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»

Слушатель Бондарчук Анна Алексеевна

Москва, 2022

# Оглавление

[Введение 3](#_Toc106468218)

[**1. Аналитическая часть** 5](#_Toc106468219)

[**2. Практическая часть** 2](#_Toc106468220)0

[Заключение 2](#_Toc106468221)9

Список литературы …………………………………………………………….31

# **Введение**

Композиционные материалы — это искусственно созданные материалы, состоящие из нескольких других с четкой границей между ними. Композиты обладают теми свойствами, которые не наблюдаются у компонентов по отдельности. При этом композиты являются монолитным материалом, т. е. компоненты материала неотделимы друг от друга без разрушения конструкции в целом. Яркий пример композита - железобетон. Бетон прекрасно сопротивляется сжатию, но плохо растяжению. Стальная арматура внутри бетона компенсирует его неспособность сопротивляться сжатию, формируя тем самым новые, уникальные свойства. Современные композиты изготавливаются из других материалов: полимеры, керамика, стеклянные и углеродные волокна, но данный принцип сохраняется. У такого подхода есть и недостаток: даже если мы знаем характеристики исходных компонентов, определить характеристики композита, состоящего из этих компонентов, достаточно проблематично. Для решения этой проблемы есть два пути: физические испытания образцов материалов, или прогнозирование характеристик. Суть прогнозирования заключается в симуляции представительного элемента объема композита, на основе данных о характеристиках входящих компонентов (связующего и армирующего компонента).

**На входе** имеются данные о начальных свойствах компонентов композиционных материалов (количество связующего, наполнителя, температурный режим отверждения и т.д.). На выходе необходимо спрогнозировать ряд конечных свойств получаемых композиционных материалов. Кейс основан на реальных производственных задачах Центра НТИ «Цифровое материаловедение: новые материалы и вещества» (структурное подразделение МГТУ им. Н.Э. Баумана).

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

В процессе исследовательской работы были разработаны несколько моделей, способные прогнозировать модули упругости при растяжении и прочности при растяжении, а также была создана нейронная сеть, предлагающая соотношение «матрицы - наполнитель».

# **1. Аналитическая часть**

**1.1. Постановка задачи**

Композиционный материал - неоднородный сплошной материал, состоящий из двух или более компонентов, среди которых можно выделить армирующие элементы, обеспечивающие необходимые механические характеристики материала, и матрицу (или связующее), обеспечивающую совместную работу армирующих элементов.

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.

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.

Главное преимущество композиционных материалов в том, что материал и конструкция создается одновременно. Исключение препреги которые являются полуфабрикатом для изготовления конструкций.

Композиционные материалы создаются под выполнение конкретных задач, поэтому не могут вмещать в себя все возможные преимущества, но, проектируя новый композит, ему задают характеристики значительно превосходящие характеристики традиционных материалов.

Преимущества композиционных материалов:

* высокая удельная прочность
* высокая жёсткость (модуль упругости 130…140 ГПа)
* высокая износостойкость
* высокая усталостная прочность

Большинство классов композитов (но не все) обладают недостатками:

* Высокая стоимость
* Анизотропия свойств, то есть их зависимость свойств от выбора направления измерения. Например, модуль упругости однонаправленного углепластика вдоль волокон в 10-15 раз выше, чем в поперечном.
* Низкая ударная вязкость
* Высокий удельный объём
* Гигроскопичность
* Токсичность

Для исследовательской работы были предоставлены 2 файла: X\_bp.xlsx (с данными о параметрах базальтопластика, 1024 строки и 11 столбцов) и X\_nup.xlsx (данными нашивок углепластика, 1041 строки и 4 столбцов).

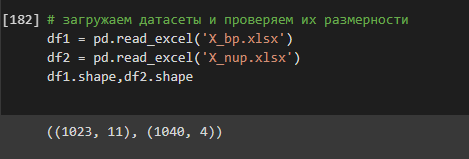
Базальт – природный материал вулканического происхождения – находит все большее применение в различных отраслях техники. Базальтовые волокна по всем основным характеристикам заметно превосходят волокна из традиционного алюмоборосиликатного стекла, приближаясь по своим показателям к более дорогим высокомодульным волокнам из магнийсиликатного стекла.

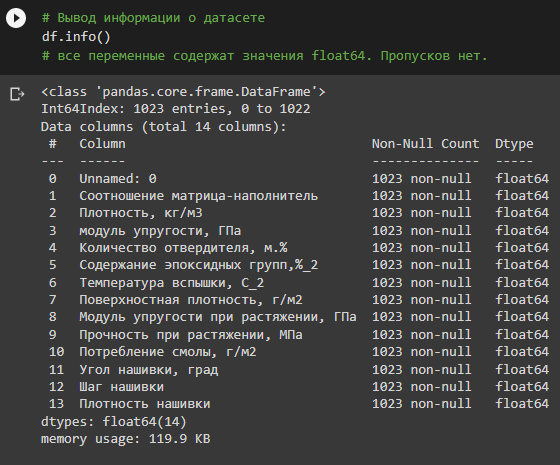
Базальтовые волокна обладают более высокой температуроустойчивостью и очень низкой гигроскопичностью. Гигроскопичность базальтовых волокон составляет менее 1%, а стеклянных – 10–20%. Прочность базальтовых волокон практически не изменяется после выдержки при 100%-ной относительной влажности в течение более 60 сут, в то время как прочность алюмоборосиликатных волокон уменьшается на 30%. Большим достоинством базальтовых волокон является кислото- и щелочеустойчивость. Процент сохранения прочности на разрыв базальтового волокна после выдержки в течение 3-х часов в кипящих щелочи и кислоте составляет, соответственно, 89 и 81%. На основе непрерывного базальтового волокна созданы ткани различных структур – ткани полотняного переплетения из ровинга, прошивная со стекловолокном из ровинга и некрученых нитей, сатинового переплетения из крученых нитей, трикотажное полотно на основе супертонкого волокна. Базальтопластики – полимерные композиционные материалы, применение в которых базальтового наполнителя по сравнению со стекловолокном позволит снизить стоимость композиционного материала, его гигроскопичность, повысить температуру эксплуатации.

Углеплатстики — полимерные композитные материалы из переплетённых нитей углеродного волокна, расположенных в матрице из полимерных (например, эпоксидных) смол. Плотность — от 1450 кг/м³ до 2000 кг/м³. Эти материалы отличаются высокой прочностью, жёсткостью и малой массой, часто прочнее стали, и гораздо легче. По удельным характеристикам превосходит высокопрочную сталь, например, легированную конструкционную сталь 25ХГСА.

Вследствие дороговизны, при экономии средств и отсутствии необходимости получения максимальных характеристик этот материал применяют в качестве усиливающих дополнений в основном материале конструкции.

