

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 Боркунов Юрий Александрович

Москва, 2023 год

# **Содержание**

1. Введение…………………………………………………………………3
2. Аналитическая часть
   1. Постановка задачи
   2. Описание используемых методов………………………
   3. Разведочный анализ данных………………………………
3. Практическая часть
   1. Предобработка данных……………………………………
   2. Разработка, обучение и тестирование моделей машинного обучения
   3. Написание нейронной сети, которая будет рекомендовать соотношение матрица-наполнитель………………………
   4. Разработка приложения……………………………………
   5. Создание удаленного репозитория и загрузка результатов работы на него…………………………………………………
4. Заключение……………………………………………………………
5. Список литературы…………………………………………………

Введение

Композиционный материал - неоднородный сплошной материал, состоящий из двух или более компонентов, среди которых можно выделить армирующие элементы, обеспечивающие необходимые механические характеристики материала, и матрицу, обеспечивающую совместную работу армирующих элементов. Механическое поведение композита определяется соотношением свойств армирующих элементов и матрицы, а также прочностью связи между ними. Эффективность и работоспособность материала зависят от правильного выбора исходных компонентов и технологии их совмещения, призванной обеспечить прочную связь между компонентами при сохранении их первоначальных характеристик. В результате совмещения армирующих элементов и матрицы образуется комплекс свойств композита, не только отражающий исходные характеристики его компонентов, но и включающий свойства, которыми изолированные компоненты не обладают. В частности, наличие границ раздела между армирующими элементами и матрицей существенно повышает трещиностойкость материала, и в композитах, в отличие от металлов, повышение статической прочности приводит не к снижению, а, как правило, к повышению характеристик вязкости разрушения.

Преимущества композиционных материалов:

-высокая удельная прочность;

-высокая жёсткость (модуль упругости 130…140 ГПа);

-высокая износостойкость;

-высокая усталостная прочность;

Из КМ возможно изготовить размеростабильные конструкции, причём, разные классы композитов могут обладать одним или несколькими преимуществами.

Наиболее частые недостатки композиционных материалов:

-высокая стоимость;

-анизотропия свойств;

-повышенная наукоёмкость производства, необходимость специального дорогостоящего оборудования и сырья, а следовательно развитого промышленного производства и научной базы страны.

Классификация композиционных материалов.

Композиты - многокомпонентные материалы, состоящие из полимерной, металлической., углеродной, керамической или др. основы (матрицы), армированной наполнителями из волокон, нитевидных кристаллов, тонкодиспeрсных частиц и др. Путем подбора состава и свойств наполнителя и матрицы (связующего), их соотношения, ориентации наполнителя можно получить материалы с требуемым сочетанием эксплуатационных и технологических свойств. Использование в одном материале нескольких матриц (полиматричные композиционные материалы) или наполнителей различной природы (гибридные композиционные материалы) значительно расширяет возможности регулирования свойств композиционных материалов. Армирующие наполнители воспринимают основную долю нагрузки композиционных материалов.

По структуре наполнителя композиционные материалы подразделяют на волокнистые (армированы волокнами и нитевидными кристаллами), слоистые (армированы пленками, пластинками, слоистыми наполнителями), дисперсноармированные, или дисперсно-упрочненные (с наполнителем в виде тонкодисперсных частиц). Матрица в композиционных материалах обеспечивает монолитность материала, передачу и распределение напряжения в наполнителе, определяет тепло-, влаго-, огне- и хим. стойкость.

По природе матричного материала различают полимерные, металлические, углеродные, керамические и др. композиты.

Основным недостатком композиционных материалов с одно и двумерным армированием является низкое сопротивление межслойному сдвигу и поперечному обрыву. Этого лишены материалы с объемным армированием.

методов изготовления полимерных и металлических волокнистых и слоистых композиционных материалов - выращивание кристаллов наполнителя в матрице непосредственно в процессе изготовления деталей. Такой метод применяют, напр., при создании эвтектических жаропрочных сплавов на основе Ni и Со. Легирование расплавов карбидными и интерметаллическими соединениями, образующими при охлаждении в контролируемых условиях волокнистые или пластинчатые кристаллы, приводит к упрочнению сплавов и позволяет повысить температуру их эксплуатации на 60-80oС. Композиционные материалы на основе углерода сочетают низкую плотность с высокой теплопроводностью, хим. стойкостью, постоянством размеров при резких перепадах температур, а также с возрастанием прочности и модуля упругости при нагреве до 2000 °С в инертной среде. Высокопрочные композиционные материалы на основе керамики получают при армировании волокнистыми наполнителями, а также металлическими и керамическими дисперсными частицами. Армирование непрерывными волокнами SiC позволяет получать композиционные материалы, характеризующиеся повышенной вязкостью, прочностью на изгиб и высокой стойкостью к окислению при высоких температурах.

Использование композитов в качестве конструкционных, теплозащитных, антифрикционных, радио - и электротехнических и др. материалов позволяет снизить массу конструкции, повысить ресурсы и мощности машин и агрегатов, создать принципиально новые узлы, детали и конструкции. Все виды композиционные материалы применяют в химической, текстильной, горнорудной, металлургической промышленности, машиностроении, на транспорте, для изготовления спортивного снаряжения и др.

Экономическая эффективность применения композиционных материалов.

Области применения композиционных материалов не ограничены. Они применяются в авиации для высоконагруженных деталей (обшивки, лонжеронов, нервюр, панелей, лопаток компрессора и турбины и т. д.), в космической технике для узлов силовых конструкций аппаратов, для элементов жесткости, панелей, в автомобилестроении для облегчения кузовов, рессор, рам, панелей кузовов, бамперов и т. д., в горной промышленности (буровой инструмент, детали комбайнов и т. д.), в гражданском строительстве (пролеты мостов, элементы сборных конструкций высотных сооружений и т. д.) и в других областях народного хозяйства.

