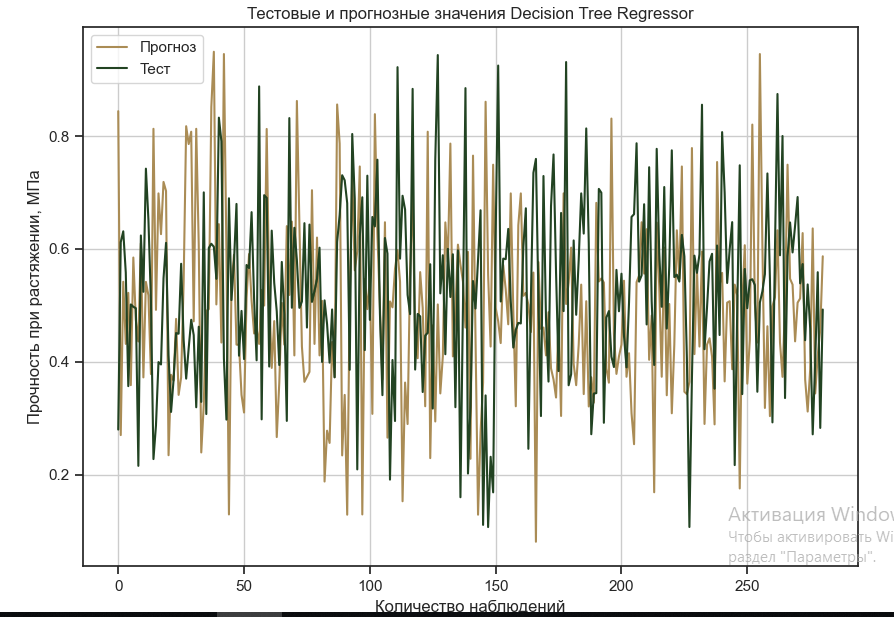


Рисунок 5 - график дерева принятия решений для Прочность при растяжении

**1.2.5 Случайный лес** — это множество решающих деревьев. Универсальный алгоритм машинного обучения с учителем, представитель ансамблевых методов. Если точность дерева решений оказалось недостаточной, мы можем множество моделей собрать в коллектив.

Достоинства метода: не переобучается; не требует предобработки входных данных; эффективно обрабатывает пропущенные данные, данные с большим числом классов и признаков; имеет высокую точность предсказания и внутреннюю оценку обобщающей способности модели, а также высокую параллелизуемость и масштабируемость.

Недостатки метода: построение занимает много времени; сложно интерпретируемый; не обладает возможностью экстраполяции; может недо обучаться; трудоёмко прогнозируемый; иногда работает хуже, чем линейные методы.

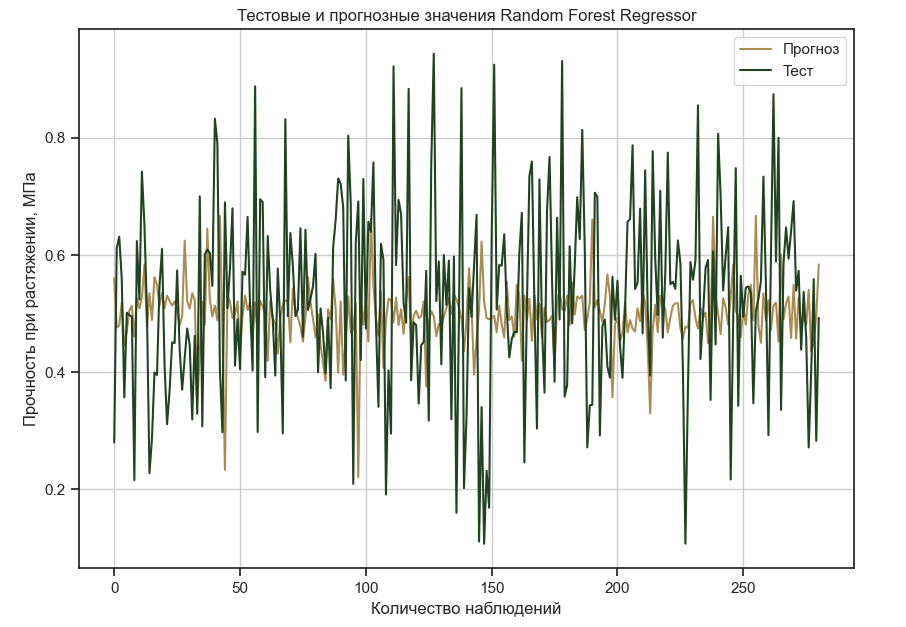


Рисунок 6 – тестовые и прогнозные значения модели Методом случайный лес для 'Прочность при растяжении, МПа '

**1.2.6 Градиентный бустинг**  — это ансамбль деревьев решений, обученный с использованием градиентного бустинга. В основе данного алгоритма лежит итеративное обучение деревьев решений с целью минимизировать функцию потерь. Основная идея градиентного бустинга: строятся последовательно несколько базовых классификаторов, каждый из которых как можно лучше компенсирует недостатки предыдущих. Финальный классификатор является линейной композицией этих базовых классификаторов.

Достоинства метода: новые алгоритмы учатся на ошибках предыдущих; требуется меньше итераций, чтобы приблизиться к фактическим прогнозам; наблюдения выбираются на основе ошибки; прост в настройке темпа обучения и применения; легко интерпретируем.