Цель работы: разработать модель для прогноза модуля упругости при растяжении, прочности при растяжении и соотношения «матрица-наполнитель». Для этого нужно объединить 2 файла. Часть информации (17 строк таблицы способов компоновки композитов) не имеют соответствующих строк в таблице соотношений и свойств используемых компонентов композитов, поэтому были удалены.

Рисунок 1 — Размерность исходных датасетов

Рисунок 2 — информация об объединенном датасете

**1.2 Описание используемых методов**

Машинное обучение ([англ.](https://ru.wikipedia.org/wiki/Английский_язык) machine learning, ML) — класс методов [искусственного интеллекта](https://ru.wikipedia.org/wiki/Искусственный_интеллект), характерной чертой которых является не прямое решение задачи, а обучение за счёт применения решений множества сходных задач. Для построения таких методов используются средства [математической статистики](https://ru.wikipedia.org/wiki/Математическая_статистика), [численных методов](https://ru.wikipedia.org/wiki/Численные_методы), [математического анализа](https://ru.wikipedia.org/wiki/Математический_анализ), [методов оптимизации](https://ru.wikipedia.org/wiki/Оптимизация_(математика)), [теории вероятностей](https://ru.wikipedia.org/wiki/Теория_вероятностей), [теории графов](https://ru.wikipedia.org/wiki/Теория_графов), различные техники работы с [данными в цифровой форме](https://ru.wikipedia.org/wiki/Данные_(вычислительная_техника)).

Машинное обучение включает в себя следующие подходы:

● Обучение с учителем (supervised learning) - наиболее распространенный и изученный тип машинного обучения, ведь легче обучить машину с выбранными данными. В зависимости от того, что вы хотите предсказать, обучение с учителем может использоваться для решения задач классификации и регрессии.

● Обучение без учителя (unsupervised learning) — обучение на неразмеченных данных с целью обнаружения закономерностей в данных и их группировки. Обучение без учителя решает задач и кластеризации и уменьшения размерности.

● Обучение с частичным привлечением учителя (semi-supervised learning) - тип машинного обучения, требующий использования системы вознаграждения/штрафа.

● Обучение с подкреплением (reinforcement learning) - способ машинного обучения, при котором система обучается, взаимодействуя с некоторой средой.

Типы задач машинного обучения

Все задачи, решаемые с помощью ML, относятся к одной из следующих категорий.  
 Задача регрессии – прогноз на основе выборки объектов с различными признаками, к примеру цены квартиры, стоимости ценной бумаги через год и проч.

Задача классификации – получение категориального ответа на основе набора признаков. Имеет конечное количество ответов (как правило, в формате «да» или «нет»): есть ли на фотографии цветок, является ли изображение человеческим лицом.

Задача кластеризации – распределение данных на группы: разделение покупателей интернет-магазина по частоте покупок, отнесение автомобилей к определенному классу (легковой, грузовой, автобус и т.п.).

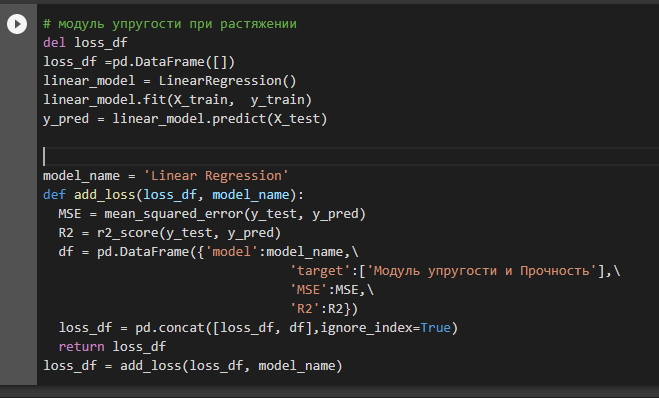
Задача уменьшения размерности – сведение большого числа признаков к меньшему (обычно 2–3) для удобства их визуализации.

Задача выявления аномалий – отделение аномалий от стандартных случаев. На первый взгляд она совпадает с задачей классификации, но есть существенное отличие: так как аномалии – редкое явление, обучающих примеров, мало или нет, поэтому методы классификации здесь не работают. Такой задачей является, например, выявление мошеннических действий с банковскими картами.

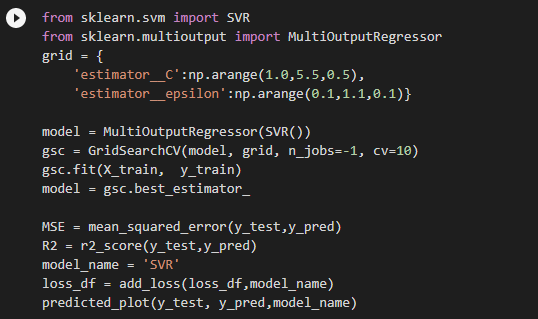
Поставленная задача в рамках приведенных категорий ML относится к задачам регрессии.

В работе были использованы следующие методы:

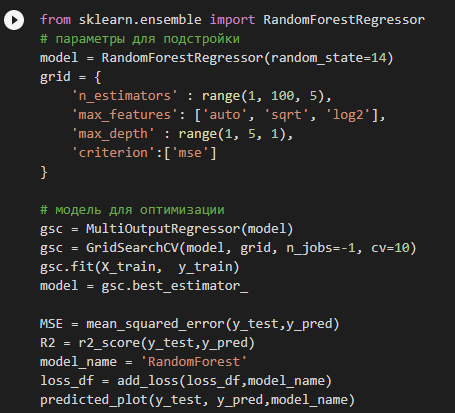
**Линейная регрессия (Linear regression)** — это алгоритм машинного обучения, основанный на контролируемом обучении, рассматривающий зависимость между одной входной и выходными переменными. Это один из самых простых и эффективных инструментов статистического моделирования. Она определяет зависимость переменных с помощью линии наилучшего соответствия. Главное достоинство метода: быстрота и простота в реализации и интерпретации. Но этот метод требует прямой связь между переменными.

Рисунок 3 — фрагмент кода, линейная регрессия

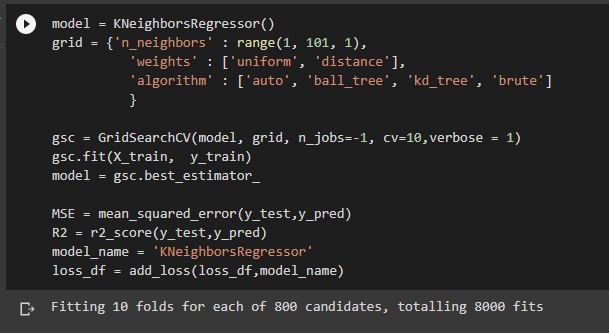
**Метод опорных векторов (Support Vector Regression, SVR)** – это контролируемое обучение моделей с использованием схожих алгоритмов для анализа данных и распознавания шаблонов. Каждый объект данных представляется как вектор (точка) в пространстве. Он создаёт линию или гиперплоскость, которая разделяет данные на классы. Метод хорошо работает на небольших датасетах, однако он неустойчив к шуму и сложнее в интерпретации.