Применение композиционных материалов обеспечивает новый качественный скачек в увеличении мощности двигателей, энергетических и транспортных установок, уменьшении массы машин и приборов. Композиционные материалы с неметаллической матрицей, а именно полимерные карбоволокниты используют в судо- и автомобилестроении (кузова гоночных машин, шасси, гребные винты); из них изготовляют подшипники, панели отопления, спортивный инвентарь, части ЭВМ. Высокомодульные карбоволокниты применяют для изготовления деталей авиационной техники, аппаратуры для химической промышленности, в рентгеновском оборудовании и другом. Карбоволокниты с углеродной матрицей заменяют различные типы графитов. Они применяются для тепловой защиты, дисков авиационных тормозов, химически стойкой аппаратуры. Изделия из бороволокнитов применяют в авиационной и космической технике (профили, панели, роторы и лопатки компрессоров, лопасти винтов, трансмиссионные валы вертолетов и т. д.). Органоволокниты применяют в качестве изоляционного и конструкционного материала в электрорадиопромышленности, авиационной технике и т. д.

2. Аналитическая часть

2.1.Постановка задачи

В процессе выполнения данной выпускной квалификационной работы необходимо продемонстрировать навыки, полученные в данном курсе на примере решения задачи прогнозирования конечных свойств новых композиционных материалов.

А, именно:

-провести анализ и предобработку данных;

-описать методы которые предполагается использовать для решения поставленной задачи;

-разработать, обучить и протестировать модель;

-написать нейронную сеть, которая будет рекомендовать соотношение матрица-наполнитель;

-разработать приложение.

2.2.Описание используемых методов

Перед нами стоит задача регрессии – прогноз на основе выборки объектов с различными признаками. Поэтому в рамках решения данной задачи были применены следующие методы:

- Линейная регрессия;

- Метод k-ближайших соседей;

- Случайный лес.

Линейная регрессия в Машинном обучении – это подход к моделированию отношений между Целевой переменной (Target Variable) и одной или несколькими "предсказывающими" переменными (Predictor Variable). Проще говоря, это «линия наилучшего соответствия», которая помогает спрогнозировать положение других точек в будущем.

Метод k-ближайших соседей - метрический алгоритм для автоматической [классификации](https://ru.wikipedia.org/wiki/%D0%97%D0%B0%D0%B4%D0%B0%D1%87%D0%B0_%D0%BA%D0%BB%D0%B0%D1%81%D1%81%D0%B8%D1%84%D0%B8%D0%BA%D0%B0%D1%86%D0%B8%D0%B8) объектов или [регрессии](https://ru.wikipedia.org/wiki/%D0%A0%D0%B5%D0%B3%D1%80%D0%B5%D1%81%D1%81%D0%B8%D1%8F_(%D0%BC%D0%B0%D1%82%D0%B5%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0)). На интуитивном уровне суть метода проста: посмотри на соседей вокруг, какие из них преобладают, таковым ты и являешься.

Случайный лес, как и следует из его названия, состоит из большого количества отдельных деревьев решений, которые работают как [ансамбль методов](https://ru.wikipedia.org/wiki/%D0%90%D0%BD%D1%81%D0%B0%D0%BC%D0%B1%D0%BB%D1%8C_%D0%BC%D0%B5%D1%82%D0%BE%D0%B4%D0%BE%D0%B2_(%D0%BE%D0%B1%D1%83%D1%87%D0%B5%D0%BD%D0%B8%D0%B5_%D0%BC%D0%B0%D1%88%D0%B8%D0%BD)). Каждое дерево в случайном лесу возвращает прогноз класса, и класс с наибольшим количеством голосов становится прогнозом леса. Фундаментальная концепция в основе случайного леса проста, но сильна — это мудрость толпы. Большое число относительно некоррелированных деревьев, работающих совместно, будет превосходить любую из их отдельных составляющих.

2.3.Разведочный анализ данных

В исходных данных мы имеем датасет с двумя Excel-таблицами. Далее с ними мы и будем работать.

Описательная статистика датасета.

