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 Глумова Влада Владиславовна

Москва, 2022

Содержание

[Введение 3](#_Toc101111306)

[1. Аналитическая часть 7](#_Toc101111307)

[1.1. Постановка задачи. 7](#_Toc101111308)

[1.2. Описание используемых методов 9](#_Toc101111309)

[1.3. Разведочный анализ данных 16](#_Toc101111310)

[2. Практическая часть 17](#_Toc101111311)

[2.1. Предобработка данных 17](#_Toc101111312)

[2.2. Разработка и обучение модели 40](#_Toc101111313)

[2.3. Тестирование модели 41](#_Toc101111314)

[2.4. Написать нейронную сеть, которая будет рекомендовать соотношение матрица-наполнитель. 44](#_Toc101111315)

[2.5. Разработка приложения 48](#_Toc101111316)

[2.6. Создание удаленного репозитория и загрузка результатов работы на него. 49](#_Toc101111317)

[Заключение 50](#_Toc101111318)

[Библиографический список 51](#_Toc101111319)

# Введение

Выпускная квалификационная работа разработана и зафиксирована в данном документе слушателем курса «Data Science» Глумовой Владой Владиславовной. В ходе работы слушателем выполнен ряд прогнозов с использованием библиотек среды Phyton, проведена работа по разработке программы для получения прогнозов посредством построенных моделей, изучена специфика производства композитных материалов.

Данная работа не может быть использована как конечный продукт без согласия Глумовой Влады Владиславовны, все представленные далее материалы носят исключительно ознакомительный характер, являются учебной работой студента, не претендуют на научную ценность. Все материалы, использованные при подготовке работы были получены из открытых источников, в том числе после приобретения материала в личное пользование студента.

В целях формирования понимания по проблематике вопроса были изучены требования к ВКР, где были изложены общие сведения по рассматриваемой задаче, а также внешние источники, раскрывающие данный вопрос с иными подходами.

В соответствии с изданием "Проектирование и прогнозирование механических свойств однонаправленного слоя из композиционного материала", композиционными материалами (КМ) называют материалы, состоящие из двух и более нерастворимых друг в друге компонентов с четко обозначенной границей раздела и прочным взаимодействием по всей площади контакта. Один из компонентов КМ является непрерывной фазой и называется матрицей, в которой размещены нерастворимые в ней материалы другой природы, называемые армирующими или наполнителями.

Композиционные материалы обладают комплексом свойств, которыми обладают компоненты, и свойствами, которыми отдельные компоненты не обладают.

По типу матрицы композиционные материалы подразделяются на отверждающие (реактопласты), термопластичные композиционные материалы (ТКМ), композиционные материалы на металлической матрице.

По природе наполнителей композиционные материалы подразделяются на минеральнонаполненные (стекло, базальт), углероднонаполненные (углеродные волокна), органонаполненные (органические волокна), металлонаполненные. По форме частиц наполнителя различают композиционные материалы с дисперсными наполнителями (порошки, микросферы, чешуйки), композиционные материалы, наполненные короткими волокнами, и КМ с непрерывными волокнами.

По распределению наполнителей в матрице различают изотропные и анизотропные композиционные материалы. К первым относятся КМ с дисперсными наполнителями и наполненные короткими волокнами. Ко вторым - КМ с непрерывными волокнами.

Отдельную классификационную группу образуют гибридные КМ, армированные системами волокон различной природы (стеклянные, органические, углеродные и т.д.).

Армирующие компоненты композиционных материалов в виде волокон, нитей, тканей и др. сами по себе не являются конструкционными материалами. Для того чтобы получить армированный пластик, необходимо соединить волокна друг с другом, обеспечить передачу усилий между ними, зафиксировать форму изделий и взаимное расположение армирующей компоненты, т. е. создать монолитную структуру композита. Для этого используются различные связующие. В качестве связующих применяются полимерные, металлические и керамические материалы.

После формования изделия из композиционных материалов отвержденное связующее называется матрицей. Часто эти понятия являются синонимами.

Пропитанный связующим армирующий каркас при формировании обычно подвергается действию значительного давления. После получения требуемой степени пропитки полимерное связующее отверждается. Отверждение происходит при нормальной или повышенной температуре в одну или несколько стадий в зависимости от вида армирующего материала, типа связующего, способа изготовления композита, необходимой скорости отверждения, степени отверждения и требований к свойствам конечного продукта. Отверждение при высокой температуре увеличивает термостойкость матрицы и композита, однако при этом возрастает их хрупкость.

Необходимо указать, что используемые в рамках рассматриваемого данной работой Dataset "нашивки" имеют отношение к "наполнителям", используемым в композитных материалах. Так, например, при разработке технологии создания элементов газотурбинного двигателя из полимерных композиционных материалов применяется автоматизированная нашивка ровингом, представленная волокнистыми наполнителями: углеродными волокнами HTS 45 12R и IMS 65 24K (TohoTenax, Япония), и армидным волокном Русар-С (ООО НПП "Термотекс", г. Мытищи, Россия).

Приведенная информация является минимально необходимым контекстом для формирования понимания поставленной перед студентами задачи.

В условиях расширения разнообразия материалов, используемых при проектировании нового композита возрастает потребность в определении характеристики нового композита с наименьшими финансовыми затратами. Для решения данной задачи как правило используют два пути: физические испытания образцов материалов, или прогнозирование характеристик, в том числе на основе физико-математических моделей. Суть прогнозирования заключается в симуляции представительного элемента объема композита, на основе данных о характеристиках входящих компонентов (связующего и армирующего компонента). В рамках текущей работы мы попробуем спроектировать модель, способную с высокой долей вероятности прогнозировать модули упругости при растяжении и прочности при растяжении, а также напишем нейронную сеть, которая будет рекомендовать соотношение матрица-наполнитель. Но необходимо отметить, что входящие данные фактически не разделяют связующего, матрицы и наполнителя, давая данные напрямую из эксперимента, что может существенно сократить возможность достоверного прогнозирования конечных характеристик продукта.

Теоретическая часть позволяет заключить, что исходный набор данных должен иметь более низкоуровневую структуру, спустить нас до исходных составляющих формирования будущего КМ.

# 1. Аналитическая часть

# 1.1. Постановка задачи.

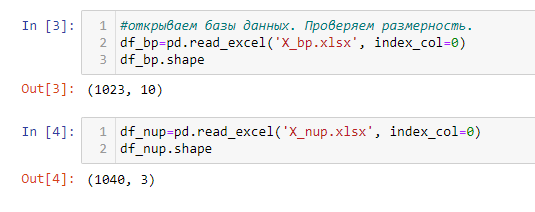
Имеются данные о начальных свойствах компонентов композиционных материалов (количество связующего, наполнителя, температурный режим отверждения и т.д.). На выходе необходимо спрогнозировать ряд конечных свойств получаемых композиционных материалов. Кейс основан на реальных производственных задачах Центра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

Датасет со свойствами композитов требуется объединить по индексу, тип объединения INNER. Датасет:

<https://drive.google.com/file/d/1B1s5gBlvgU81H9GGolLQVw_SOi-vyNf2/view?usp=sharing>

Анализ представленного Датасета демонстрирует следующее противоречие:

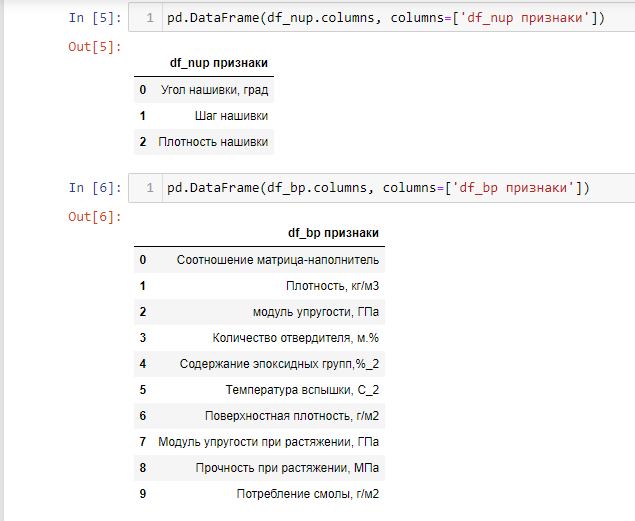
Рис. 1



Представленное изображение демонстрирует конфликт в данных, используемых для построения прогноза. Размерность большего документа 1023 строки при 10 столбцах, меньший документ имеет больше экспериментов 1040 при 3-х характеристиках.

Основные элементы, представленные в Датасете:

Рис. 2



При слиянии двух файлов получен датасет размерностью 1023 на 13.   
Со следующими характеристиками:

Таблица 1

|  | **mean** | **std** | **min** | **max** |
| --- | --- | --- | --- | --- |
| 1 | 2 | 3 | 4 | 5 |
| Соотношение матрица-наполнитель | 2.93 | 0.91 | 0.389 | 5.59 |
| Плотность, кг/м3 | 1975.73 | 73.72 | 1731.76 | 2207.77 |
| модуль упругости, ГПа | 739.92 | 330.23 | 2.436 | 1911.53 |
| Количество отвердителя, м.% | 110.57 | 28.29 | 17.74 | 198.95 |
| Содержание эпоксидных групп,%\_2 | 22.24 | 2.406 | 14.25 | 33.0 |
| Температура вспышки, С\_2 | 285.88 | 40.94 | 100 | 413.27 |
| Поверхностная плотность, г/м2 | 482.73 | 281.31 | 0.60 | 1399.54 |
| Модуль упругости при растяжении, ГПа | 73.32 | 3.11 | 64.05 | 82.68 |
| Прочность при растяжении, МПа | 2466.92 | 485.62 | 1036.85 | 3848.43 |
| Потребление смолы, г/м2 | 218.42 | 59.73 | 33.80 | 414.59 |
| Угол нашивки, град | 44.25 | 45.01 | 0 | 90 |
| Шаг нашивки | 6.89 | 2.56 | 0 | 14.44 |
| Плотность нашивки | 57.15 | 12.35 | 0 | 103.98 |

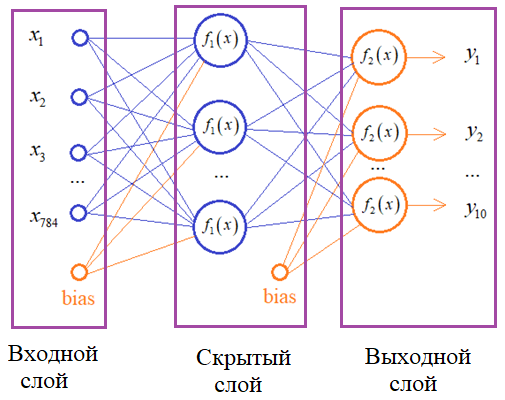
Данный датасет требует предобработки, что будет реализовано далее.