Рисунок 4 — фрагмент кода, SVR

**Случайный лес (RandomForest)** — это универсальный алгоритм машинного обучения с учителем, представитель ансамблевых методов.  Если точность дерева решений оказалось недостаточной, можно множество моделей собрать в коллектив. Метод не подвержен переобучению, эффективен при работе с пропущенными данными, имеет высокую точность предсказания, но его применение требует много времени.

Рисунок 5 — фрагмент кода, RandomForest

**Метод К-ближайших соседей (kNN - k Nearest Neighbours)** — этот алгоритм находит расстояния между запросом и всеми примерами в данных, выбирая определенное количество примеров (k), наиболее близких к запросу, затем голосует за наиболее часто встречающуюся метку (в случае задачи классификации) или усредняет метки (в случае задачи регрессии). Метод универсален, прост в реализации и понимании полученных результатов и не чувствителен к выбросам. При этом он трудоемок и имеет высокую зависимость результатов от выбранной метрики.

Рисунок 6 — фрагмент кода, метод k-ближайших соседей

В качестве метрики качества модели использовалась **MSE (Mean Squared Error)** или средняя квадратичная ошибка. Она 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**

Разведочный анализ (Exploratory Data Analysis) — это процесс анализа данных, способствующий выдвижению гипотез, когда отсутсвтует представление о связи между переменными или оно недостаточно. Важно отметить, что разведочный анализ— это не набор готовых методов анализа, а философия того, как следует анализировать данные, что искать и как интерпретировать.

В разведочном анализе часто используется статистическая графика — графическая техника, используемая для визуализации количественных данных.

Результаты разведочного анализа не используются для выработки управленческих решений. Их назначение — помощь в разработке наилучшей стратегии углубленного анализа, выдвижение гипотез, уточнение особенностей применения тех или иных математических методов и моделей. Без разведочного анализа углубленный анализ данных будет производиться практически «вслепую».

Это необходимо для того, чтобы:

* максимизировать понимание набора данных;
* выявить базовую структуру;
* извлечь важные переменные;
* выявить выбросы и аномалии;
* проверить исходные предположения;
* разработать скупые модели;
* определять оптимальные параметры.

Причина сильной зависимости в графике заключается в том, что основная роль разведочного анализа — открывать глаза, а графика дает аналитикам возможность заставлять данные раскрывать свои структурные секреты и быть всегда готовыми получить новое, часто непредвиденное, понимание данных. В сочетании с естественными возможностями распознавания образов, которыми все мы обладаем, графика открывает большие шансы для этого.

Можно назвать следующие распространенные методы использования графики в разведочном анализе:

* предоставление необработанных данных с помощью гистограмм, диаграмм рассеяния, вероятностных графиков;
* вычисление простых статистических данных: добавление линии среднего значения и стандартного отклонения; сопоставление нескольких графиков.

К основным методам разведочного анализа относятся:

* процедура анализа распределений переменных;
* [корреляционный анализ](https://wiki.loginom.ru/articles/correlation-analysis.html) c целью поиска [коэффициентов](https://wiki.loginom.ru/articles/correlation-coefficient.html), превосходящих по величине определенные пороговые значения;
* [факторный анализ](https://wiki.loginom.ru/articles/factorial-analysis.html);
* [дискриминантный анализ](https://wiki.loginom.ru/articles/linear-discriminant-analysis.html);
* многомерное шкалирование;
* визуальный анализ.

Корреляционный анализ (Correlation analysis) — это совокупность методов обработки данных с целью обнаружения статистической взаимосвязи между случайными величинами или [признаками](https://wiki.loginom.ru/articles/attribute.html). Впервые элементы корреляционного анализа в научных исследованиях начал применять французский палеонтолог [Жорж Кювье](https://ru.wikipedia.org/wiki/Кювье,_Жорж_Леопольд), который и ввел в научный обиход термин «корреляция». Значительный вклад в развитие теории корреляционного анализ внесли [Френсис Гальтон](https://ru.wikipedia.org/wiki/Гальтон,_Фрэнсис), [Карл Пирсон](https://ru.wikipedia.org/wiki/Пирсон,_Карл), [Чарльз Спирмен](https://ru.wikipedia.org/wiki/Спирмен,_Чарльз_Эдвард), [Морис Кендалл](https://ru.wikipedia.org/wiki/Кендалл,_Морис_Джордж) и другие.

Корреляционный анализ для двух случайных величин заключает в себе:

* построение корреляционного поля и составление корреляционной таблицы;
* вычисление выборочных [коэффициентов корреляции](https://wiki.loginom.ru/articles/correlation-coefficient.html) и корреляционных отношений;
* проверка статистической [гипотезы](https://wiki.loginom.ru/articles/hypothesis.html) значимости связи.

Методами корреляционного анализа решаются следующие задачи:

* Взаимосвязь. Установление наличия зависимости между двумя признаками и определение ее силы.
* Прогнозирование. Предсказание поведения одного признака на основе изменения другого, коррелирующего с первым.
* Отбор переменных. Корреляционный анализ позволяет производить выбор набора входных переменных для аналитической модели, в наименьшей степени коррелирующих между собой и в наибольшей степени коррелирующих с [выходной переменной](https://wiki.loginom.ru/articles/output-variable.html). Это позволяет сделать работу аналитических моделей более точной и устойчивой.

Сила корреляционной связи между двумя переменными характеризуется с помощью коэффициента корреляции.

Конкретный вид (линейная, экспоненциальная и т.д.) корреляционной связи и ее параметры устанавливается в рамках [регрессионного анализа](https://wiki.loginom.ru/articles/regression-analysis.html).

Факторный анализ (Factorial analysis) — направление [математической статистики](https://wiki.loginom.ru/articles/mathematical-statistics.html), раздел многомерного статистического анализа, объединяющий вычислительные методы, которые в ряде случаев позволяют получить компактное описание исследуемых явлений на основе обработки больших массивов информации. Основные идеи факторного анализа были заложены английским психологом и антропологом [Фрэнсисом Гальтоном](https://ru.wikipedia.org/wiki/Гальтон,_Фрэнсис) во второй половине XIX в.