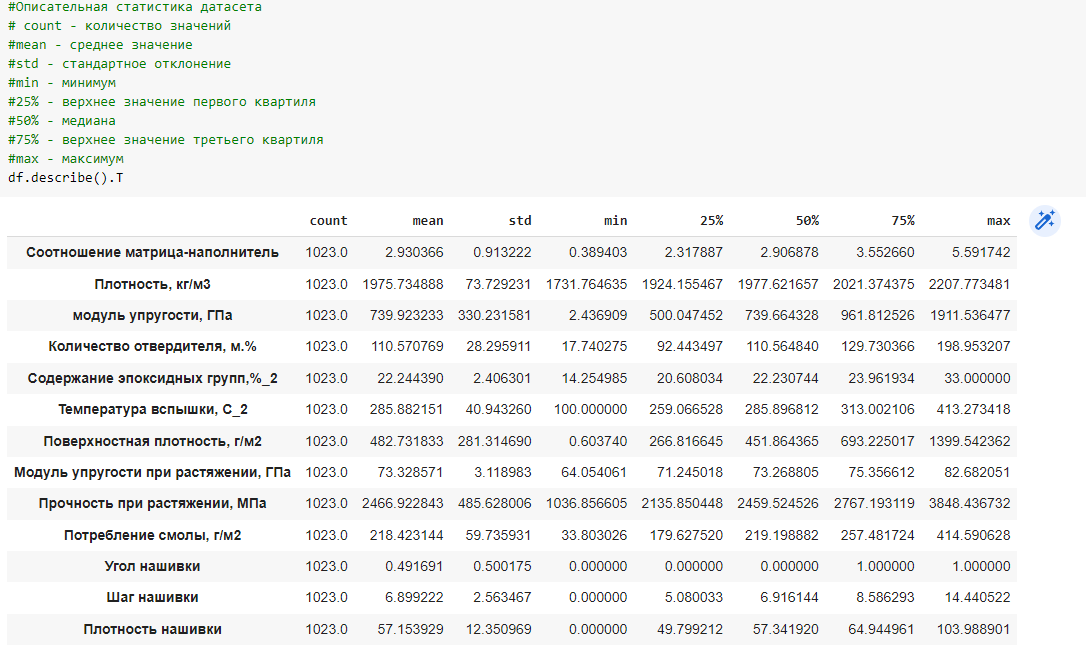


Рисунок1- таблица с основными параметрами датасета

Произведено построение диаграмм с целью выявления характера взаимосвязи между переменными и последующего подбора моделей машинного обучения, подходящих для решения задачи.

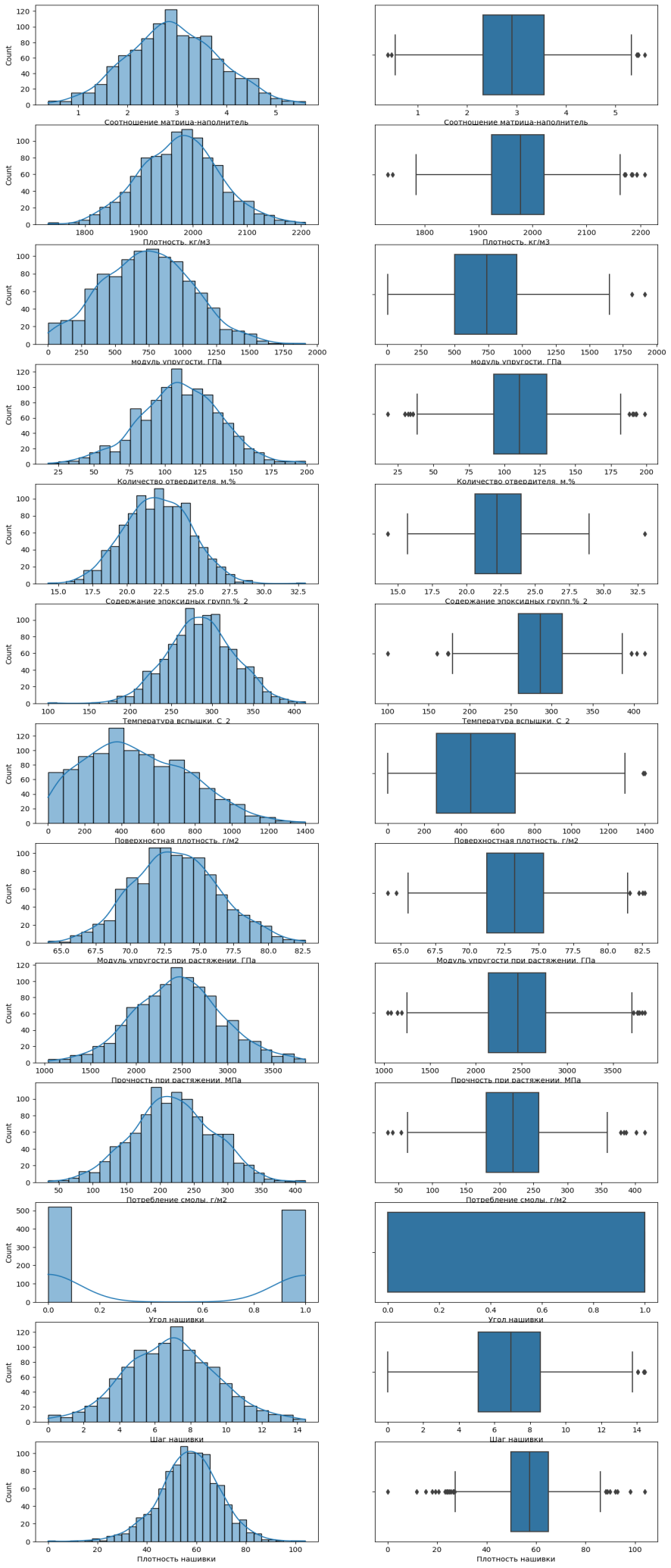


Рисунок 2 - Гистограммы распределения параметров переменных и "ящик с усами"