Недостатки метода: необходимо тщательно выбирать критерии остановки, иначе это может привести к переобучению; наблюдения с наибольшей ошибкой появляются чаще; слабее и менее гибок чем нейронные сети.

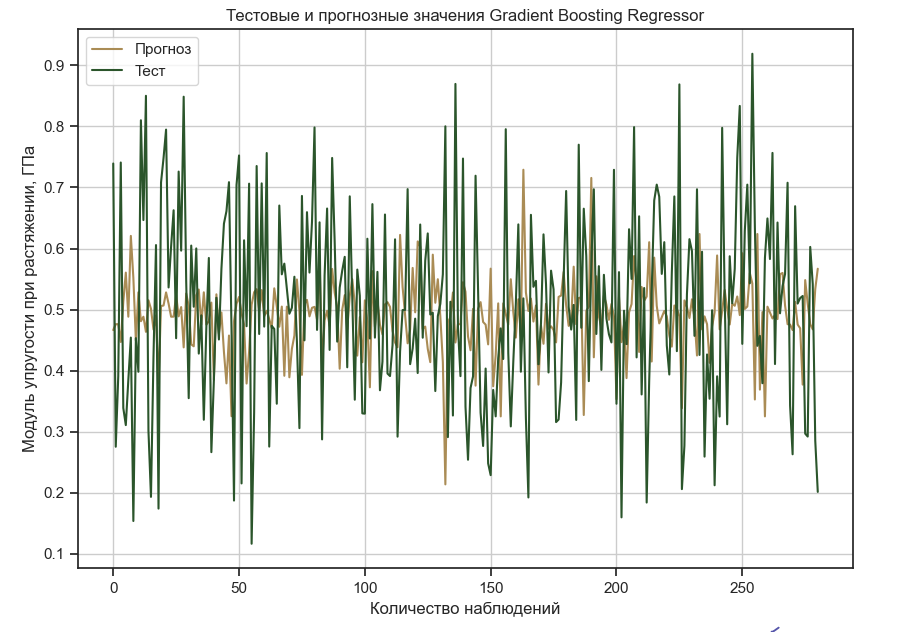


Рисунок 5 – тестовые и прогнозные значения модели Методом градиентного бустинга для 'Модуль упругости при растяжении, ГПа'

**1.2.7 Стохастический градиентный спуск** — это простой, но очень эффективный подход к подгонке линейных классификаторов и регрессоров под выпуклые функции потерь. Этот подход подразумевает корректировку весов нейронной сети, используя аппроксимацию градиента функционала, вычисленную только на одном случайном обучающем примере из выборки.

Достоинства метода: эффективен; прост в реализации; имеет множество возможностей для настройки кода; способен обучаться на избыточно больших выборках.

Недостатки метода: требует ряд гиперпараметров; чувствителен к масштабированию функций; может не сходиться или сходиться слишком медленно; функционал многоэкстремален; процесс может "застрять" в одном из локальных минимумов; возможно переобучение.

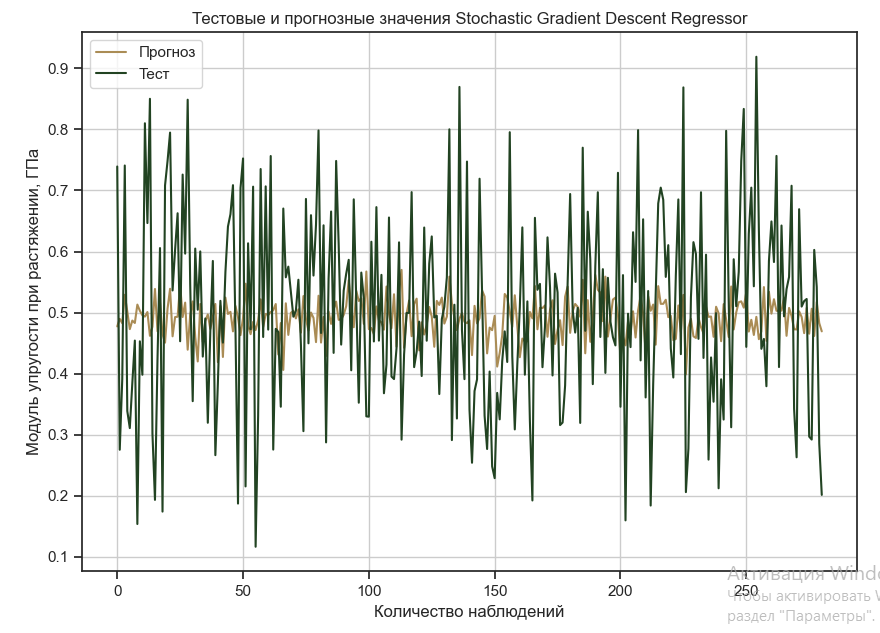


Рисунок 6 – тестовые и прогнозные значения модели Методом Стохастический градиентный спуск для 'Модуль упругости при растяжении'

Немного расскажем об используемых метриках качества моделей: **R2 или коэффициент детерминации** измеряет долю дисперсии, объяснённую моделью, в общей дисперсии целевой переменной. 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.