От других средств подобного «сжатия информации» (например, распространенных методов статистической группировки объектов) факторный анализ отличается тем, что не опирается на заранее заданный, априорный перечень факторов, влияющих на исследуемые процессы или объекты, а, наоборот, при соблюдении определенных правил и предосторожностей помогает обнаружить наиболее важные из них, причем скрытые. Задача факторного анализа состоит в том, чтобы выявить скрытые обобщенные факторы, которые в достаточной для данного исследования степени объясняют изменения изучаемого показателя. Выявленные факторы позволяют строить аналитические модели с относительно небольшим числом [независимых переменных](https://wiki.loginom.ru/articles/input-variable.html), что упрощает их реализацию и интерпретацию пользователем, снижает вычислительные затраты и время, требуемое на получение решений, а следовательно, повышает оперативность принятия решений на основе результатов анализа.

Дискриминантный линейный анализ (Linear discriminant analysis) и линейный дискриминант Фишера — методы статистики и [машинного обучения](https://wiki.loginom.ru/articles/machine-learning.html) для нахождения линейных комбинаций признаков, наилучшим образом разделяющих два или более класса объектов или событий. Полученная комбинация может быть использована в качестве линейного классификатора или для сокращения размерности пространства признаков перед последующей к[лассификацией](https://wiki.loginom.ru/articles/classification.html). Основные идеи дискриминантного анализа были сформулированы [Роналдом Фишером](https://ru.wikipedia.org/wiki/Фишер,_Роналд) в 1936 г.

Дискриминантный линейный анализ представляет собой раздел многомерного статистического анализа, содержанием которого является разработка методов решения задач различения (дискриминации) объектов наблюдения по набору [признаков](https://wiki.loginom.ru/articles/attribute.html). Иными словами, он позволяет изучать различия между двумя и более группами объектов по нескольким признакам одновременно.

Дискриминантный линейный анализ тесно связан с [дисперсионным анализом](https://wiki.loginom.ru/articles/anova.html) и [регрессионным анализом](https://wiki.loginom.ru/articles/regression-analysis.html), также пытающимися выразить какую-либо зависимую переменную через линейную комбинацию других признаков или измерений. Но этих двух методах [зависимая переменная](https://wiki.loginom.ru/articles/output-variable.html) — численная величина, а в дискриминантном линейном анализе она является величиной номинальной ([меткой класса](https://wiki.loginom.ru/articles/class-label.html)). Также дискриминантный линейный анализ имеет схожие черты с [методом главных компонент](https://wiki.loginom.ru/articles/principal-component-analysis.html) и [факторным анализом](https://wiki.loginom.ru/articles/factorial-analysis.html), которые ищут линейные комбинации величин, наилучшим образом описывающие данные.

Можно выделить три вида задач дискриминантного анализа:

* определение дискриминирующих признаков (т.е. признаков, которые позволяют отнести наблюдение к той или иной группе);
* построение дискриминирующей функции;
* [прогнозирование](https://wiki.loginom.ru/articles/forecasting.html) будущих событий, связанных с попаданием объекта в ту или иную группу на основе значений его признака.

Основной целью дискриминации является поиск линейной комбинации признаков (называемых дискриминантными признаками), которые позволили бы наилучшим образом разделить рассматриваемые группы.

Дискриминантный линейный анализ широко используется для решения задач классификации и распознавания образов, понижения размерности входных данных. Хотя он и работает с информацией, которая определяет принадлежность объекта к одному из классов, но сам по себе классификатором не является, а используется как часть линейной классификационной модели.

Преимущество метода — сравнительная простота реализации и интерпретации результатов. Недостаток — чувствительность к распределению исходных данных, когда даже небольшое их изменение приводит к значительным изменениям результатов классификации.

В заключение отметим, что предварительное исследование данных может служить лишь первым этапом в процессе их анализа, и пока результаты не подтверждены на других [выборках](https://wiki.loginom.ru/articles/sample.html) или на независимом множестве данных, их следует воспринимать самое большее как гипотезу. Если результаты разведочного анализа говорят в пользу некоторой модели, то ее правильность можно затем проверить, применив к новым данным.

# **2. Практическая часть**

**2.1 Предобработка данных**

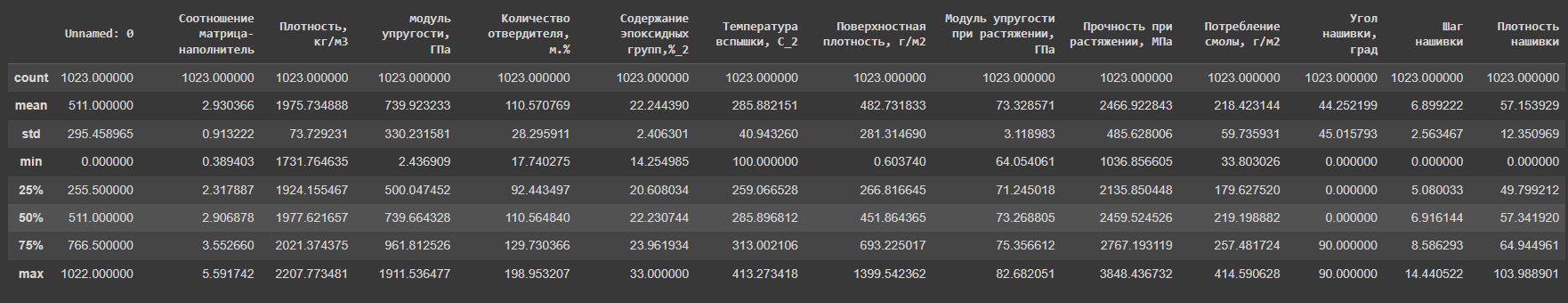
Прежде чем передать данные в работу моделей машинного обучения, необходимо обработать и очистить их. Но безосновательно удалять что-либо тоже неправильно. Именно поэтому сначала набор данных надо изучить.

Приступим к выполнению задачи по прогнозированию конечных свойств новых композиционных материалов.