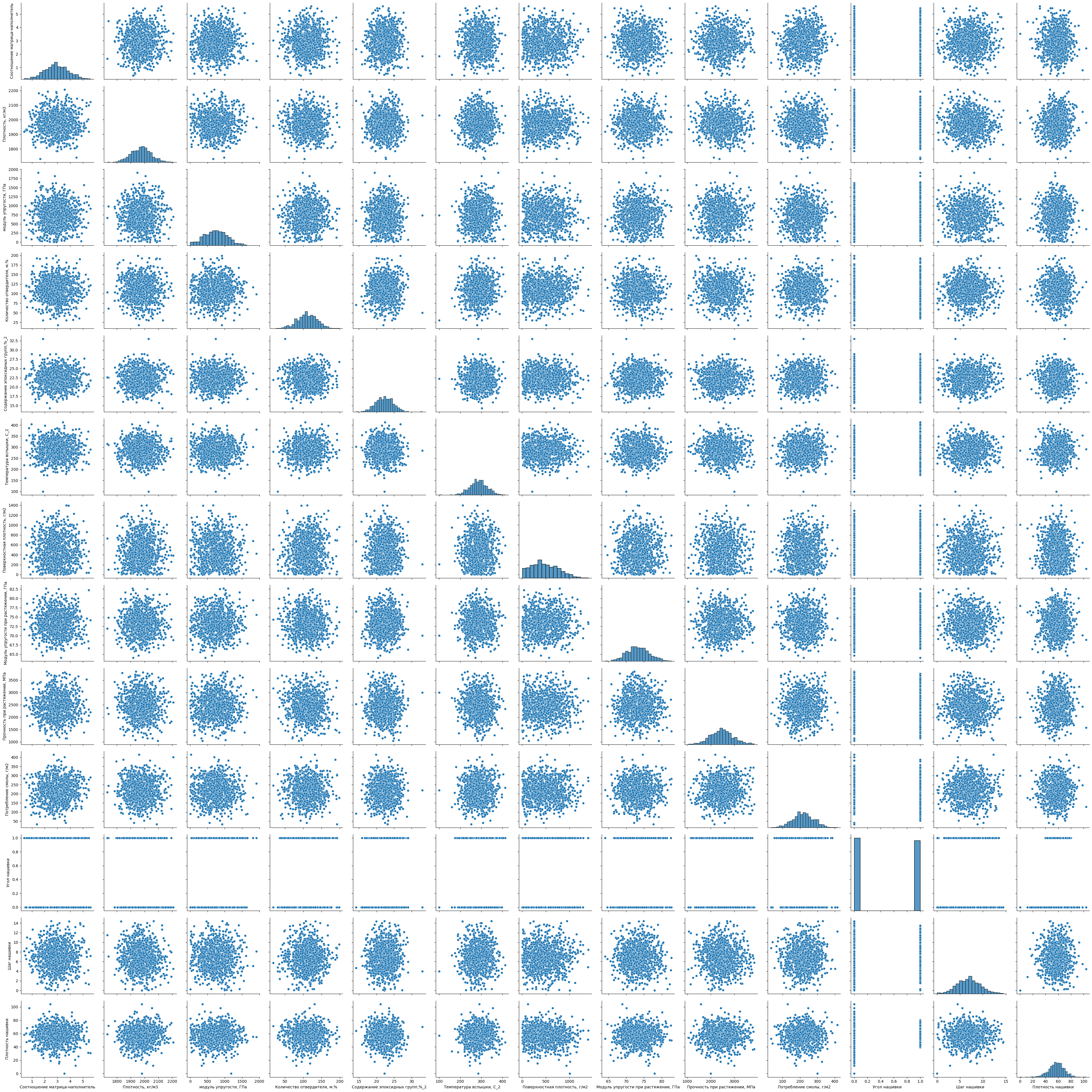


Рисунок 3 - Попарные графики рассеяния точек

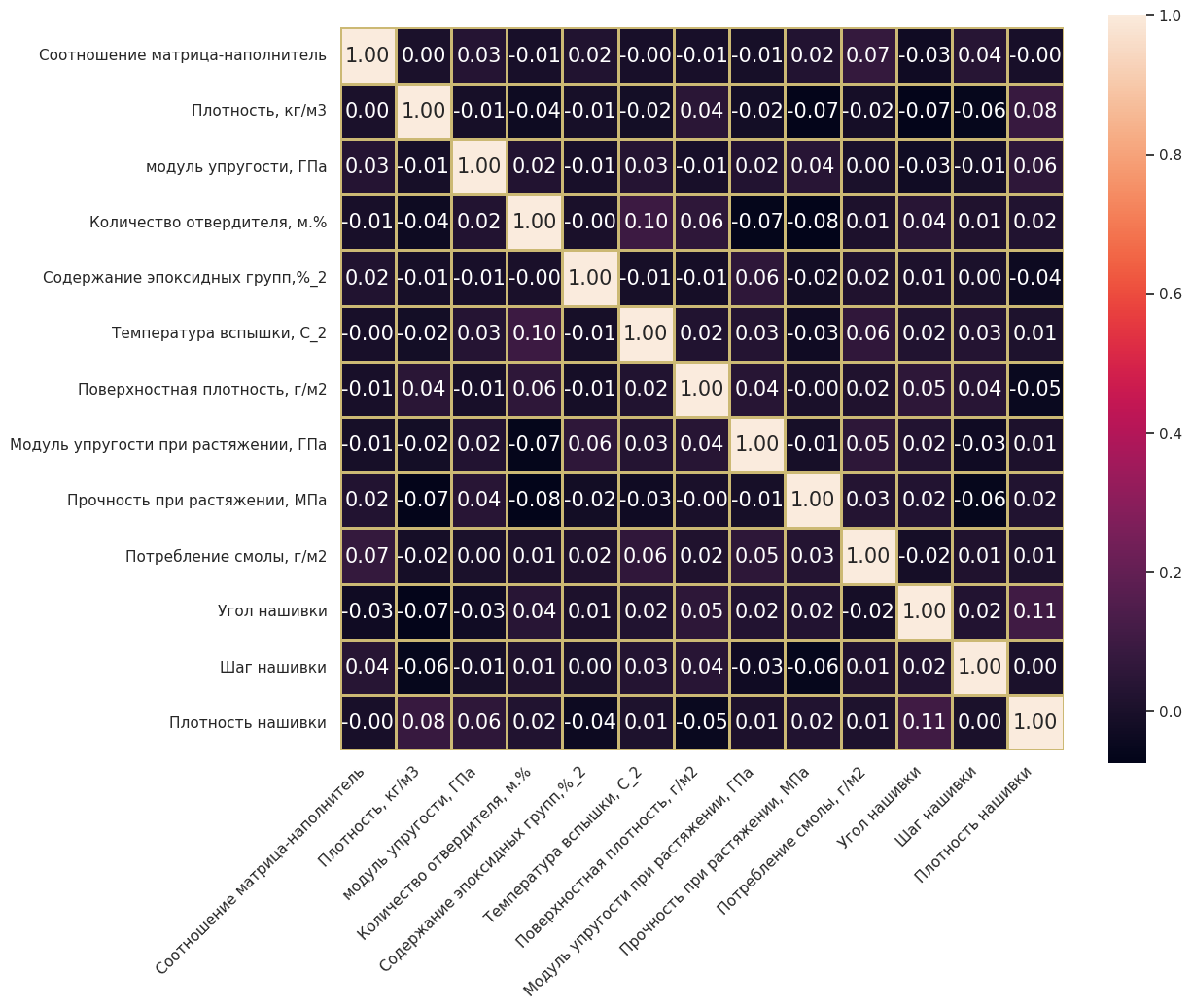


Рисунок 4 - Корреляционная матрица,представляющую собой

тепловую карту, в которой измеряется линейная зависимость между парой признаков

Произведено обнаружение выбросов методом трех сигм и межквартильного расстояния. Правило трех сигм (3-sigma rule) - правило, утверждающее, что вероятность того, что случайная величина отклонится от своего математического ожидания более чем на три среднеквадратических отклонения. Межквартильный диапазон набора данных, часто сокращенно IQR, представляет собой разницу между первым квартилем (25-й процентиль) и третьим квартилем (75-й процентиль) набора данных.

Далее осуществлена очистка данных от выбросов методом межквартильного расстояния.

3.Практическая часть

3.1.Предобработка данных

Произведено объединение двух Excel-таблиц с данными в один датасет.

В параметре "угол нашивки,Град" только 2 значения, поэтому для удобства мы определили переменные как 0 и 1 и переименовали в “угол нашивки”.