# 1.2. Описание используемых методов

В рамках решения задач выпускной квалификационной работы используются библиотеки numpy, pandas, matplotlib.pyplot, seaborn, sklearn, tensorflow, в частности keras. В качестве моделей для построения прогнозов используются нейронные сети Sequential и MLPRegressor, а также DummyRegressor, LinearRegression, Ridge, Lasso, SVR, SGDRegressor, KNeighborsRegressor, DecisionTreeRegressor, RandomForestRegressor, GradientBoostingRegressor.

Нейронные сети, используемые в рамках работы принципиально имеют схожую архитектуру и базируются на следующей схеме:

Рис. 3



В рамках прогноза, строящегося в текущей работе выходной слой всегда будет содержать один нейрон, на вход подается 12 признаков. Внутреннее содержание слове при прогнозе используется следующее:

Input(shape=(12,)), # Входной слой

keras.layers.Dense(units=48, activation='relu'), # Скрытый слой

keras.layers.Dense(units=24, activation='relu'), # Скрытый слой

keras.layers.Dense(units=12, activation='relu'), # Скрытый слой

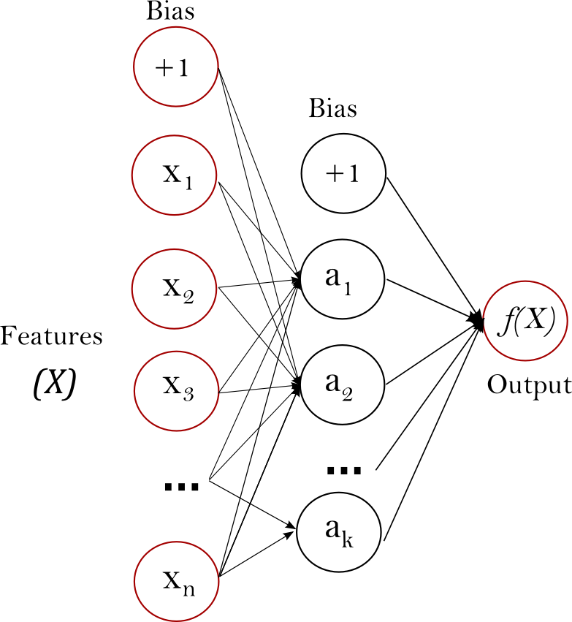
keras.layers.Dense(units=6, activation='relu')

Для MLPRegressor использованы эти же показатели, но в синтаксисе sklearn.neural\_network. Число нейронов на каждом скрытом слое подобрано с наибольшей эффективностью текущих моделей.

MLPRegressor – многослойный персептрон.

**Многослойный персептрон (MLP)** — это алгоритм обучения с учителем, который изучает функцию f(⋅):Rm→Ro обучением на наборе данных, где m — количество измерений для ввода и o- количество размеров для вывода. Учитывая набор функций X=x1,x2,…,xm и цель y, он может изучить аппроксиматор нелинейной функции для классификации или регрессии. Он отличается от логистической регрессии тем, что между входным и выходным слоями может быть один или несколько нелинейных слоев, называемых скрытыми слоями.

Рис. 4



Самый левый слой, известный как входной, состоит из набора нейронов xi|x1,x2,…,xm представляющие входные функции. Каждый нейрон в скрытом слое преобразует значения из предыдущего слоя с взвешенным линейным суммированием w1x1+w2x2+…+wmxm, за которой следует нелинейная функция активации g(⋅):R→R — как функция гиперболического загара. Выходной слой получает значения из последнего скрытого слоя и преобразует их в выходные значения.

Модуль содержит публичные атрибуты coefs\_ и intercepts\_. coefs\_список весовых матриц, где весовая матрица с индексомi представляет собой веса между слоями i и слой i+1. intercepts\_ список векторов смещения, где вектор с индексом i представляет значения смещения, добавленные к слою i+1.

Преимущества многослойного перцептрона:

* Возможность изучать нелинейные модели.
* Возможность изучения моделей в режиме реального времени (онлайн-обучение) с использованием partial\_fit.

К недостаткам многослойного персептрона (MLP) можно отнести:

* MLP со скрытыми слоями имеют невыпуклую функцию потерь, когда существует более одного локального минимума. Поэтому разные инициализации случайных весов могут привести к разной точности проверки.
* MLP требует настройки ряда гиперпараметров, таких как количество скрытых нейронов, слоев и итераций.
* MLP чувствителен к масштабированию функций.

Sequential – последовательная модель нейронного обучения библиотеки Keras. Модель Sequential представляет собой линейный стек слоев. Фактически является удобной оболочкой для построения моделей любой сложности с использованием инструментария библитеки Keras. Sequential API хорош для создания моделей глубокого обучения в большинстве случаев, но есть некоторые ограничения. Например, в нем непросто задать модели с множеством входов или выходов, а также переиспользующие свои слои. Выше описаны слои, использованные в работе, в нашем случае не требуется прогнозировать несколько выходов, так что мы можем использовать модель Sequential API. Далее кратко о ее компонентах.

Активационная функция Relu на момент написания работы является наиболее распространенным выбором функции активации в мире глубокого обучения. Она имеет линейную составляющую в своем содержании, легче подстраивается под внешне разрозненные данные. В рамках текущей работы были перебраны разные линейные и нелинейные активационные функции для построения прогноза, но только Relu смогла выдать результат приближенный к требуемому.

В качестве оптимизатора выбран Adam - это метод стохастического градиентного спуска, основанный на адаптивной оценке моментов первого и второго порядка.

Согласно Kingma et al., 2014, метод "вычислительно эффективен, имеет мало требований к памяти, инвариантен к диагональному масштабированию градиентов и хорошо подходит для задач, которые велики с точки зрения данных/параметров". Удобный, надежный оптимизатор для нейронной сети, показывает хорошие результаты с нейронными сетями Keras.

DummyRegressor – он же «фиктивный регрессор». Фиктивный регрессор - это регрессор, который дает прогноз на основе простых стратегий, не обращая внимания на входные данные. Как и в случае с фиктивным классификатором, библиотека sklearn также предоставляет фиктивный регрессор, который используется для настройки базовой линии для сравнения других существующих регрессоров, а именно пуассоновского регрессора, линейной регрессии, гребневой регрессии и многих других. В нашей работе он будет использоваться как база для оценки эффективности других моделей.

LinearRegression — используемая в статистике регрессионная модель зависимости одной (объясняемой, зависимой) переменной y от другой или нескольких других переменных (факторов, регрессоров, независимых переменных) x с линейной функцией зависимости. Линейная регрессия относится к задаче определения «линии наилучшего соответствия» через набор точек данных и стала простым предшественником нелинейных методов, которые используют для обучения нейронных сетей. В этой статье покажем вам примеры линейной регрессии.

Цель линейной регрессии — поиск линии, которая наилучшим образом соответствует этим точкам. Напомним, что общее уравнение для прямой есть f (x) = m⋅x + b, где m — наклон линии, а b — его y-сдвиг. Таким образом, решение линейной регрессии определяет значения для m и b, так что f (x) приближается как можно ближе к y. Линейная регрессия обычно является первым алгоритмом машинного обучения, с которым сталкивается каждый ученый. Ее мы будем использовать для проверки наличия линейных зависимостей.

Ridge, она же гребневая регрессия, и Lasso - это два простых метода, позволяющих уменьшить сложность модели и предотвратить переопределение, которое может возникнуть в результате простой линейной регрессии.

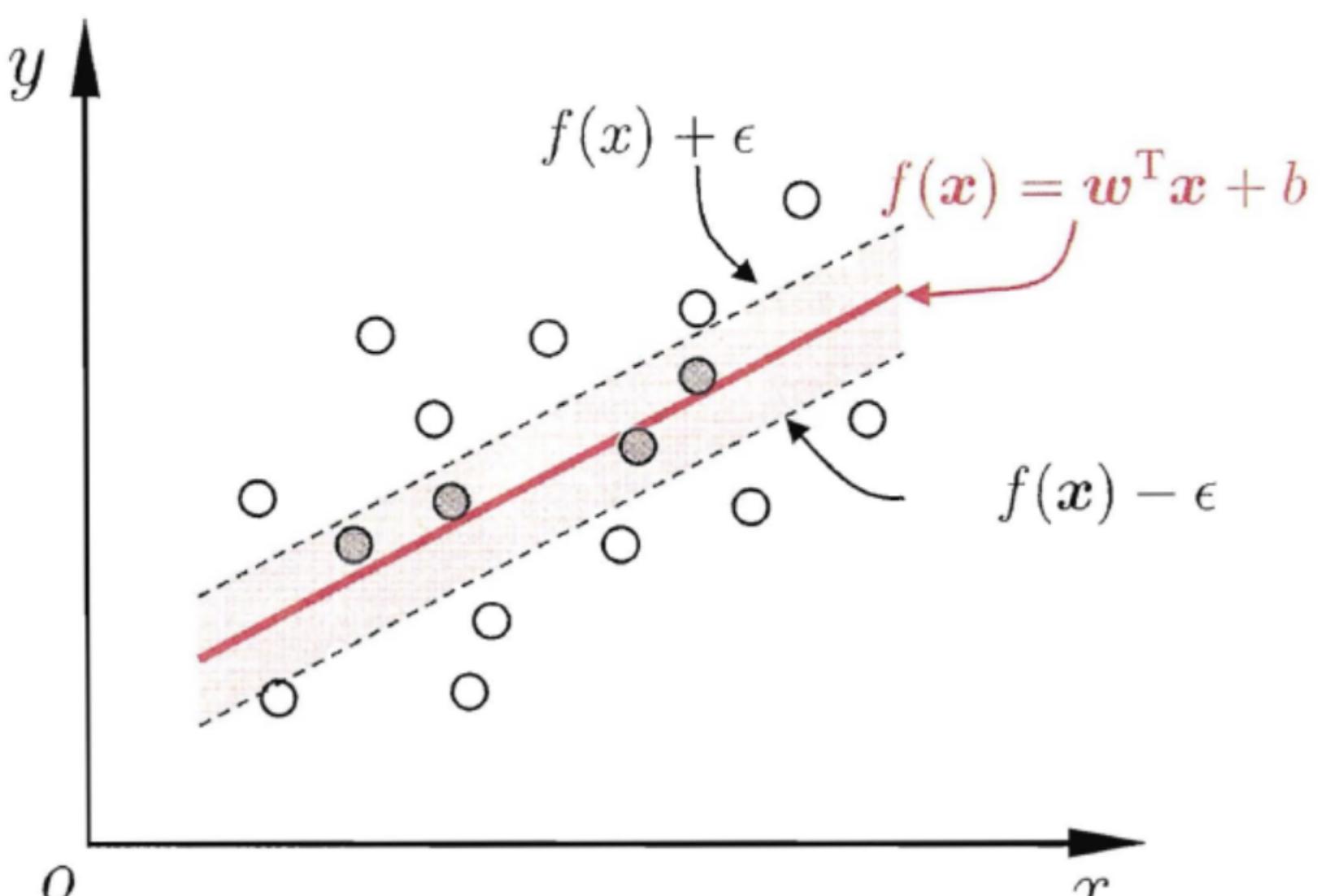
Ридж-регрессия или гребневая регрессия (ridge regression) - это один из методов понижения размерности. Часто его применяют для борьбы с переизбыточностью данных, когда независимые переменные коррелируют друг с другом (т.е. имеет место мультиколлинеарность). Следствием этого является плохая обусловленность матрицы XTX и неустойчивость оценок коэффициентов регрессии. Оценки, например, могут иметь неправильный знак или значения, которые намного превосходят те, которые приемлемы из физических или практических соображений. Регрессия добавляет в процессе прогнозирования штраф за некорректный прогноз, тем самым улучшая конечный результат.