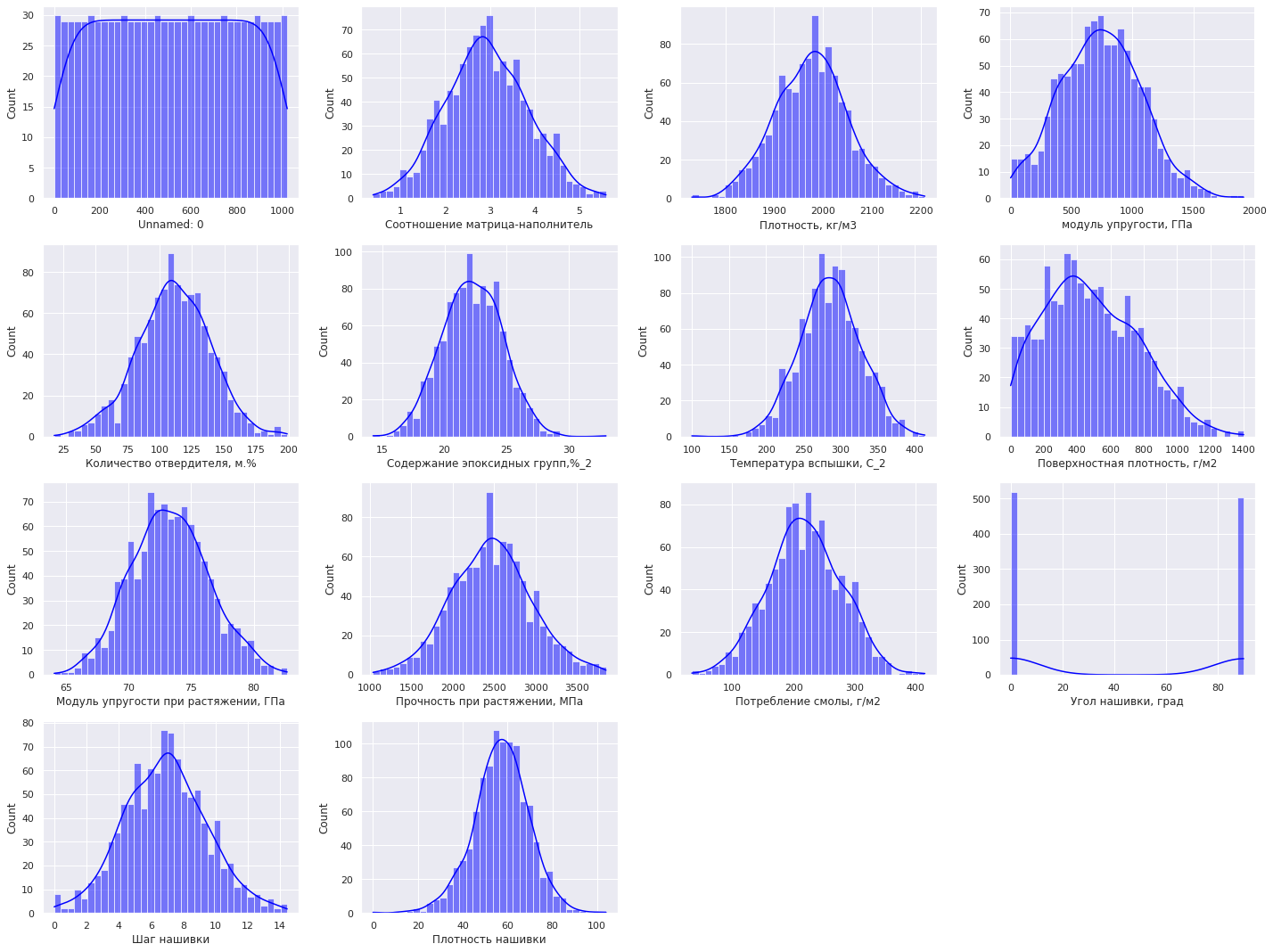
**На входе** имеются данные о начальных свойствах компонентов композиционных материалов (количество связующего, наполнителя, температурный режим отверждения и т.д.). На выходе необходимо спрогнозировать ряд конечных свойств получаемых композиционных материалов. Кейс основан на реальных производственных задачах Центра НТИ «Цифровое материаловедение: новые материалы и вещества» (структурное подразделение МГТУ им. Н.Э. Баумана).

В качестве инструментов разведочного анализа будем использовать: оценку статистических характеристик датасета; гистограммы распределения переменных; диаграммы ящика с усами; попарные графики рассеяния точек (несколько вариантов); тепловую карту

Данные описательной статистики объединенного дататсета до начала обработки выглядят как показано на рисунке 7.

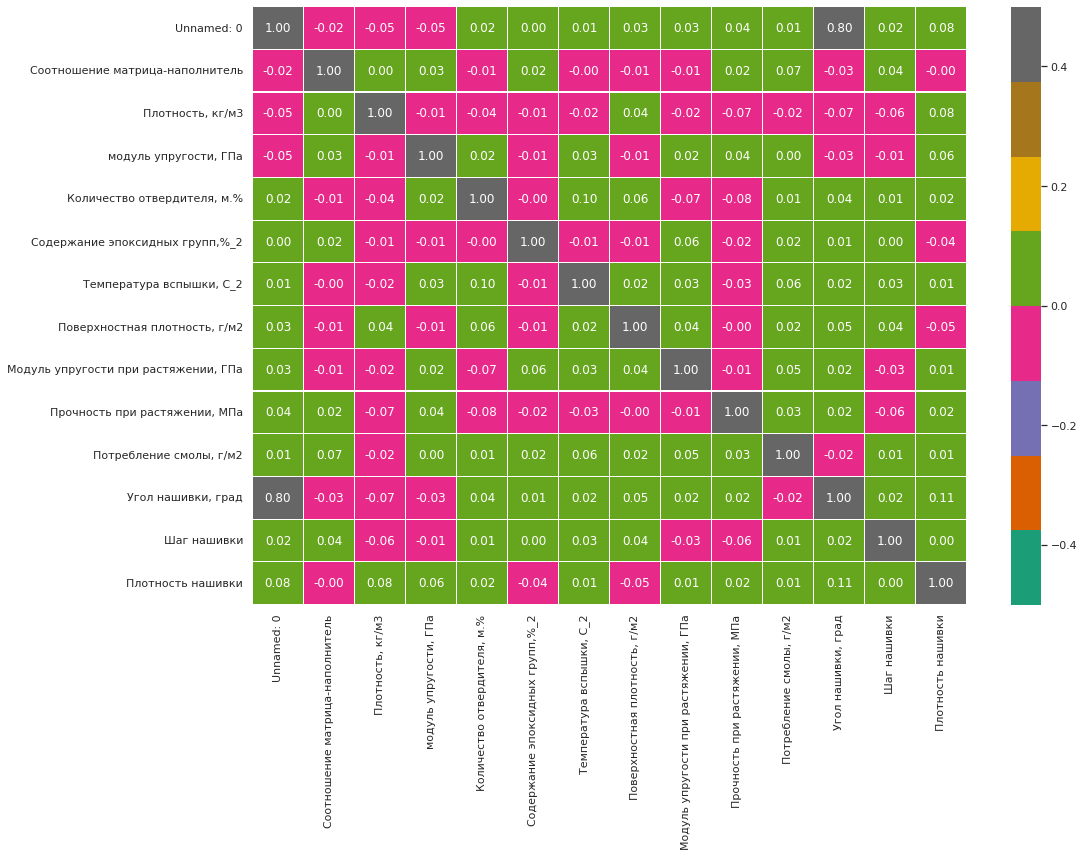
Рисунок 7 — Описательная статистика объединенного датасета

Для изучения распределений частот значений переменных используем гистограммы. Видно очень слабую корреляцию между переменными.

Рисунок 8 - Гистограммы

По представленным гистограммам видно, что параметры имеют весьма большой разброс.