Рисунок 5 – Объединённый датасет

Произведена нормализация данных методом MinMaxScaler

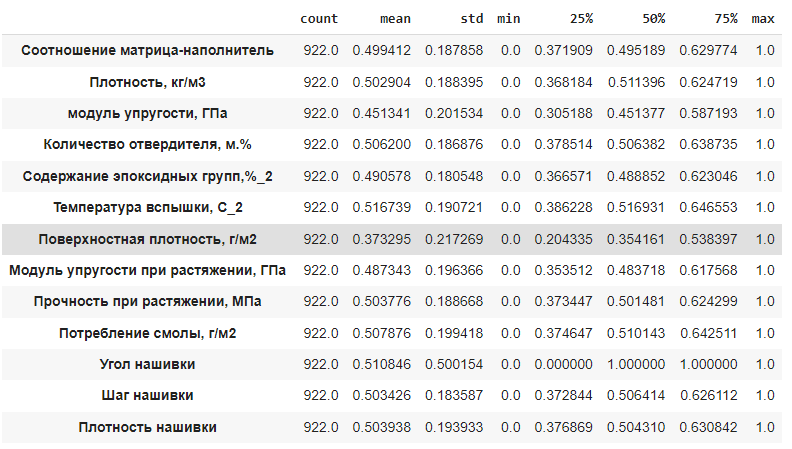


Рисунок 6 – Нормализованный датасет

3.2.Разработка и обучение модели

Перед нами стоит задача регрессии – прогноз на основе выборки объектов с различными признаками. Построим модели для прогноза модуля упругости при растяжении и прочности при растяжении.

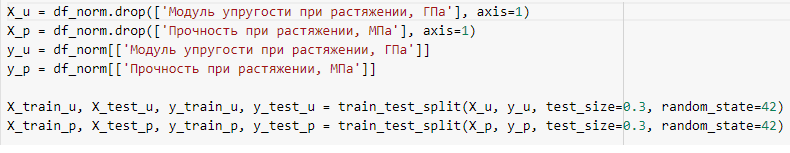


Рисунок 6 - Разбиваем данные на обучающую и тестовую выборки

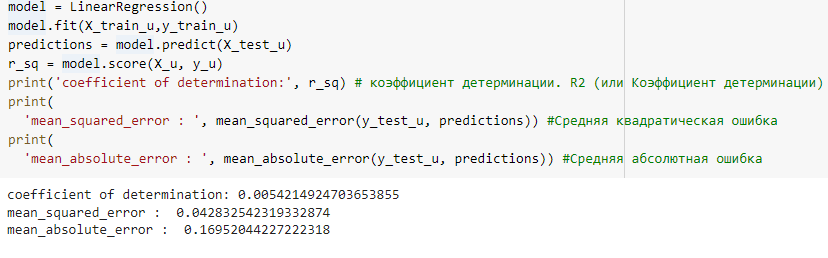
Метод линейной регрессии.

Определим несколько параметров:

- R2(или Коэффициент детерминации)- это статистическая мера, которая показывает степень вариации зависимой переменной из-за независимой переменной;

- Средняя квадратическая ошибка;

- Средняя абсолютная ошибка.



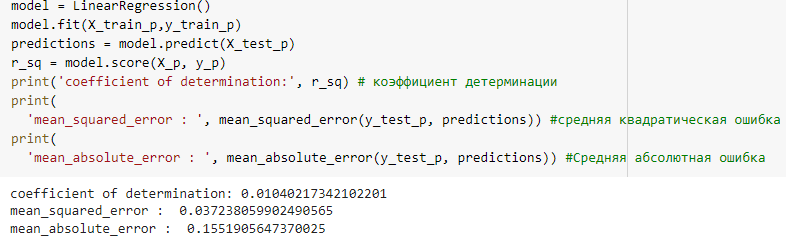


Рисунок 7 – Оценки производительности модели для модуля упругости при растяжении и прочности при растяжении

Получим прогнозы и выведем соответствующие графики.