Лассо очень похожа на ридж регрессию. В ней штраф — это сумма модулей значений коэффициентов. Ридж регрессия снижает размер коэффициентов, а лассо сокращает многие до 0. Это позволяет снизить размерность (ридж) и выбрать важные предикторы (лассо).

SVR(Support Vector Regression) – разновидность метода опорных векторов, регрессия опорных векторов. Регрессия SVR заключается в поиске такой плоскости регрессии, чтобы все данные в выборке были наиболее близки к плоскости. Традиционный метод регрессии считает, что прогноз верен тогда и только тогда, когда регрессия f (x) полностью равна y.

Регрессия опорных векторов считает, что до тех пор, пока f (x) не слишком сильно отклоняется от y, прогноз можно считать правильным без вычисления потерь. При этом все, что остается за пределами плоскости (тени) потери модели:

Рис. 5



Предоставленная база данных не позволяет дать четкий ответ на вопрос какие модели следует использовать при построении прогноза. В условиях неопределенности при отсутствии слишком большого объема данных необходимо подобрать такой вариант модели, который наиболее полно сможет описать предложенную зависимость. Регрессия опорных векторов также может предоставить эту возможность.

SGDRegressor - стохастический градиентный спуск. Это простой, но очень эффективный подход к подгонке линейных классификаторов и регрессоров под выпуклые функции потерь, такие как (линейные) Метод опорных векторов и логистическая регрессия. Класс SGDRegressor реализует простую процедуру обучения стохастическим градиентным спуском, которая поддерживает различные функции потерь и штрафы для соответствия моделям линейной регрессии. SGDRegressor хорошо подходят для задач регрессии с большим количеством учебных образцов (> 10.000) или значительными отклонениями, что присутствует в нашей задаче и будет видно на этапе поиска взаимосвязей.

KNeighborsRegressor – К ближайших соседей. Наша задача в чистом виде не относится к методам классификации, но может иметь данные взаимосвязи, что нельзя исключать при анализе данных. Несмотря на свою простоту, функция «Ближайшие соседи» успешно справляется с большим количеством задач классификации и регрессии, включая рукописные цифры и сцены спутниковых изображений. Будучи непараметрическим методом, он часто бывает успешным в ситуациях классификации, когда граница решения очень нерегулярна.   
С учетом неоднозначности нашей задачи, использование данного метода может принести неожиданные результаты.

DecisionTreeRegressor, оно же Дерево принятия решений – 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. Алгоритм дерева решений подпадает под категорию контролируемых алгоритмов обучения. Он работает как для непрерывных, так и для категориальных выходных переменных. 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.

RandomForestRegressor или Случайный лес - это множество решающих деревьев. В задаче регрессии их ответы усредняются, в задаче классификации принимается решение голосованием по большинству. Все деревья строятся независимо по следующей схеме:

1. Выбирается подвыборка обучающей выборки размера samplesize – по ней строится дерево (для каждого дерева — своя подвыборка).
2. Для построения каждого расщепления в дереве просматриваем max\_features случайных признаков (для каждого нового расщепления — свои случайные признаки).
3. Выбирается наилучший признак и расщепление по нему (по заранее заданному критерию).

Дерево строится, как правило, до исчерпания выборки (пока в листьях не останутся представители только одного класса), но в современных реализациях есть параметры, которые ограничивают высоту дерева, число объектов в листьях и число объектов в подвыборке, при котором проводится расщепление.

Как видно из описания, Случайный лес это более сложная структура по сравнению с Деревом решений, которая может показать большую эффективность на одинаковой выборке. Тем не менее порой лучшим вариантом является Дерево решений, а не Случайный лес, если результаты обеих моделей примерно схожи.

GradientBoostingRegressor. Градиентный бустинг — это техника машинного обучения для задач классификации и регрессии, которая строит модель предсказания в форме ансамбля слабых предсказывающих моделей, обычно деревьев решений. Один из самых эффективных инструментов для решения задач машинного обучения. Данный инструмент мы будем использовать в ситуациях, когда прогноз демонстрирует существенно неудовлетворительные результаты прогноза.

# 1.3. Разведочный анализ данных

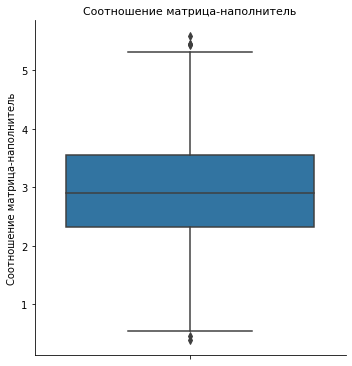
В качестве инструментов разведочного анализа используется оценка статистических характеристик базы данных (см. Таблицу 1), а также матрица попарной корреляции, тепловая карта корреляции, гистограммы нормального распределения, поиск выбросов через «ящик с усами» (catplot). После обнаружения выбросов данные значительно отличающиеся от выборки, в случае их небольшого количества (не более 5%) будут удалены.

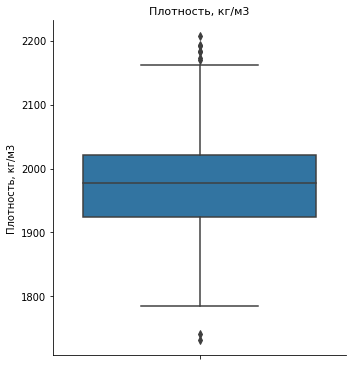
# 2. Практическая часть

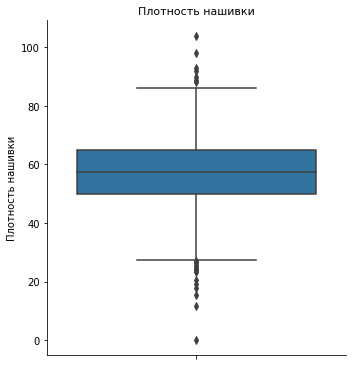
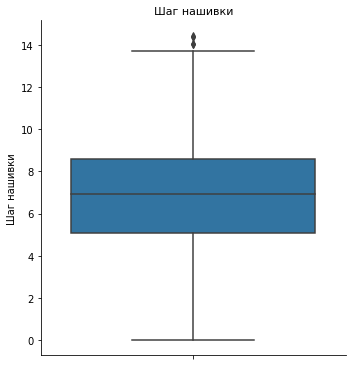
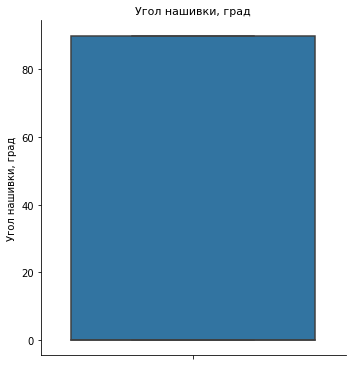
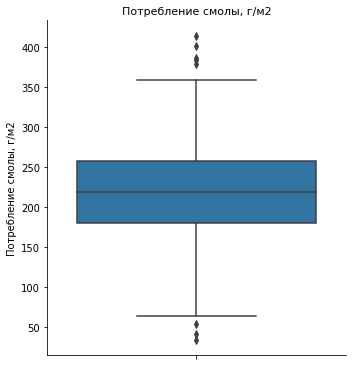
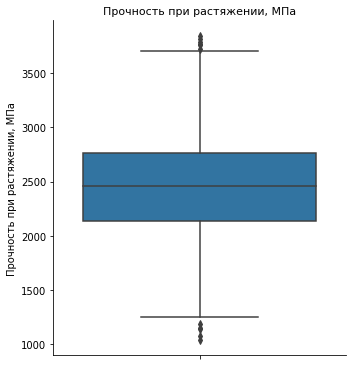
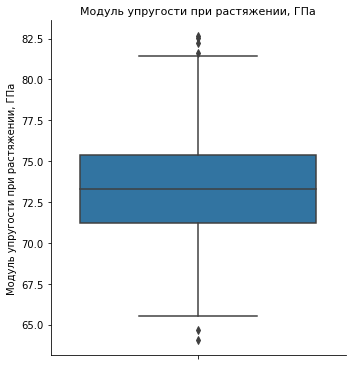
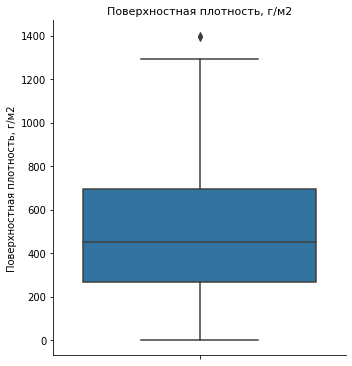
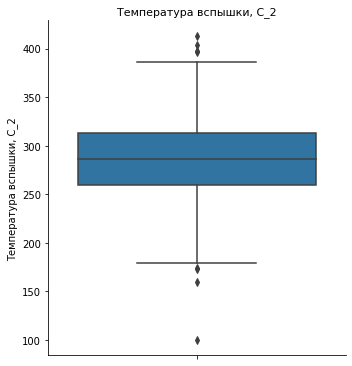
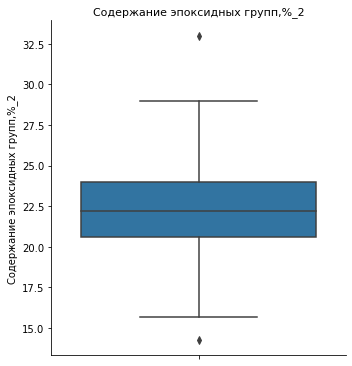
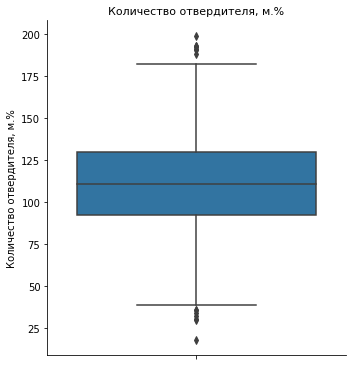
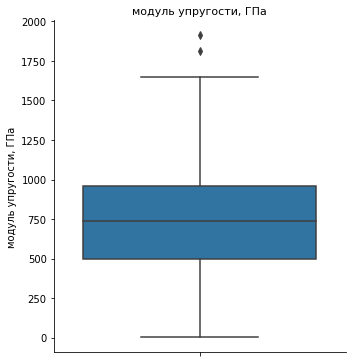
# 2.1. Предобработка данных