У целевых признаков нет сильной корреляции ни с одним из исходных признаков. Наиболее наглядно это можно увидеть на следующем графике тепловой карте.

Рисунок 9 — Тепловая карта

Визуализируем выбросы для различных признаков с помощью диаграмм «ящики с усами» (boxplot). Она показывает медиану, 25 и 75 квантили, верхнюю и нижнюю границы, а также выбросы (точки вне прямоугольника и "усов").

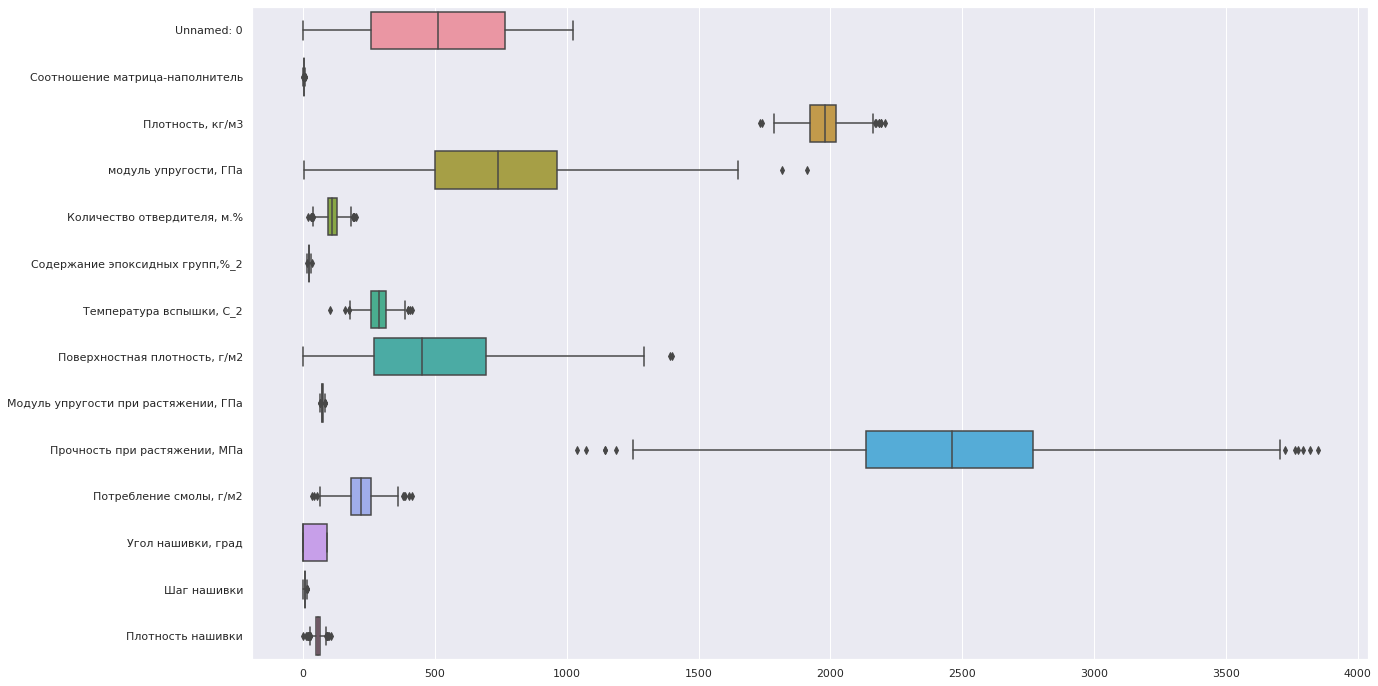
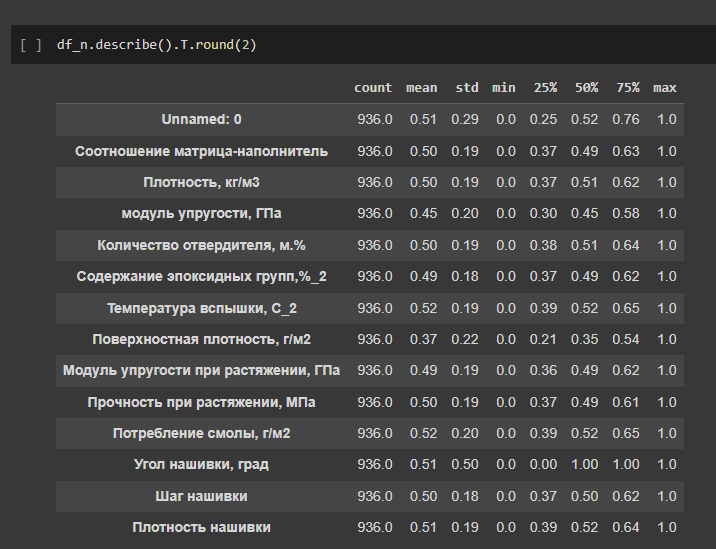
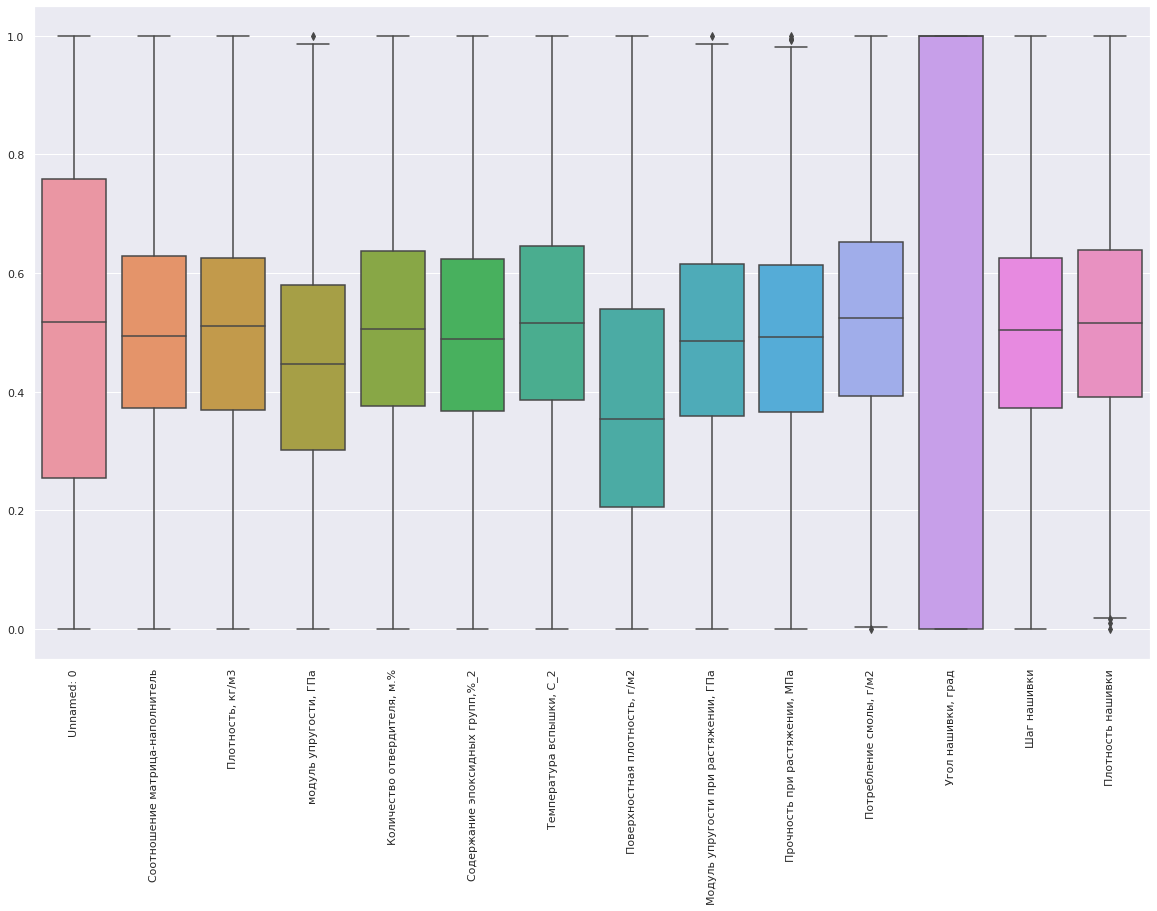