**1.3 Разведочный анализ данных**

Прежде чем передать данные в работу моделей машинного обучения, необходимо обработать и очистить их. Очевидно, что «грязные» и необработанные данные могут содержать искажения и пропущенные значения – это ненадёжно, поскольку способно привести к крайне неверным результатам по итогам моделирования. Но безосновательно удалять что-либо тоже неправильно.

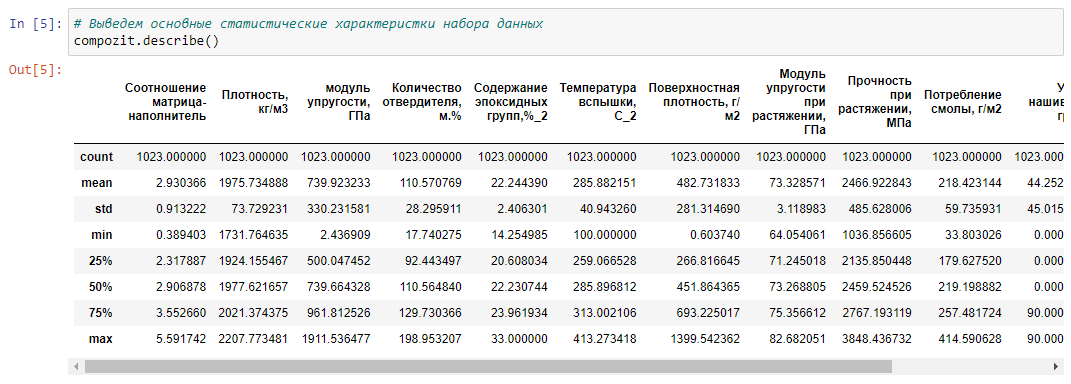


Рисунок 7 – основные статистические характеристики набора данных

Цель разведочного анализа данных — выявить закономерности в данных. Для корректной работы большинства моделей желательна сильная зависимость выходных переменных от входных и отсутствие зависимости между входными переменными.

В качестве инструментов разведочного анализа используется: 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 (несколько вариантов); описательная статистика для каждой переменной; анализ и полное исключение выбросов; проверка наличия пропусков и дубликатов; ранговая корреляция Кендалла и Пирсона.

Была проведена визуализация сырых данных и их анализ. Построены гистограммы распределения каждой из переменных и боксплоты, диаграммы "ящиков с усами", попарные графики рассеяния точек, графики квантиль-квантиль, тепловая карта без нормализации и исключения шумов.



Рисунок 8 – гистограммы распределения каждой из переменных без нормализации и исключения шумов



Рисунок 09 – график "ящиков с усами" для всех переменных

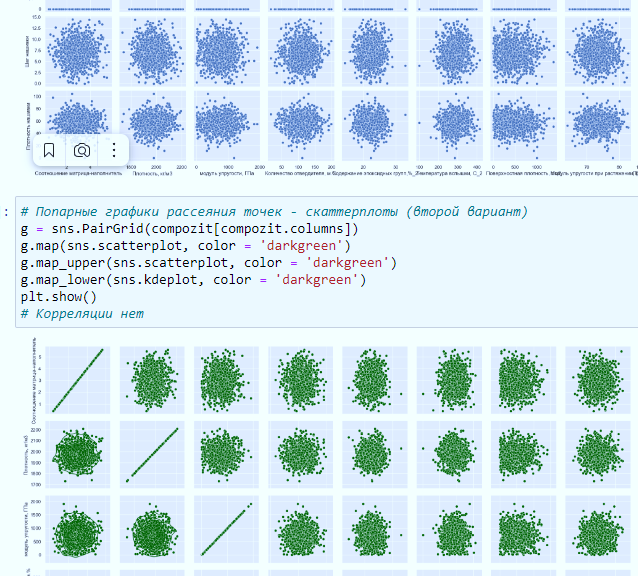


Рисунок 10 – попарные графики рассеяния точек (два разных варианта)

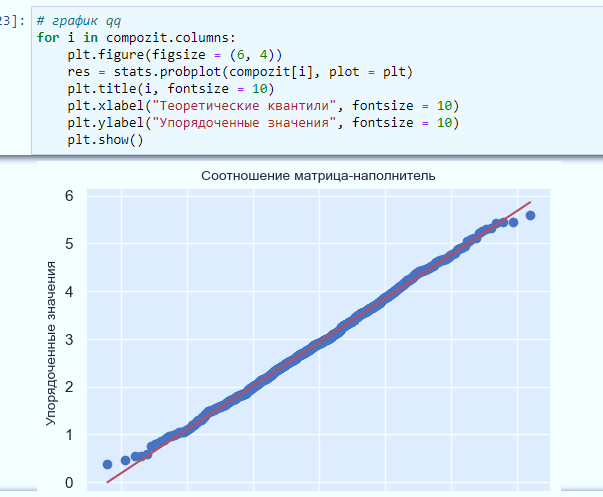


Рисунок 11 - графики «квантиль-квантиль»