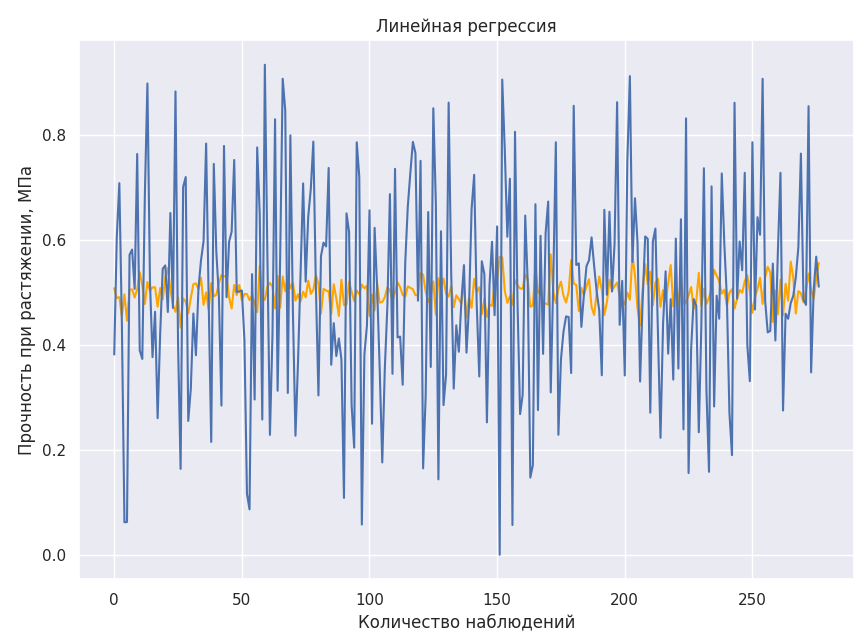
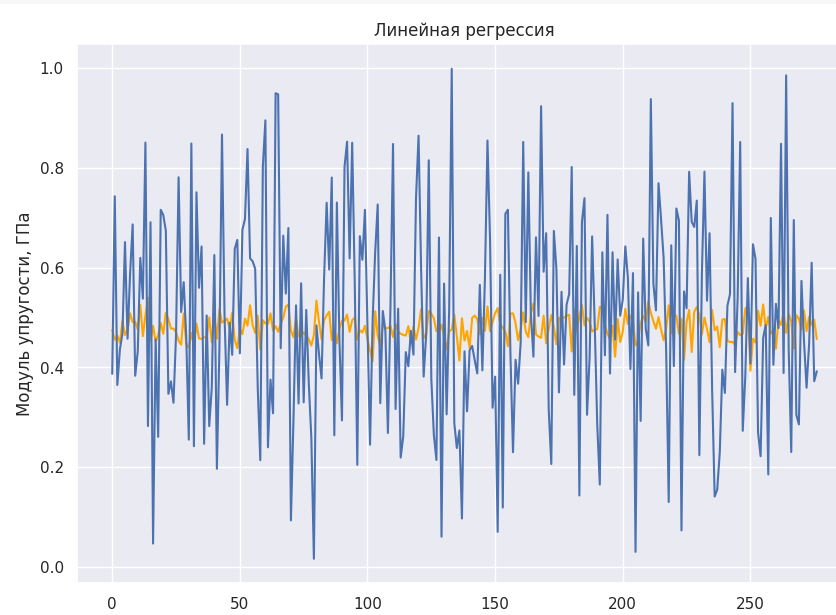


Рисунок 8 - графики прогноза и тестовых данных

Метод k-ближайших соседей.

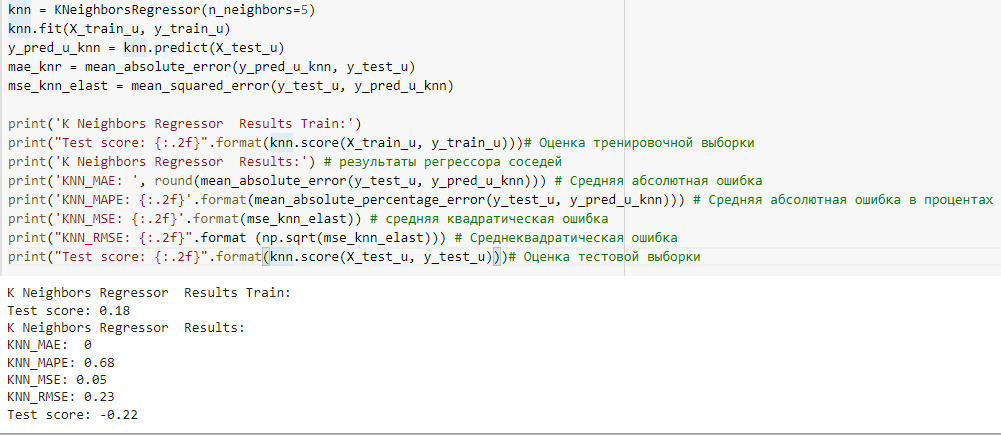


Рисунок 8 – оценочные характеристики модели

Здесь будет применена функция GridSearch — поиск лучших параметров в фиксированной сетке возможных значений.

CV – перекрёстная проверка (кросс-валидация, Cross-validation), метод, который показывает, что модель не переобучилась.

Результаты:

R2-score Модуль упругости при растяжении: -0.014

R2-score Прочность при растяжении: -0.002

R2<0-разработанная модель даёт прогноз даже хуже, чем простое усреднение.

Случайный лес.

Здесь будет применена функция RandomizedSearchCV. Она реализует метод «подгонки» и «оценки». В отличие от GridSearchCV, проверяются не все значения параметров, а из указанных распределений выбирается фиксированное количество значений параметров.