Рисунок 10 - «Ящики с усами» до нормализации данных

После удаления выбросов Размерность датасета составит 936 строк на 14 столбцов.

Применим MinMaxScaler(), чтобы нормализовать датасет. Описательная статистика после этого выглядит как показано на рисунке 11. Визуализируем новые данные в виде диаграмм «ящики с усами».

Рисунок 11 — Описательная статистика после нормализации

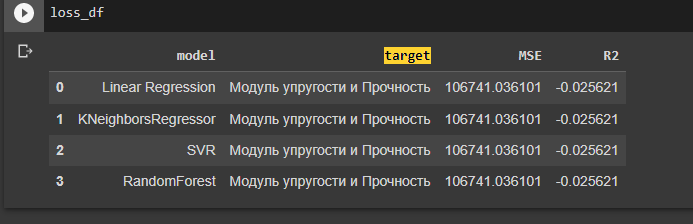
Рисунок 12 - Boxplot после нормализации

**2.2 Разработка и обучение модели**

При разработке и обучении модели использовались следующие методы:

* Метод опорных векторов (Support Vector Machines — SVM);
* Линейная регрессия;
* RandomForestRegressor;
* Метод K-ближайших соседей.

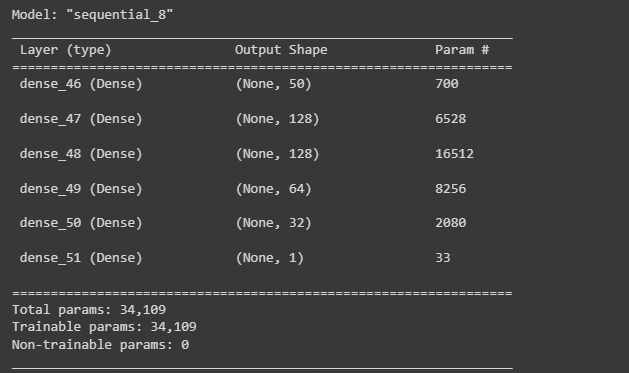
Ни один из методов не дал приемлемого результата.

Рисунок 13 — Результаты оценки точности модели

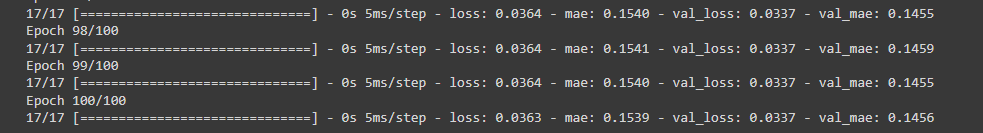
**2.4 Написание нейронной сети, которая будет рекомендовать**

**соотношение «матрица – наполнитель»**

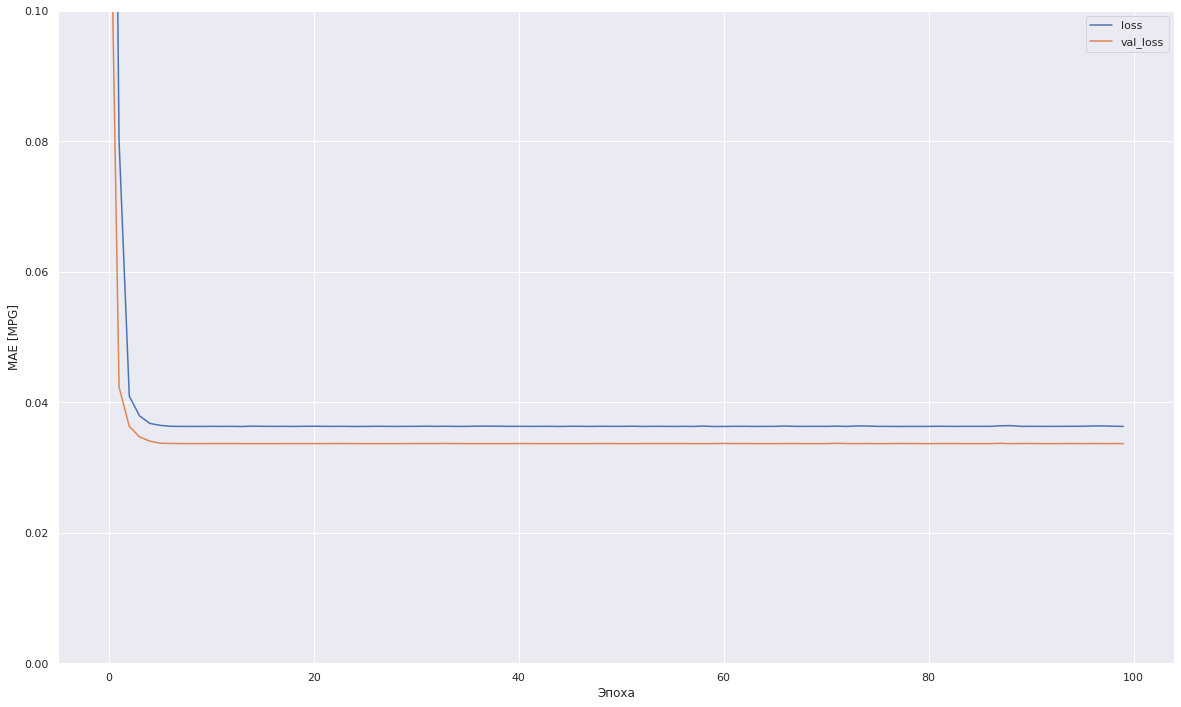
Обучение нейронной сети — это такой процесс, при котором происходит подбор оптимальных параметров модели, с точки зрения минимизации функционала ошибки.