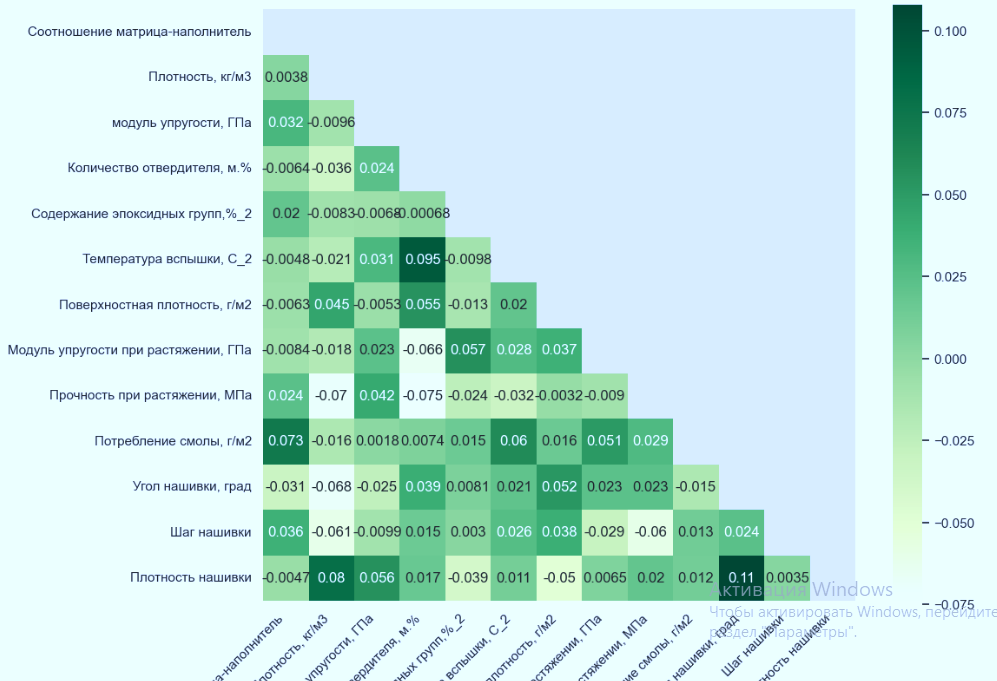


Рисунок 12 - Матрица корреляции

После обнаружения выбросов данные, значительно отличающиеся от выборки, будут полностью удалены. Для расчёта этих данных мы будем использовать метод межквартильного расстояния.

Данные объединённого датасета не имеют чётко выраженной зависимости, что подтверждает тепловая карта с матрицей корреляции и матрицы диаграмм рассеяния. Максимальная корреляция между плотностью нашивки и углом нашивки 0.11, значит, нет зависимости между этими данными. Корреляция между всеми параметрами очень близка к 0, то есть корреляционные связи между переменными не наблюдаются.

**2 Практическая часть**

**2.1 Предобработка данных**

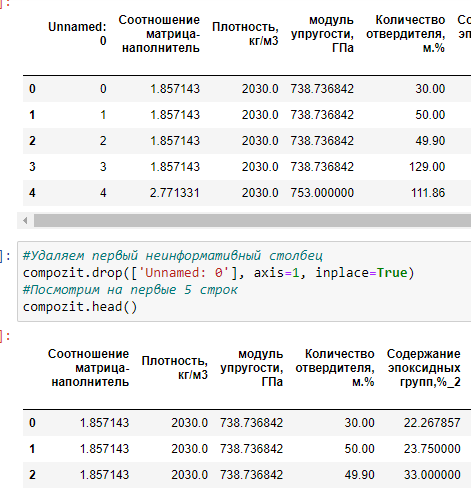
После объединения датасета, удалили первый неинформативный столбец с индексами.

Рисунок 13 – удаление первого столбца

Затем провели с помощью MinMaxScaler нормализацию всех данных.

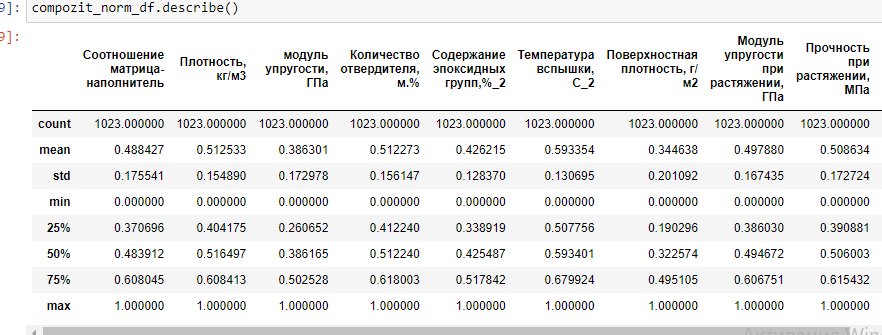


Рисунок 14 – нормализованный датасет

Потом методом межквартильного расстояния удалили выбросы.

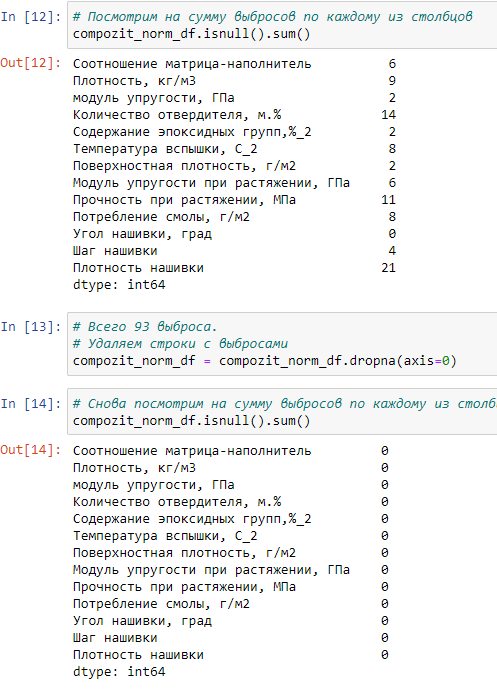


Рисунок 15 – очищенный от выбросов датасет

При работе с нейронными сетями для нормализации данных использовала Normalization.