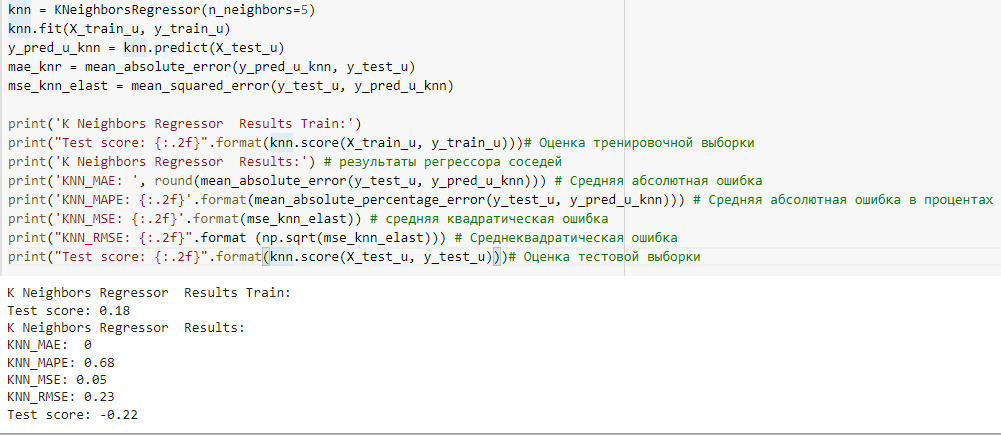


Рисунок 9 – оценочные метрики для модуля упругости при растяжении

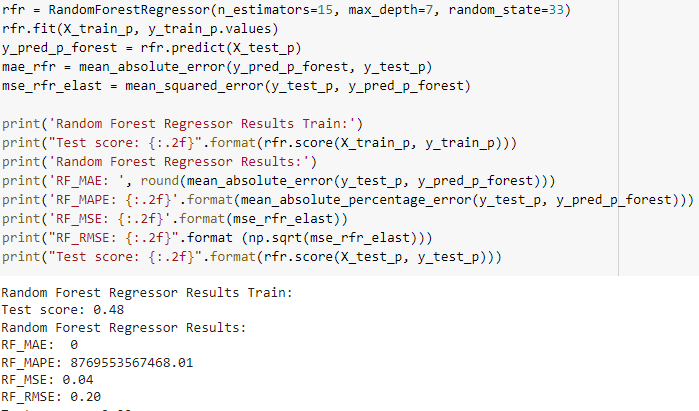


Рисунок 10 – оценочные метрики для прочности при растяжении

R2-score Модуль упругости при растяжении: -0.013

R2-score Прочность при растяжении: -0.002

3.3.Написание нейронной сети, которая будет рекомендовать соотношение матрица-наполнитель.

Сформируем выборки:

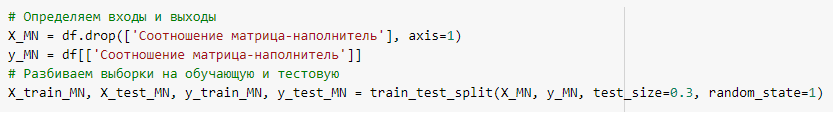


Рисунок 11 - Выборки

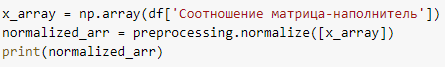


Рисунок 12 – Нормализация

Применим функцию активации SELU,она сочетает в себе оба преимущества классического RELU со свойствами самонормализации.

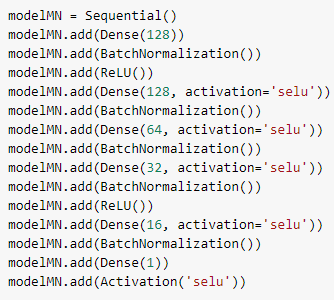


Рисунок 12 – Слои нейронной сети

Построение модели и определение её параметров.

SGD-стохастический градиентный спуск с мини-пакетами — вариант, при котором коэффициенты меняются после обсчета N элементов выборки, то есть для каждой тренировочной итерации алгоритм выбирает случайное подмножество набора данных. Частота обновления параметров выше, меньше требуется оперативной памяти, эффективность вычислений высокая.



Рисунок 13 – Формирование нейросети

Оценим модель.

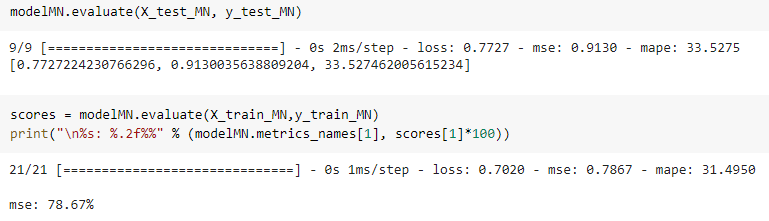


Рисунок 14 – потери, средняя квадратическая ошибка, средняя абсолютная ошибка в процентах.