Графики данных до обработки:

Рис. 6

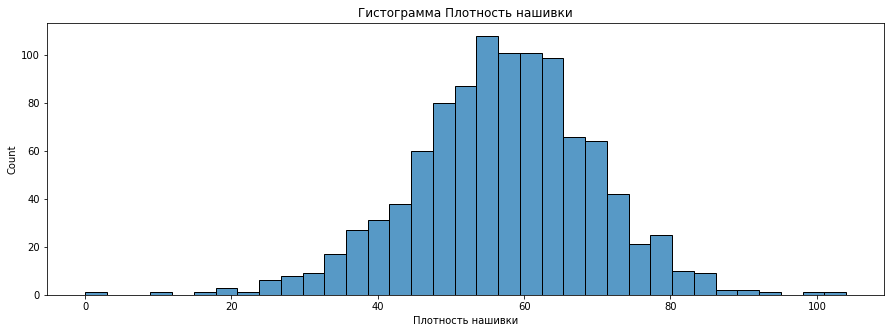
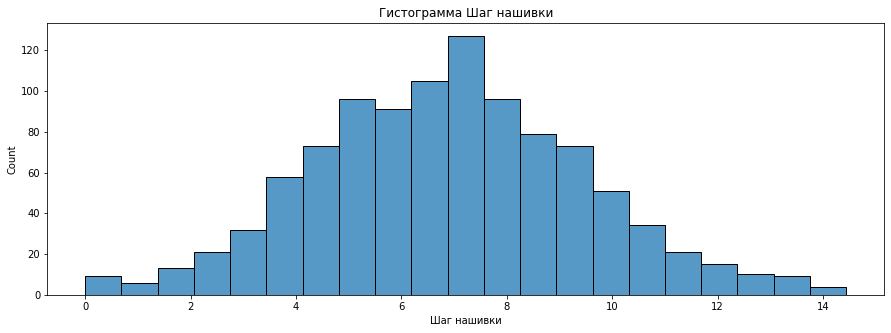
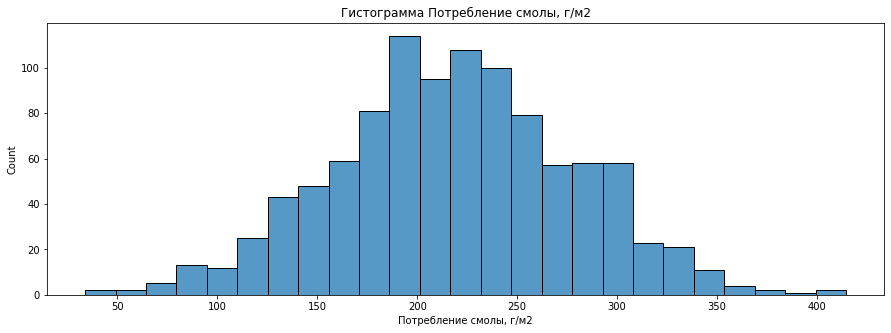
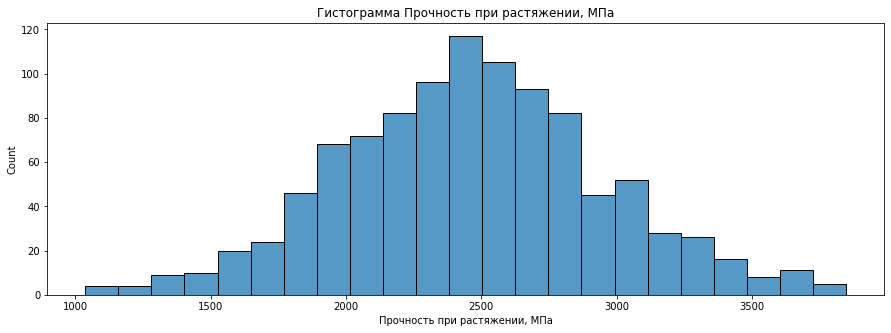
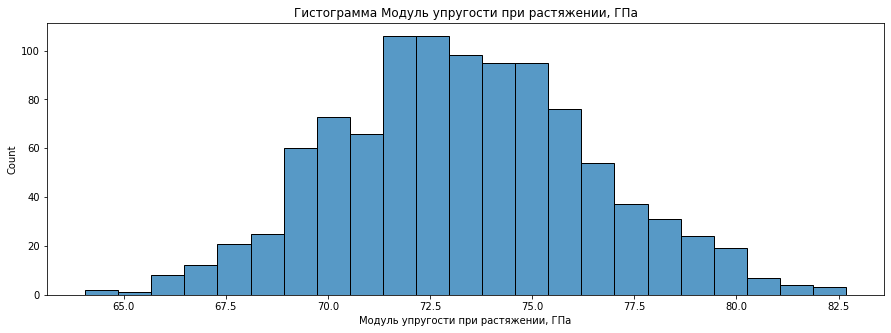
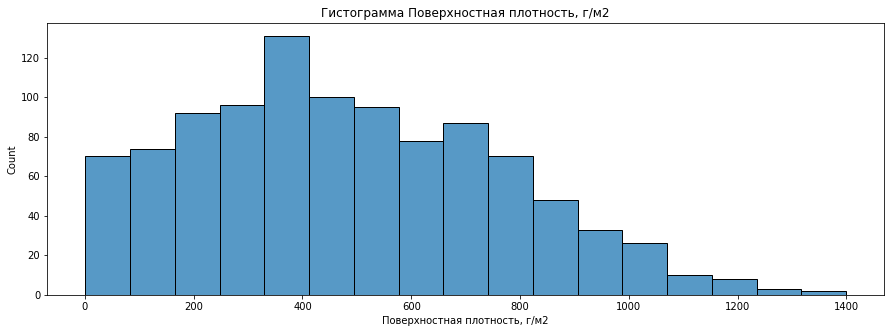
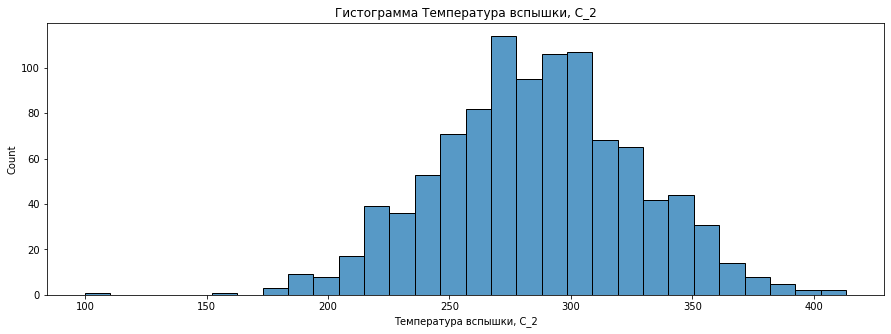
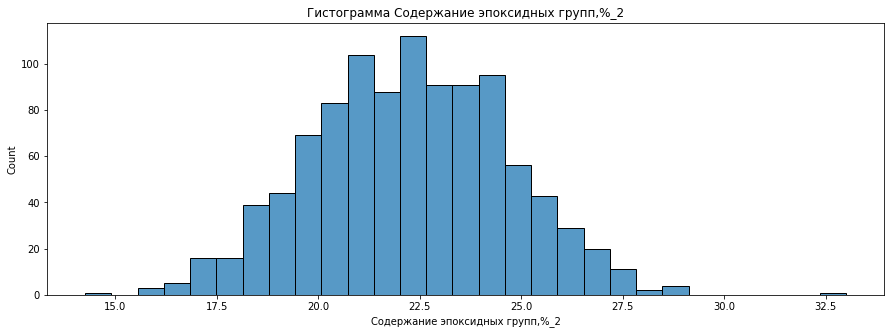
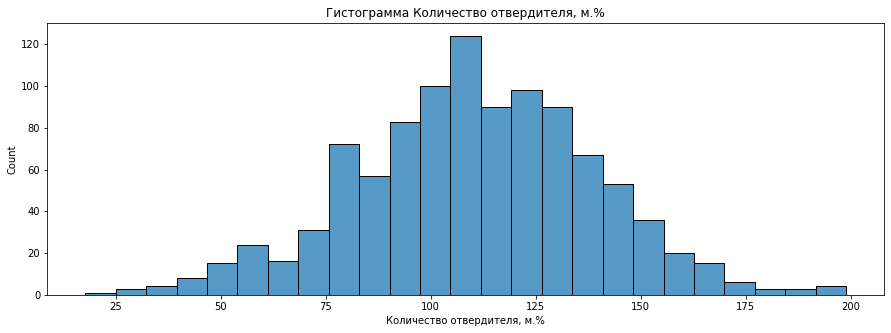
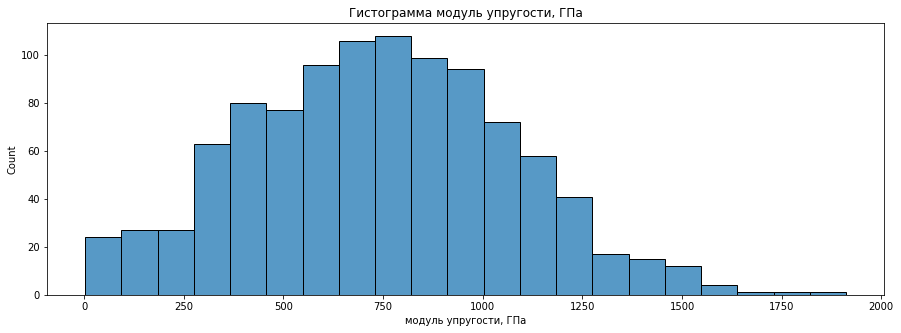
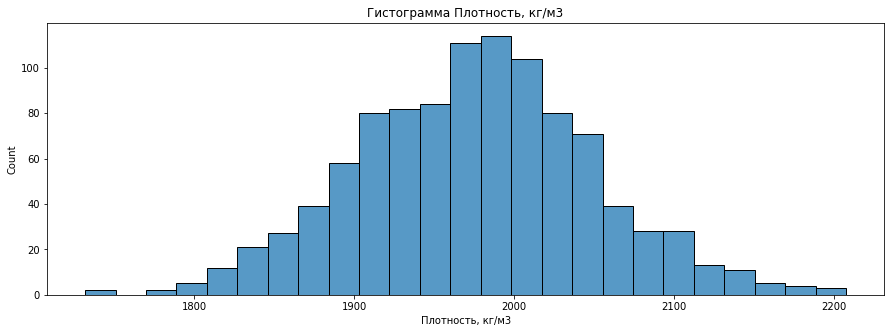
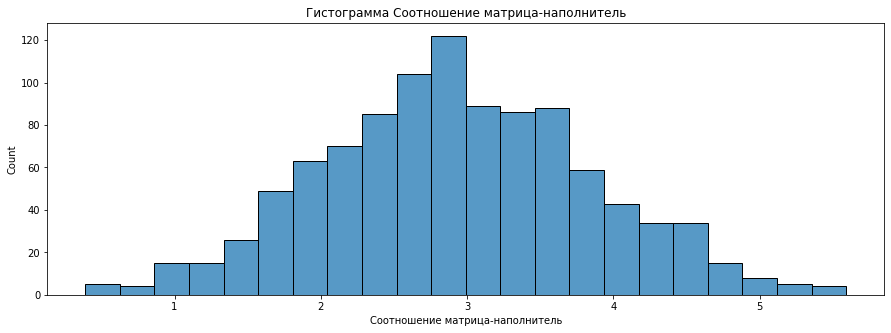