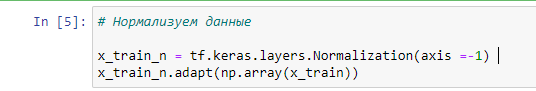


Рисунок 16 – нормализация датасета с помощью Normalization

**2.2 Разработка и обучение модели**

Разработка и обучение моделей машинного обучения осуществлялась для двух выходных параметров: «Прочность при растяжении» и «Модуль упругости при растяжении» отдельно. Для решения применила все методы, описанные выше.

Сначала создаем 1norm-модель линейной регрессии для признака «Модуль упругости при растяжении, Гпа», используя все остальные признаки

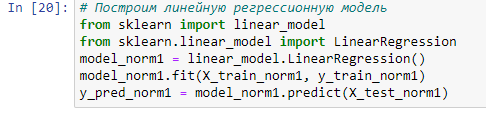


Рисунок 17 – линейная регрессионная модель

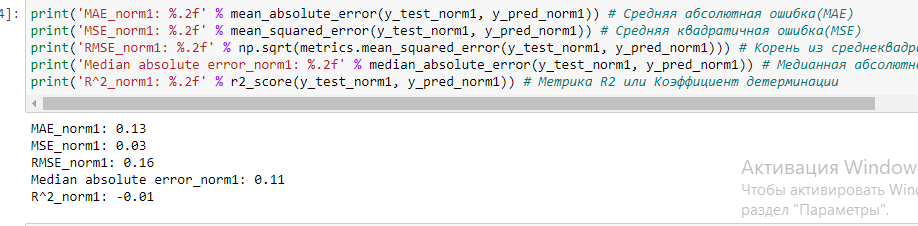


Рисунок 18 – метрики 1-й линейной регрессионной модели

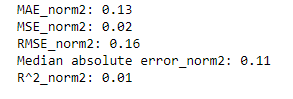
Получившаяся модель плохо выдает прогнозные значения на тестовой выборке. Поэтому необходимо построить новую модель model\_norm2, используя только 2 признака 'Содержание эпоксидных групп,%\_2' и 'Потребление смолы, г/м2'.

Рисунок 19 – метрики 2-й линейной регрессионной модели

Видно, что коэффициент детерминации стал положительным, хотя и слишком мал (0.01).

Затем попробуем построить модели линейной регрессии на StandardScaler-стандартизированных предварительно очищенных данных, используя для построения model\_std1-модели все признаки, а для model\_std2 - только 2 признака 'Содержание эпоксидных групп,%\_2' и 'Потребление смолы, г/м2'. Но коэффициент детерминации этих моделей не стал положительным.

Поэтому при создании 5-й модели линейной регрессии для определения значения "Прочность при растяжении, МПа", работаем на MinMaxScaler-нормализованных данных по одному признаку 'модуль упругости, ГПа'. Так как хороших корреляционных связей между признаками не наблюдается. Коэффициент детерминации этой модели не стал положительным.

Таким образом, не зависимо от того какие и какое количество признаков подаются на вход модели, данные модели линейной регрессии недостаточно справляются с прогнозами целевых признаков.

В 3 части своей работы построили ряд моделей, используя следующие методы: метод опорных векторов; К -ближайших соседей; дерево решений; случайный лес; градиентный бустинг; стохастический градиентный спуск.

Для прогнозирования 'Модуль упругости при растяжении, ГПа' разбила датасет на тестовую и обучающую выборки используя признаки 'Содержание эпоксидных групп,%\_2' и 'Потребление смолы, г/м2'. А для прогнозирования 'Прочность при растяжении, МПа' использовала 'модуль упругости, ГПа'. Для визуализации тестовых и прогнозных значений использовала plt.figure. При создании моделей К-ближайших соседей использовала 5 и 20 соседей. Лучше результаты оказались при меньшем количестве соседей.

Во всех созданных моделях коэффициент детерминации получился отрицательным. Таким образом, ни одна из рассмотренных моделей не годится для прогноза ‘Модуль упругости при растяжении, ГПа’ и 'Прочность при растяжении, МПа' по предоставленному набору данных.