Тестовые и прогнозные значения

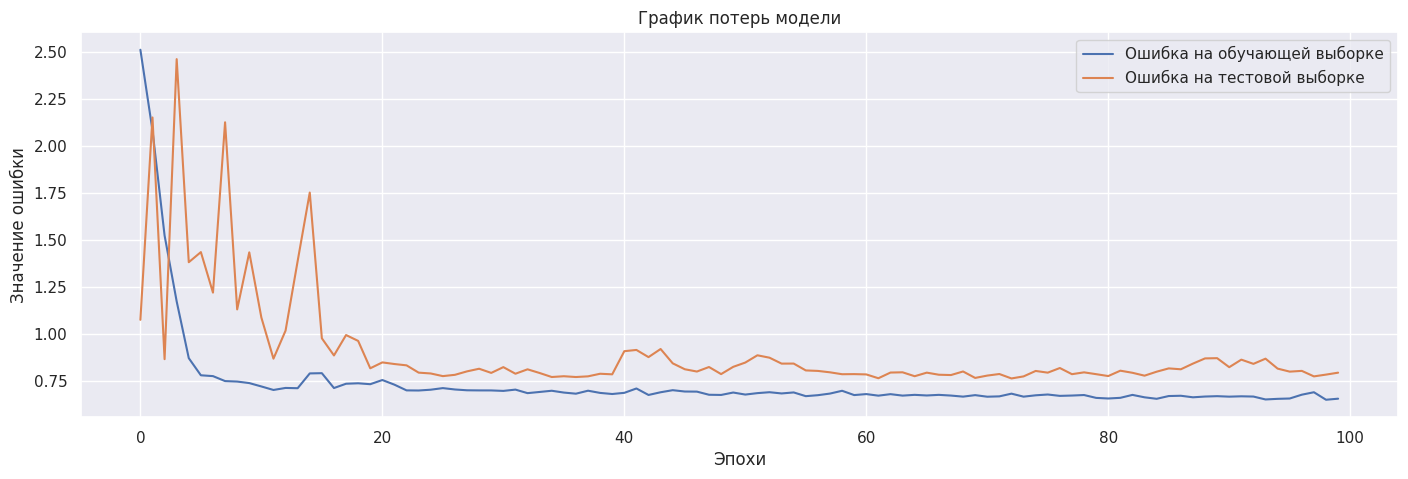


Рисунок 15

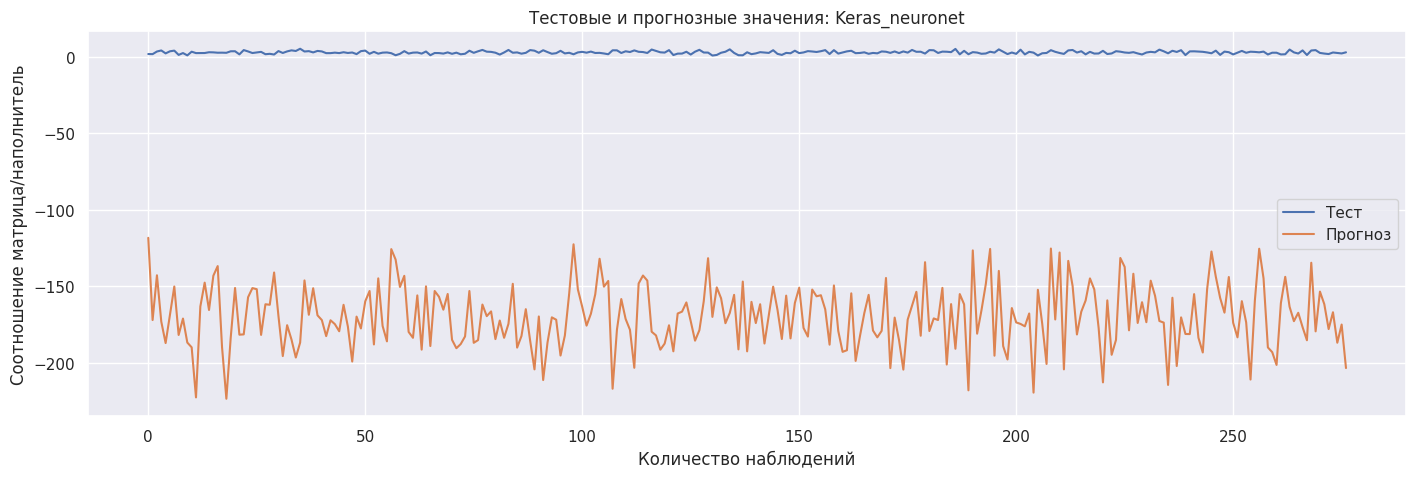


Рисунок 16

Средняя абсолютная ошибка.

9/9 [==============================] - 0s 2ms/step –

loss: 0.7727 - mse: 0.9130 - mape: 33.5275

Model MAE: [0.7727224230766296, 0.9130035638809204, 33.527462005615234]

MAE среднего значения: Соотношение матрица-наполнитель 0.741552

Заключение

В процессе выполнения данной выпускной квалификационной работы были продемонстрированы навыки, полученные в данном курсе на примере решения задачи прогнозирования конечных свойств новых композиционных материалов.

Проведён анализ и предобработка данных, описаны методы которые предполагается использовать для решения поставленной задачи, разработаны, обучены и протестированы модели, написана нейронная сеть, которая будет рекомендовать соотношение матрица-наполнитель, разработано приложение.

Были сделаны следующие выводы. 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.

Созданная нейронная сеть, которая будет рекомендовать соотношение матрица-наполнитель также далека от идеала.

Список литературы

1. ГрасД. Data Science. Наука о данных с нуля: Пер. с англ. - 2-е изд., перераб. и доп. - СПб.: БХВ-Петербурr, 2021. - 416 с.: ил.

2. Силен Дэви, Мейсман Арно, Али Мохамед. Основы Data Science и Big Data. Python и наука о данных. – СПб.: Питер, 2017. – 336 с.: ил.

3. Композиционные материалы : учебное пособие для вузов / Д. А. Иванов, А. И. Ситников, С. Д. Шляпин ; под редакцией А. А. Ильина. — Москва : Издательство Юрайт, 2019 — 253 с. — (Высшее образование). — Текст : непосредственный.

4. Документация по языку программирования python: – Режим доступа: <https://docs.python.org/3.8/index.html>.

5. Документация по библиотеке matplotlib: – Режим доступа: <https://matplotlib.org/stable/users/index.html>.

6. Документация по библиотеке seaborn: – Режим доступа: <https://seaborn.pydata.org/tutorial.html>.

7. Документация по библиотеке sklearn: – Режим доступа: <https://scikit-learn.org/stable/user_guide.html>.

8. Документация по библиотеке keras: – Режим доступа: <https://keras.io/api/>.

9. Документация по библиотеке numpy: – Режим доступа: <https://numpy.org/doc/1.22/user/index.html#user>.

19. Документация по библиотеке pandas: – Режим доступа: <https://pandas.pydata.org/docs/user_guide/index.html#user-guide>.