Гистограммы данных до обработки:

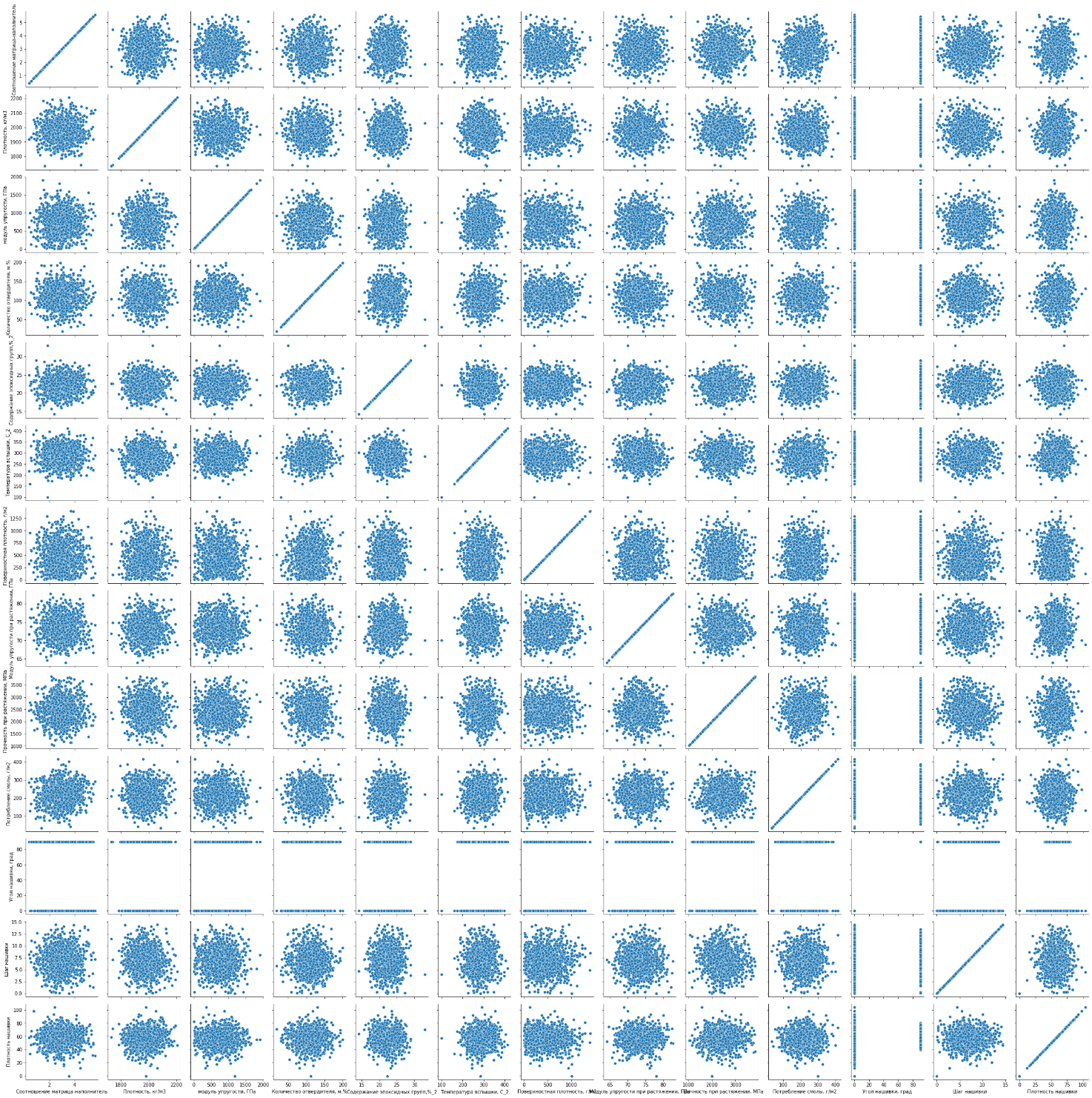
Рис. 7

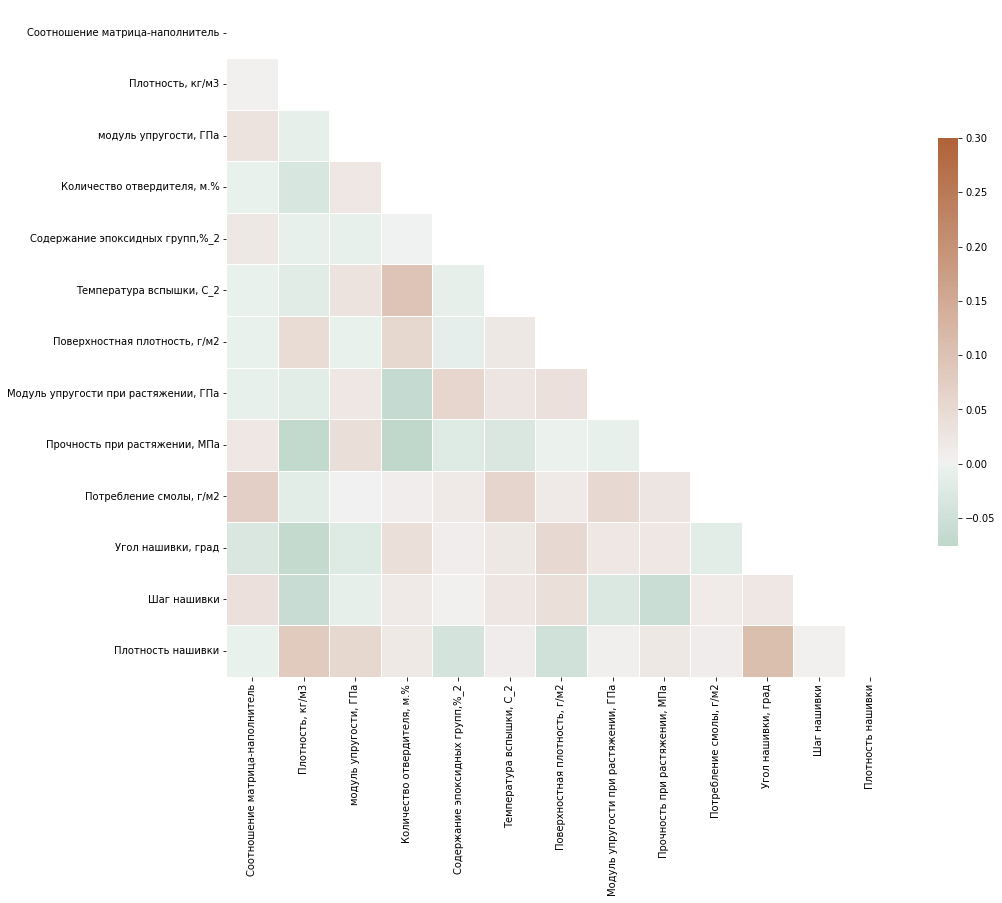


Преимущественно данные стремятся к нормальному распределению. Угол нашивки, как и отражено в датасете, имеет только два значения 90 градусов и 0 градусов, что отражает общий подход к проведению нашивки материалов, а также может быть использовано при обработке данных. Учитывая отсутствие иных показателей для угла нашивки, предлагаем в прогнозе использовать категориальный, а не непрерывный подход при анализе данного параметра.

Матрица корреляции и тепловая карта:

Рис. 8





В ходе предобработки выявлено следующее число выбросов:

Соотношение матрица-наполнитель 6

Плотность, кг/м3 9

модуль упругости, ГПа 2

Количество отвердителя, м.% 14

Содержание эпоксидных групп,%\_2 2

Температура вспышки, С\_2 8

Поверхностная плотность, г/м2 2

Модуль упругости при растяжении, ГПа 6

Прочность при растяжении, МПа 11

Потребление смолы, г/м2 8

Угол нашивки, град 0

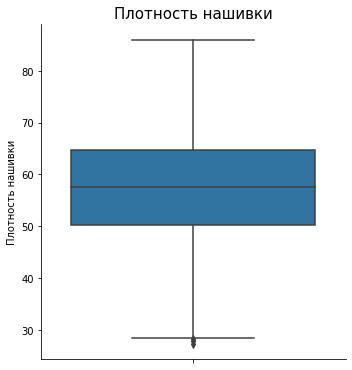
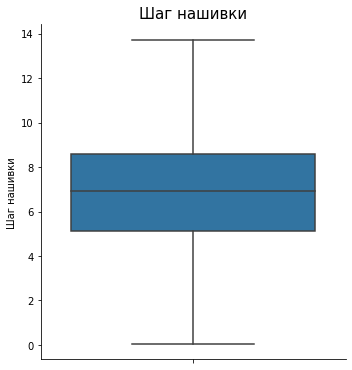
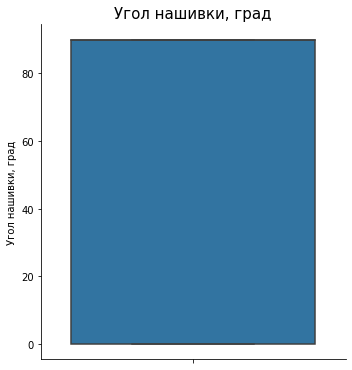
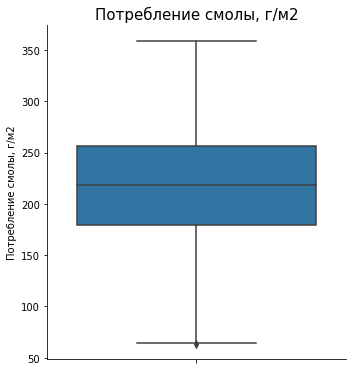
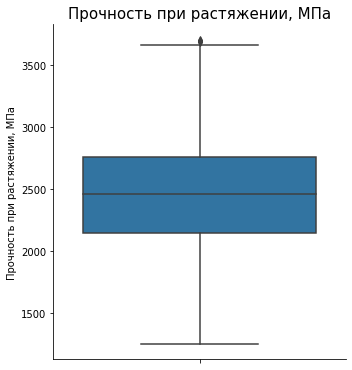
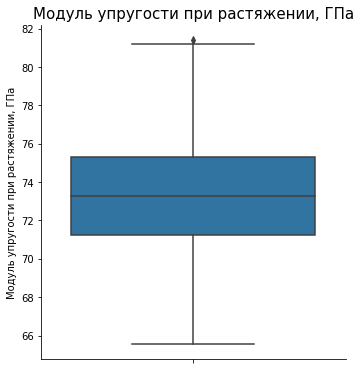
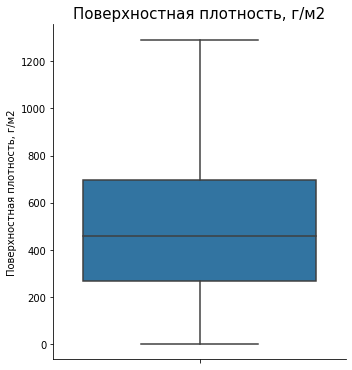
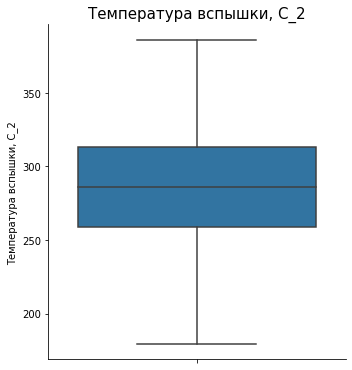
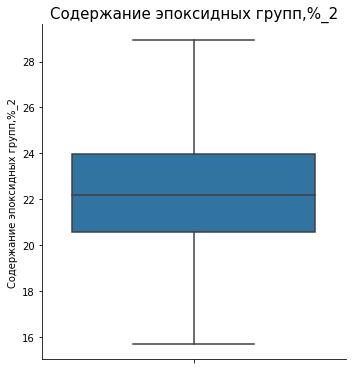
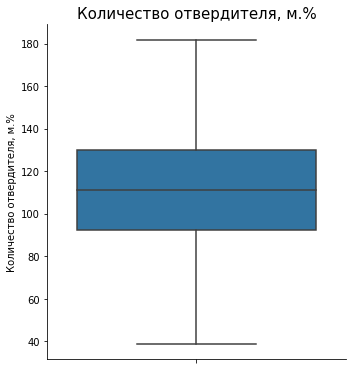
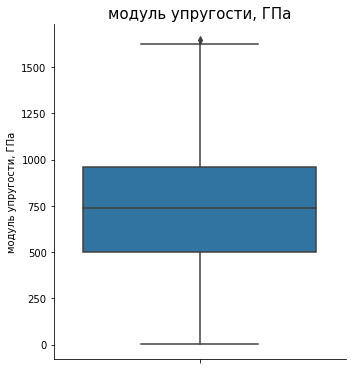
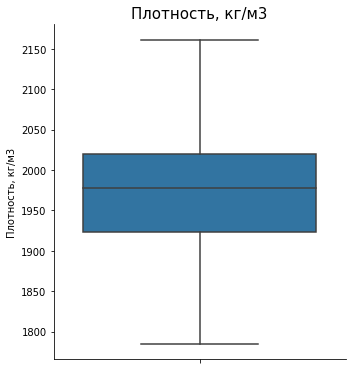
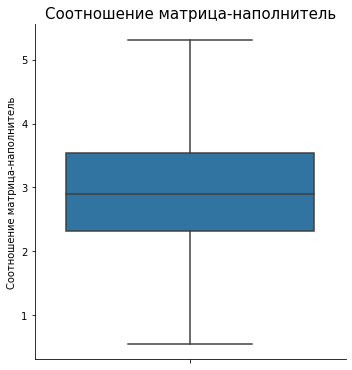
Шаг нашивки 4

Плотность нашивки 21

Всего в базе 1023 эксперимента. Предположим, что исключение данных выбросов не окажет значительного влияния на качество модели.

Ящики после удаления данных:

Рис. 9



В дальнейшем предобработка данных станет частью каждого этапа моделирования. В рамках предобработки данных мы будем менять угол нашивки на категориальный признак, т.к. данный элемент базы имеет всего два значения 0 и 90. Иные показатели будут Стандартизированы с помощью функции StandardScaler.

Рассмотрим данные для каждой задачи до и после предобработки.

Модуль упругости при растяжении:

До обработки:

Таблица 2

|  | **min** | **max** | **mean** | **std** |
| --- | --- | --- | --- | --- |
| **1** | **2** | **3** | **4** | **5** |
| Соотношение матрица-наполнитель | 0.547391 | 5.314144 | 2.876888 | 0.880431 |
| Плотность, кг/м3 | 1801.940695 | 2160.000000 | 1974.191536 | 69.133748 |
| модуль упругости, ГПа | 2.436909 | 1628.000000 | 740.164541 | 328.999060 |
| Количество отвердителя, м.% | 38.668500 | 181.828448 | 112.382558 | 27.103285 |
| Содержание эпоксидных групп,%\_2 | 15.695894 | 28.955094 | 22.135760 | 2.367513 |
| Температура вспышки, С\_2 | 186.508613 | 386.067992 | 286.075009 | 39.172566 |
| Поверхностная плотность, г/м2 | 0.603740 | 1291.340115 | 485.309834 | 284.252109 |
| Потребление смолы, г/м2 | 63.685698 | 356.925975 | 215.531927 | 58.472463 |
| Угол нашивки, град | 0.000000 | 90.000000 | 45.343511 | 45.033078 |
| Шаг нашивки | 0.145034 | 13.732404 | 6.925298 | 2.541034 |
| Плотность нашивки | 28.237746 | 86.012427 | 57.332232 | 11.208720 |

После обработки:

Таблица 3

|  | **min** | **max** | **mean** | **std** |
| --- | --- | --- | --- | --- |
| 1 | 2 | 3 | 4 | 5 |
| Соотношение матрица-наполнитель | -2.647882 | 2.770369 | 0.000000 | 1.000764 |
| Плотность, кг/м3 | -2.493464 | 2.689721 | 0.000000 | 1.000764 |
| модуль упругости, ГПа | -2.244053 | 2.700658 | 0.000000 | 1.000764 |
| Количество отвердителя, м.% | -2.721825 | 2.564227 | 0.000000 | 1.000764 |
| Содержание эпоксидных групп,%\_2 | -2.722176 | 2.882580 | 0.000000 | 1.000764 |
| Температура вспышки, С\_2 | -2.543680 | 2.554579 | 0.000000 | 1.000764 |
| Поверхностная плотность, г/м2 | -1.706501 | 2.837785 | 0.000000 | 1.000764 |
| Потребление смолы, г/м2 | -2.598869 | 2.419979 | 0.000000 | 1.000764 |
| Шаг нашивки | -2.670349 | 2.680920 | 0.000000 | 1.000764 |
| Плотность нашивки | -2.597685 | 2.560695 | 0.000000 | 1.000764 |
| Угол нашивки, град | 0.000000 | 1.000000 | 0.503817 | 0.500368 |

Прочность при растяжении:

До обработки:

Таблица 4

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **min** | **max** | **mean** | **std** |
| 1 | 2 | 3 | 4 | 5 |
| Соотношение матрица-наполнитель | 0.547391 | 5.314144 | 2.876888 | 0.880431 |
| Плотность, кг/м3 | 1801.940695 | 2160.000000 | 1974.191536 | 69.133748 |
| модуль упругости, ГПа | 2.436909 | 1628.000000 | 740.164541 | 328.999060 |
| Количество отвердителя, м.% | 38.668500 | 181.828448 | 112.382558 | 27.103285 |
| Содержание эпоксидных групп,%\_2 | 15.695894 | 28.955094 | 22.135760 | 2.367513 |
| Температура вспышки, С\_2 | 186.508613 | 386.067992 | 286.075009 | 39.172566 |
| Поверхностная плотность, г/м2 | 0.603740 | 1291.340115 | 485.309834 | 284.252109 |
| Потребление смолы, г/м2 | 63.685698 | 356.925975 | 215.531927 | 58.472463 |
| Угол нашивки, град | 0.000000 | 90.000000 | 45.343511 | 45.033078 |
| Шаг нашивки | 0.145034 | 13.732404 | 6.925298 | 2.541034 |
| Плотность нашивки | 28.237746 | 86.012427 | 57.332232 | 11.208720 |

После обработки:

Таблица 5

|  | **min** | **max** | **mean** | **std** |
| --- | --- | --- | --- | --- |
| 1 | 2 | 3 | 4 | 5 |
| Соотношение матрица-наполнитель | -2.647882 | 2.770369 | 0.000000 | 1.000764 |
| Плотность, кг/м3 | -2.493464 | 2.689721 | 0.000000 | 1.000764 |
| модуль упругости, ГПа | -2.244053 | 2.700658 | 0.000000 | 1.000764 |
| Количество отвердителя, м.% | -2.721825 | 2.564227 | 0.000000 | 1.000764 |
| Содержание эпоксидных групп,%\_2 | -2.722176 | 2.882580 | 0.000000 | 1.000764 |
| Температура вспышки, С\_2 | -2.543680 | 2.554579 | 0.000000 | 1.000764 |
| Поверхностная плотность, г/м2 | -1.706501 | 2.837785 | 0.000000 | 1.000764 |
| Потребление смолы, г/м2 | -2.598869 | 2.419979 | 0.000000 | 1.000764 |
| Шаг нашивки | -2.670349 | 2.680920 | 0.000000 | 1.000764 |
| Плотность нашивки | -2.597685 | 2.560695 | 0.000000 | 1.000764 |
| Угол нашивки, град | 0.000000 | 1.000000 | 0.503817 | 0.500368 |