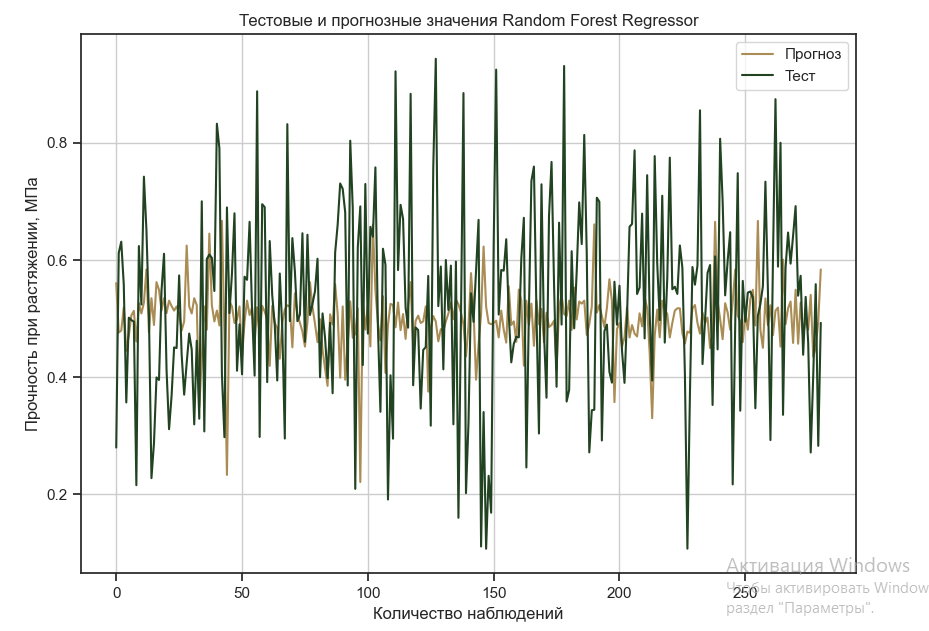


Рисунок 20 – Тестовые и прогнозные значения Прочности при растяжении методом Random Forest Regressor

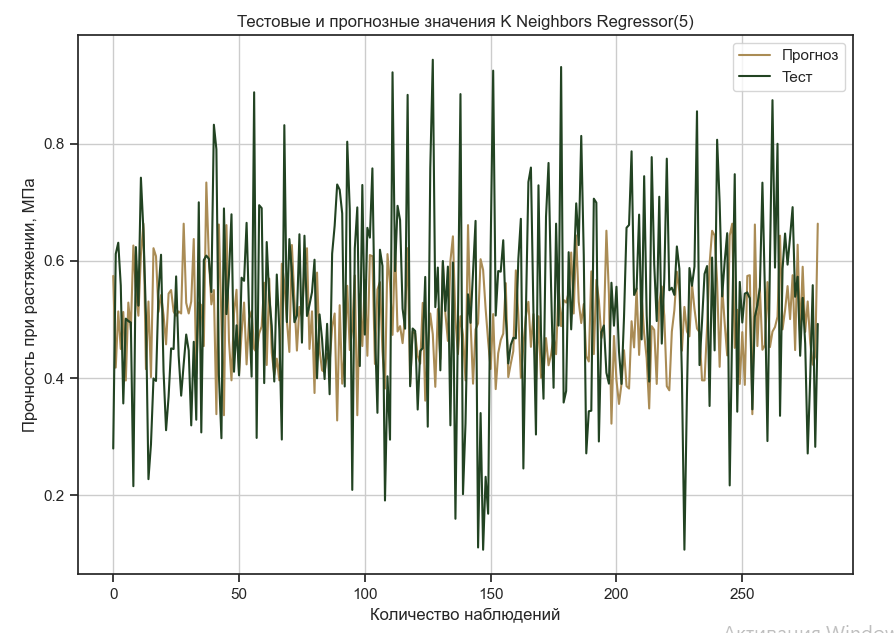


Рисунок 21 – Тестовые и прогнозные значения Прочности при растяжении методом К-Ближайших соседей

**2.3 Создание нейронной сети**

Обучение нейронной сети — это такой процесс, при котором происходит подбор оптимальных параметров модели, с точки зрения минимизации функционала ошибки. Начнём стоить нейронную сеть с помощью класса Sequential. Функции активации слоев relu, выходного слоя sigmoid.

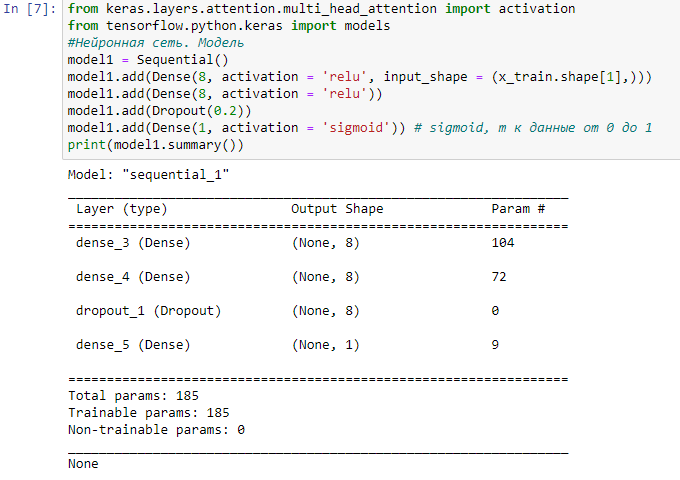


Рисунок 22 – Модель нейронной сети

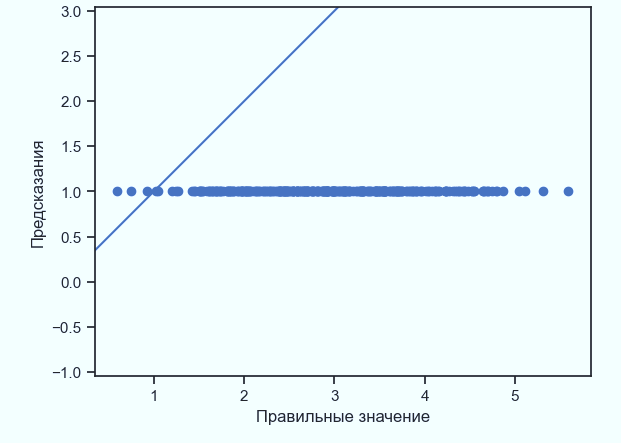
Экспериментировала с количеством слоев, количеством нейронов, эпох. Коэффициент детерминации везде оказывался отрицательным. И визуализация показывает несостоятельность моделей.