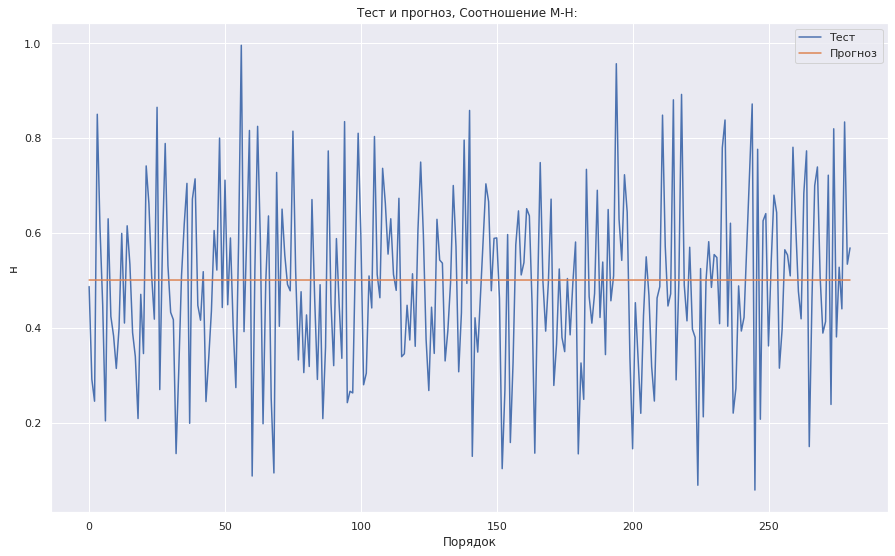
Рисунок 14 — Построение нейронной сети

Проведем обучение нейронной сети в течение 100 эпох, на последних эпохах получим данные:

Рисунок 15 — Обучение нейросети, фрагмент

Сформируем график потерь модели, а также зададим функцию для визуализации тестовых и прогнозных значений для результатов модели.

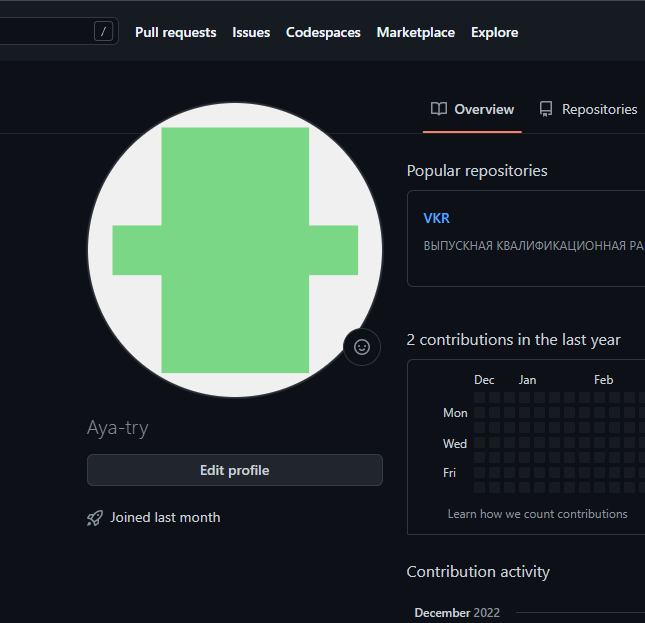
Рисунок 16 — График потерь модели

Рисунок 17 — Тестовые и прогнозные значения модели

**2.5 Создание удаленного репозитория и загрузка проекта**

Репозиторий был создан на github.com по адресу: https://github.com/Aya-try/VKR/

В репозиторий загружены файлы с ВКР.

Рисунок 18 — Фрагмент профиля на github.com

# **Заключение**

При решении данной задачи не удалось достигнуть желаемых результатов, так как признаки плохо между собой коррелируют ни один из признаков не приближается близко к параметрам упругости и прочности.

Можно сделать вывод, что невозможно определить из свойств материалов соотношение «матрица – наполнитель». Однако это не указывает на невозможность прогнозирования характеристик композитных материалов, но может указывать на недостатки базы данных и подходов, использованных в данной работе.

В целом прогнозирование конечных свойств и характеристик композитных материалов без изучения материаловедения, погружения в вопрос экспериментального анализа характеристик композитных материалов не способно дать удовлетворительные результаты.

# Список литературы

1. Devpractice Team. Python. Визуализация данных. Matplotlib. Seaborn. Mayavi. - devpractice.ru. 2020. - 412 с.: ил.
2. Абу-Хасан Махмуд, Масленникова Л. Л.: Прогнозирование свойств композиционных материалов с учётом наноразмера частиц и акцепторных свойств катионов твёрдых фаз, статья, 2006 г.
3. Бизли Д. Python. Подробный справочник: учебное пособие. – Пер. с англ. – СПб.: Символ-Плюс, 2010. – 864 с., ил.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: 12.12.2022)
5. Документация по библиотеке matplotlib: – Режим доступа: <https://matplotlib.org/stable/users/index.html>. (дата обращения: 12.12.2022)
6. Документация по библиотеке numpy: – Режим доступа: [https://numpy.org/doc/1.22/user/index.html#user](_blank). (дата обращения: 12.12.2022)
7. Документация по библиотеке pandas: – Режим доступа: [https://pandas.pydata.org/docs/user\_guide/index.html#user-guide](_blank). (дата обращения: 12.12.2022)
8. Документация по библиотеке scikit-learn: – Режим доступа: <https://scikit-learn.org/stable/user_guide.html>. (дата обращения: 12.12.2022)
9. Документация по библиотеке Tensorflow: – Режим доступа: <https://www.tensorflow.org/overview> (дата обращения: 12.12.2022)
10. Документация по языку программирования python: – Режим доступа:  <https://docs.python.org/3.8/index.html>. (дата обращения: 12.12.2022)
11. Иванов Д.А., Ситников А.И., Шляпин С.Д – Композиционные материалы: учебное пособие для вузов, 2019. 13 с.
12. Композиционные материалы волокнистого строения.К., 1980.
13. Конкин А.А., Углеродные и другие жаростойкие волокнистые материалы, М., 1974.
14. Композиционные материалы, пер. с англ., т.1-8, М., 1998.
15. Наполнители для полимерных композиционных материалов, пер. с англ., М., 2005.
16. 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.
17. 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>. (дата обращения 12.12.2022)