Соотношения матрица-наполнитель:

До обработки:

Таблица 6

|  | **min** | **max** | **mean** | **std** |
| --- | --- | --- | --- | --- |
| 1 | 2 | 3 | 4 | 5 |
| Плотность, кг/м3 | 1801.940695 | 2160.000000 | 1974.191536 | 69.133748 |
| модуль упругости, ГПа | 2.436909 | 1628.000000 | 740.164541 | 328.999060 |
| Количество отвердителя, м.% | 38.668500 | 181.828448 | 112.382558 | 27.103285 |
| Содержание эпоксидных групп,%\_2 | 15.695894 | 28.955094 | 22.135760 | 2.367513 |
| Температура вспышки, С\_2 | 186.508613 | 386.067992 | 286.075009 | 39.172566 |
| Поверхностная плотность, г/м2 | 0.603740 | 1291.340115 | 485.309834 | 284.252109 |
| Модуль упругости при растяжении, ГПа | 65.793845 | 81.417126 | 73.345263 | 3.010881 |
| Прочность при растяжении, МПа | 1250.392802 | 3705.672523 | 2459.535981 | 464.807054 |
| Потребление смолы, г/м2 | 63.685698 | 356.925975 | 215.531927 | 58.472463 |
| Угол нашивки, град | 0.000000 | 90.000000 | 45.343511 | 45.033078 |
| Шаг нашивки | 0.145034 | 13.732404 | 6.925298 | 2.541034 |
| Плотность нашивки | 28.237746 | 86.012427 | 57.332232 | 11.208720 |

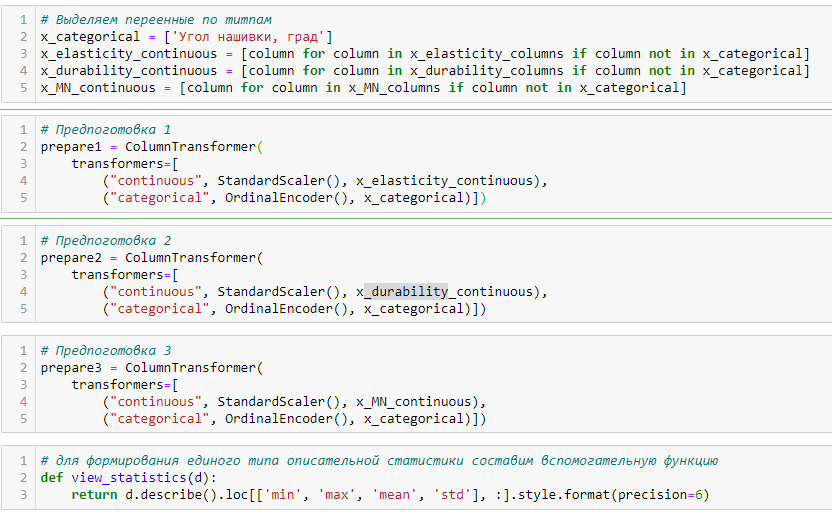
После обработки:

Таблица 7

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **min** | **max** | **mean** | **std** |
| 1 | 2 | 3 | 4 | 5 |
| Плотность, кг/м3 | -2.493464 | 2.689721 | 0.000000 | 1.000764 |
| модуль упругости, ГПа | -2.244053 | 2.700658 | 0.000000 | 1.000764 |
| Количество отвердителя, м.% | -2.721825 | 2.564227 | 0.000000 | 1.000764 |
| Содержание эпоксидных групп,%\_2 | -2.722176 | 2.882580 | 0.000000 | 1.000764 |
| Температура вспышки, С\_2 | -2.543680 | 2.554579 | 0.000000 | 1.000764 |
| Поверхностная плотность, г/м2 | -1.706501 | 2.837785 | 0.000000 | 1.000764 |
| Модуль упругости при растяжении, ГПа | -2.509959 | 2.682946 | 0.000000 | 1.000764 |
| Прочность при растяжении, МПа | -2.603375 | 2.683025 | -0.000000 | 1.000764 |
| Потребление смолы, г/м2 | -2.598869 | 2.419979 | 0.000000 | 1.000764 |
| Шаг нашивки | -2.670349 | 2.680920 | 0.000000 | 1.000764 |
| Плотность нашивки | -2.597685 | 2.560695 | 0.000000 | 1.000764 |
| Угол нашивки, град | 0.000000 | 1.000000 | 0.503817 | 0.500368 |

Этап предобработки вынесен в отдельный инструмент для повышения удобства программирования дальнейшей оптимизации входящего потока данных от пользователя при построении программы.

В качестве оптимизатора во всех моделях используется следующий код:



Данный код позволяет применять к данным единый подход при обработке для любой модели, а также выдает единый формат представления данных в постобработанном и предобработанном форматах, что позволило включить в данную работу выше представленные таблицы.

# 2.2. Разработка и обучение модели

В рамках решения задач прогнозирования Модуля упругости при растяжении, Прочности при растяжении и Соотношения матрица-наполнитель будут использованы:

* DummyRegressor
* LinearRegression
* Ridge
* Lasso
* SVR
* SGDRegressor
* GradientBoostingRegressor
* RandomForestRegressor)
* MLPRegressor

Sequential

# 2.3. Тестирование модели

Модуль упругости при растяжении:

Таблица 8

|  | **R2** | **RMSE** | **MAE** | **MAPE** | **max\_error** |
| --- | --- | --- | --- | --- | --- |
| 1 | 2 | 3 | 4 | 5 | 6 |
| DummyRegressor | -0.012059 | -2.998439 | -2.442075 | -0.033391 | -7.113542 |
| Ridge | -0.012754 | -2.999374 | -2.443103 | -0.033406 | -7.114296 |
| Lasso | -0.012059 | -2.998439 | -2.442075 | -0.033391 | -7.113542 |
| SVR | -0.011873 | -2.998499 | -2.444446 | -0.033439 | **-7.112214** |
| SGDRegressor | -0.012099 | -2.998512 | -2.442082 | -0.033393 | -7.115203 |
| KNeighborsRegressor | -0.053093 | -3.056920 | -2.464994 | -0.033714 | -7.356232 |
| DecisionTreeRegressor | -0.030998 | -3.025149 | -2.466529 | -0.033725 | -7.207054 |
| RandomForestRegressor | **-0.008977** | **-2.994092** | **-2.441902** | **-0.033389** | -7.120131 |

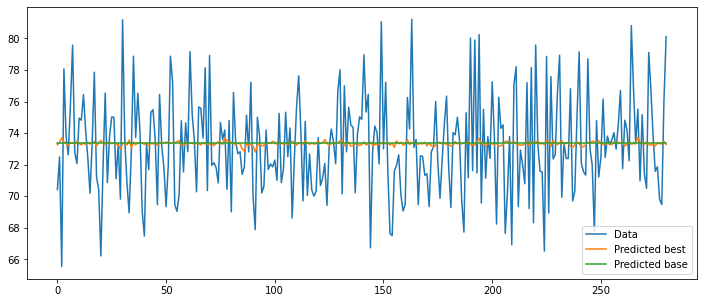
Лучшая модель – Случайный лес:

Подобранные гиперпараметры модели:

RandomForestRegressor (max\_depth=2, max\_features=1, n\_estimators=75)

Отображение лучшей модели при предсказании:

Рис. 10



Прочность при растяжении:

Таблица 9

|  | **R2** | **RMSE** | **MAE** | **MAPE** | **max\_error** |
| --- | --- | --- | --- | --- | --- |
| 1 | 2 | 3 | 4 | 5 | 6 |
| DummyRegressor | -0.022998 | -464.11 | -372.58 | -0.161845 | -1164.177 |
| Ridge | -0.023157 | -464.13 | -372.69 | -0.161880 | -1164.93 |
| Lasso | -0.022998 | -464.11 | -372.58 | -0.161845 | -1164.17 |
| SVR | -0.025665 | -464.65 | -373.54 | -0.161967 | -1166.93 |
| SGDRegressor | **-0.021077** | **-463.67** | **-372.35** | **-0.161673** | -1163.09 |
| GradientBoostingRegressor | -0.035576 | -466.60 | -374.45 | -0.162561 | -1191.89 |
| RandomForestRegressor | -0.021457 | -463.70 | -372.92 | -0.161978 | **-1162.64** |

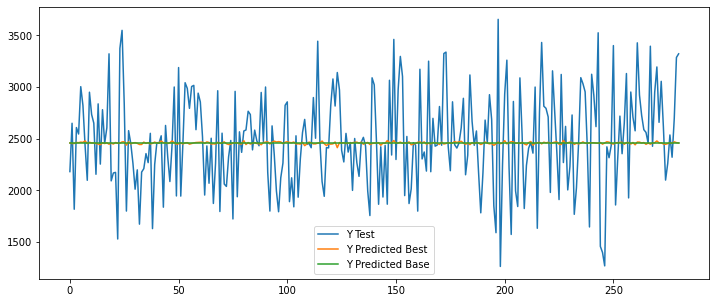
Лучшая модель – SGDRegressor:

Подобранные гиперпараметры модели:

SGDRegressor (alpha=15, l1\_ratio=0.1, penalty=l2)

Отображение лучшей модели при предсказании:

Рис. 11



Соотношения матрица-наполнитель:

Таблица 10

|  | **R2** | **RMSE** | **MAE** | **MAPE** | **max\_error** |
| --- | --- | --- | --- | --- | --- |
| 1 | 2 | 3 | 4 | 5 | 6 |
| DummyRegressor | -0.007274 | -0.878669 | -0.699270 | -0.308229 | -2.097731 |
| Ridge | -0.008287 | -0.879113 | -0.699702 | -0.308379 | -2.095200 |
| Lasso | -0.007274 | -0.878669 | -0.699270 | -0.308229 | -2.097731 |
| SVR | -0.008528 | -0.879204 | **-0.699139** | -0.304915 | -2.104737 |
| SGDRegressor | -0.006729 | -0.878468 | -0.699301 | -0.308303 | -2.095016 |
| GradientBoostingRegressor | -0.010176 | -0.879973 | -0.701529 | **-0.304588** | -2.163472 |
| RandomForestRegressor | **-0.000149** | **-0.875634** | -0.700383 | -0.308187 | **-2.080780** |

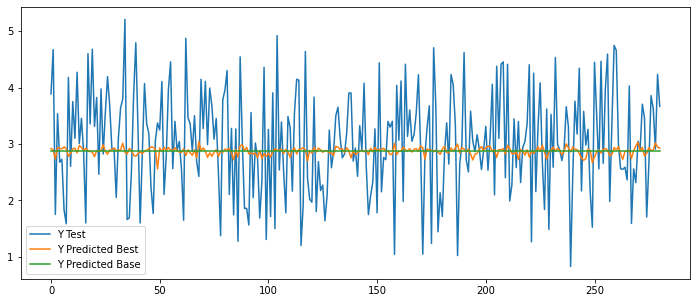
Лучшее из имеющегося прогноза – Случайный лес:

Подобранные гиперпараметры модели:

RandomForestRegressor (max\_depth=4, max\_features=1, n\_estimators=75)

Отображение лучшей модели при предсказании:

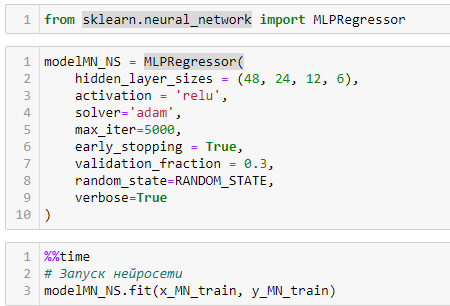
Рис. 12

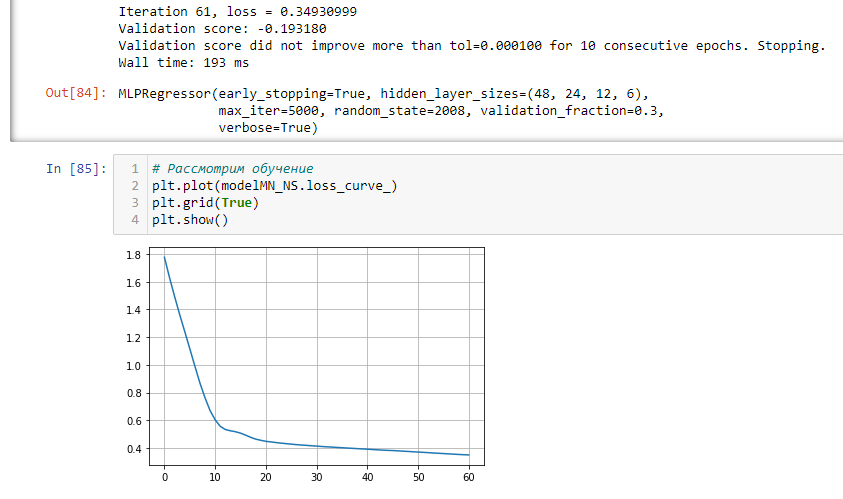


# 2.4. Написать нейронную сеть, которая будет рекомендовать соотношение матрица-наполнитель.

Нейронная сеть 1:

Рис. 13



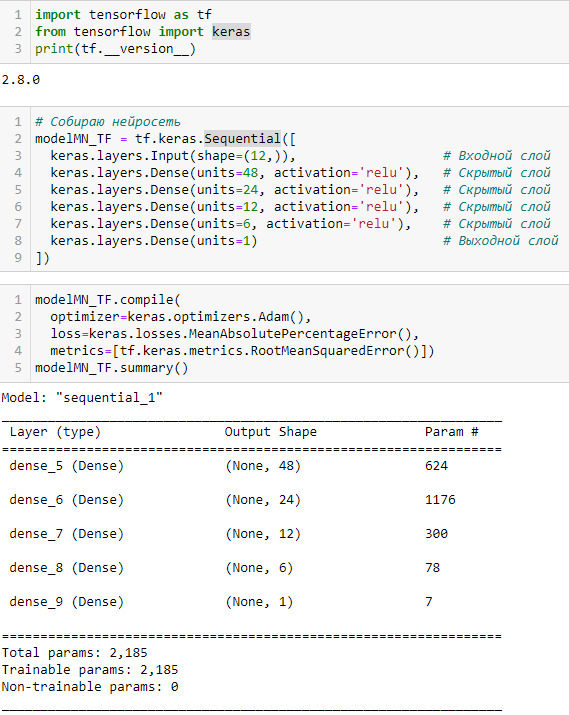


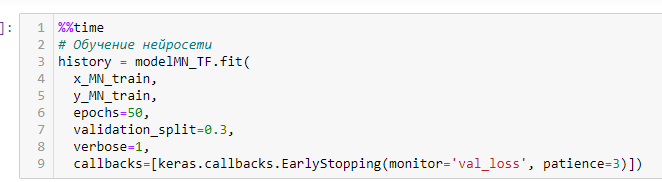


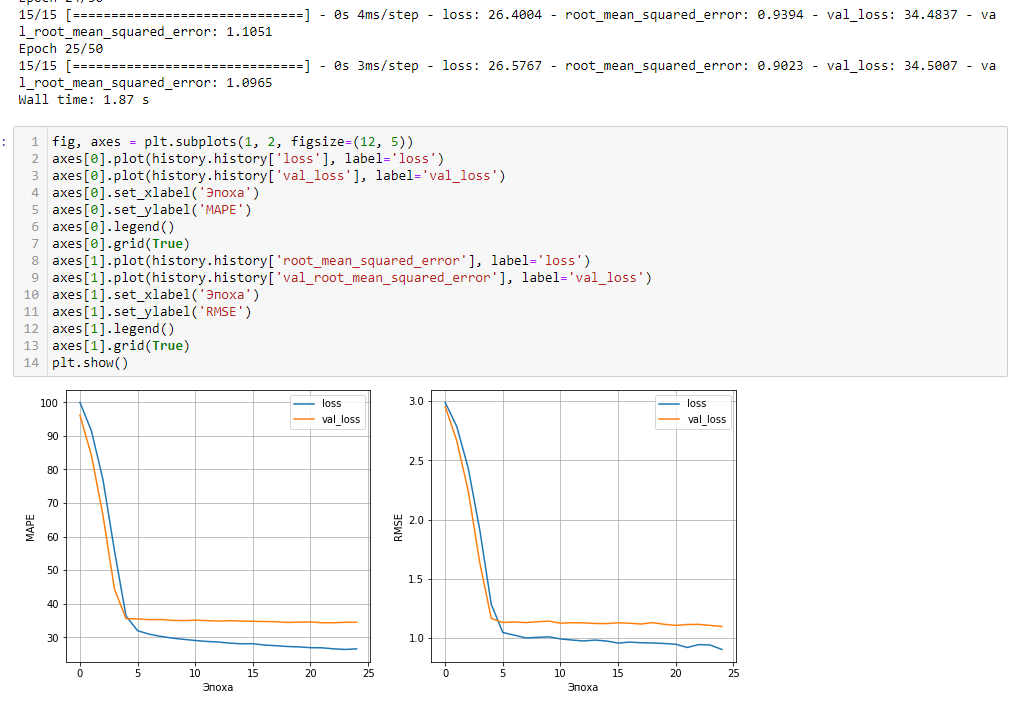
Нейросеть, построенная с помощью MLPRegressor, возвращает приближенные к исходным, но ошибка по всем метрикам значительно хуже, чем у линейных моделей. Возможно, ошибку можно уменьшить, используя другую нейросеть.

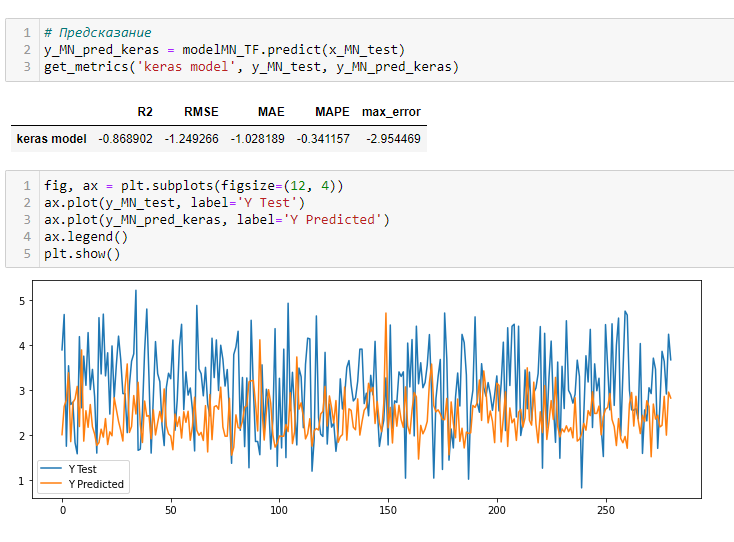
Нейронная сеть 2:

Рис. 14









Проблемы точности сохраняются. При схожем числе слоев и нейронов TensorFlow показало лучший график, но большую ошибку, чем MLPRegressor. Ошибка по любой из нейросетей, чем у рассмотренных выше моделей. Лучший прогноз по-прежнему дает случайный лес.

# 2.5. Разработка приложения

Приложение состоит из файла Flask, который содержит в себе весь основной функционал и методы обработки, папки templates, где лежат основные шаблоны страниц и блока моделей с предобработчиками для входных данных. На вкладке main необходимо выбрать тип желаемой задачи предсказание упругости и прочности или предсказание соотношения матрица/наполнитель, после чего пользователь увидит экран с запросом данных.

В запросе уже стоят данные, на которых можно проверить модель, либо данные можно заменить собственными с соблюдением требований диапазонов. Если какие-то значения диапазонов выходят за рамки допустимых пользователь получает соответствующую ошибку. Если все в норме – программа распаковывает необходимые предобработчики и модели и запускает прогноз.

На выходе пользователь видит значения упругости, прочности или соотношения матрица/наполнитель в зависимости от его запроса.

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

<https://github.com/Velervv/VKR>

# Заключение

Использованные при разработке моделей подходы не позволили получить сколь-нибудь достоверных прогнозов. 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.

В целом прогнозирование конечных свойств/характеристик композитных материалов без изучения материаловедения, погружения в вопрос экспериментального анализа характеристик композитных материалов не демонстрирует сколь-нибудь удовлетворительных результатов. Проработка моделей и построение прогнозов требует внедрения в процесс производных от имеющихся показателей для выявления иного уровня взаимосвязей.

# Библиографический список

1. Абросимов Н.А.: Методика построения разрешающей системы уравнений динамического деформирования композитных элементов конструкций (Учебно-методическое пособие), ННГУ, 2010
2. Материалы конференции: V Всероссийская научно-техническая конференция «Полимерные композиционные материалы и производственные технологии нового поколения», 19 ноября 2021 г.
3. Реутов Ю.А.: Прогнозирование свойств полимерных композиционных материалов и оценка надежности изделий из них, Диссертация на соискание ученой степени кандидата физико-математических наук, Томск 2016.
4. Абу-Хасан Махмуд, Масленникова Л. Л.: Прогнозирование свойств композиционных материалов с учетом наноразмера частиц и акцепторных свойств катионов твердых фаз, статья 2006 год