Рисунок 23 – Разброс предсказаний модели нейронной сети

**2.4 Экспорт модели для создания приложения**

В Google Colab удалось создать модель машинного обучения и экспортировать ее на локальный компьютер. Дальнейшая работа по созданию интерактивных веб-приложений была отложена в связи с нехваткой времени по изучению Flask и Streamlit. А также возникшими затруднениями при работе с этими приложениями в VS Code.

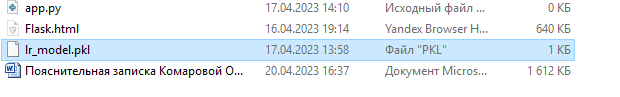


Рисунок 24 –Файл lr\_model.pkl

**2.5 Создание удаленного репозитория**

Для данного исследования был создан удаленный репозиторий на GitHub, который находится по адресу <https://github.com/GalinaSk21/BMSTU_VKR_SkorykhG> и на который были загружены материалы по проделанной работе.

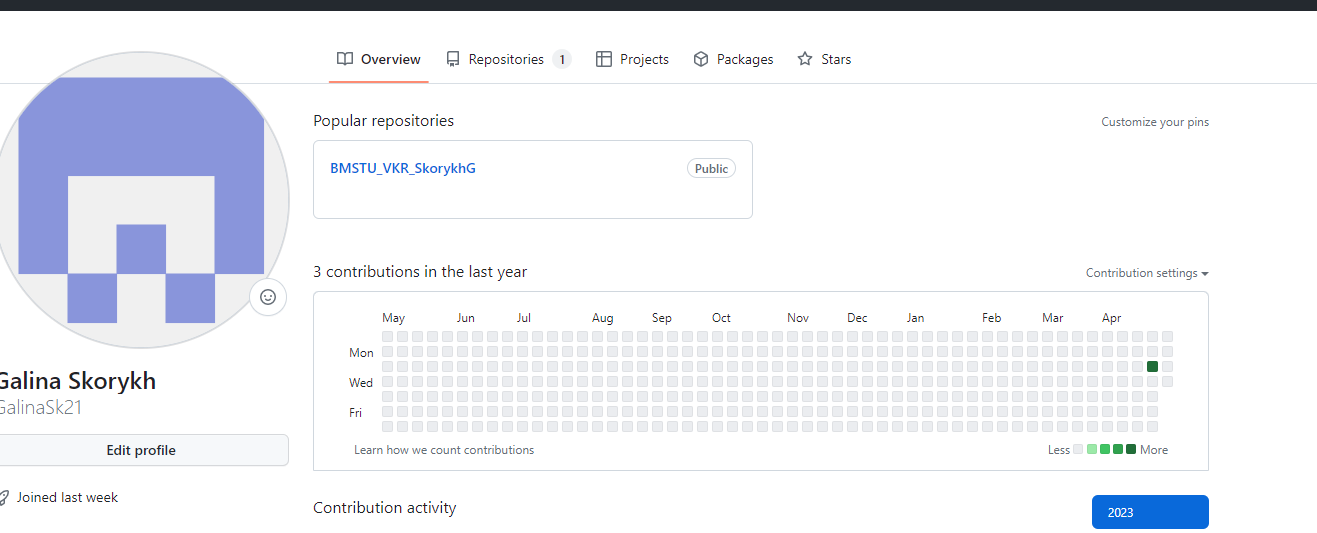


Рисунок 27 - часть страницы на github.com

**Заключение**

Данная исследовательская работа позволяет сделать некоторые основные выводы по теме. Распределение полученных данных в объединённом датасете близко к нормальному, но коэффициенты корреляции между парами признаков стремятся к нулю. Использованные при разработке моделей подходы не позволили получить сколько-нибудь достоверных прогнозов. Применённые модели регрессии не показали высокой эффективности в прогнозировании свойств композитов.

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.

Возможные причины неудачи:

- нечеткая постановка задачи, отсутствие дополнительной информации о зависимости признаков с точки зрения физики процесса. Незначимые признаки являются для модели шумом, и мешают найти зависимость целевых от значимых входных признаков;

- исследование предварительно обработанных данных. Возможно, на "сырых", не предобработанных данных можно было бы получить более качественные модели, воспользовавшись другими методами очистки и подготовки;

- недостаток знаний и опыта. Нейросети являются самым современным подходам к решению такого рода задач. Они способны находить скрытые и нелинейные зависимости в данных. Но выбор оптимальной архитектуры нейросети является неочевидной задачей.

Дальнейшие возможные пути решения этой задачи могли бы быть:

- углубиться в изучение нейросетей, попробовать различные архитектуры, параметры обучения и т.д.;

- провести отбор признаков разными методами. Испробовать методы уменьшения размерности, например метод главных компонент;

- после уменьшения размерности градиентный бустинг может улучшить свои результаты. Так же есть большой простор для подбора гиперпараметров для этого метода;

- проконсультироваться у экспертов в предметной области. Возможно, они могли бы поделиться знаниями, необходимыми для решения задачи.

**Библиографический список**

1 ГОСТ 32794-2014 Композиты полимерные. - Введ. 2015-09-01. - М.: Стандартинформ, 2015

2 ГОСТ Р 57970-2017 Композиты углеродные. Углеродные композиты, армированные углеродным волокном. – Введ. 2018-06-01. - М.: Стандартинформ, 2018

3 СП 28.13330.2012. «Защита строительных конструкций от коррозии». Актуализированная редакция СНиП 2.03.11-85. Дата введения 2013.01.01.

4 Библиотека Keras - инструмент глубокого обучения. Реализация нейронных сетей с помощью библиотек Theano и TensorFlow / пер. с англ. Слинкин А. А. - М.: ДМК Пресс, 2018. - 294 с.

5 Аллен Б. Дауни – Основы Python. Научитесь думать как программист / Аллен Б. Дауни ; пер. с англ. С. Черникова ; [науч. ред. А. Родионов]. — Москва: Манн, Иванов и Фербер, 2021. — 304 с.

6 Avdeeva A., Shlykova I., Perez M., Antonova M., Belyaeva S.Chemical properties of reinforcing fiberglass in aggressive media. MATEC Web of Conferences. 2016. Vol. 53. 01004.

7 Астапов Р.Л., Мухамадеева Р.М. Автоматизация подбора параметров машинного обучения и обучение модели машинного обучения // Актуальные научные исследования в современном мире. 2021. № 5-2 (73). С. 34-37.

8 Barabanshchikov Y., Belyaeva S., Avdeeva A.and Perez M. Fiberglass Reinforcement for Concrete (2015) Applied Mechanchanics and Materials, Pp. 475-481

9 Вичугова А.Data Preparation: полет нормальный – что такое нормализация данных и зачем она нужна [Электронный ресурс]: – Режим доступа: https://www.bigdataschool.ru/blog/нормализация-feature-transformation-data-preparation.html (дата обращения: 13.06.2022).

10 Билл Любанович. Простой Python. Современный стиль программирования. — СПб.: Питер, 2016. — 480 с.: ил. — (Серия «Бестселлеры O’Reilly»).

11 Бринк Х. Машинное обучение. / Х. Бринк, Дж. Ричардс, М. Феверолф. – СПб.: Питер,2017. 336 с.

12 Вандер Плас Дж. Python для сложных задач: наука о данных и машинное обучение. - СПб.: Питер, 2018. - 576 с.

13 Гиздатуллин А.Р., Хозин В.Г., Куклин А.Н., Хуснутдинов А.М. «Особенности испытаний и характер разрушения полимеркомпозитной арматуры».

14 Горбунов П.М., Мацкевич Ю.А., Чубарь А.В. Машинное обучение. Автоматизация подбора модели машинного обучения // Робототехника и искусственный интеллект. 2021. С. 155-160.

15 Джалилов Ш.А. Метод расчета параметров множественной линейной регрессии // Достижения науки и образования. 2020. № 3 (57). С. 24-28.

16 Джулли, Пал: Библиотека Keras - инструмент глубокого обучения / пер. с англ. А. А. Слинкин.- ДМК Пресс, 2017. – 249 с.

17 Жерон, Орельен. Прикладное машинное обучение с помощью Scikit-Learn и TensorFlow: концепции, инструменты и техники для создания интеллектуальных систем.Пер. с англ. - СпБ.: ООО Альфа-книга: 2018. - 688 с.

18 Flach P. Machine learning. The art and science of building algorithms. pp. 118-142.

19 Кузнецов И.Н. Пример решения задачи множественной регрессии с помощью Python [Электронный ресурс]: – Режим доступа: https://habr.com/ru/post/206306/ (дата обращения: 13.06.2022).

20 Makusheva N.Yu., Kolosova N.B. Comparative analysis of metal reinforcement and fibre-reinforced plastic rebar Construction of Unique Buildings and Structures, 2014, No10 (25) Pp. 60-72

21 Плас Дж. Вандер. Python для сложных задач: наука о данных и машинное обучение. / Дж. Вандер Плас. – СПб.: Питер, 2018. 576 с.

22 Сохина С.А., Немченко С.А. Машинное обучение. Методы машинного обучения // Современная наука в условиях модернизационных процессов: проблемы, реалии, перспективы. 2021. С. 165-168.

23 Субботина С.А., Шлыкова И.Д., Авдеева А.А., Одинокова Г.В., Соколова Н.В. Виды композитных материалов: стеклопластик, углепластик, базальтопластик // Синергия наук. 2017. № 18. − С. 641-645.

24 Таршхоева Ж.Т. Zзык программирования Python. Библиотеки Python // Молодой ученый. 2021. № 5 (347). С. 20-21.

25 Токарев В. Сравнение арматурных прутьев из базальтопластика и углепластика [Электронный ресурс]: – Режим доступа: http://cemgid.ru/sravnenie-armaturnyx-prutev-iz-bazaltoplastika-i-ugleplastika.html (дата обращения: 12.06.2022).

26 Щелконогов А.Н. Разработка простейших нейросетей в keras // Математическое и программное обеспечение вычислительных систем. 2019. С. 51-53.