

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 «DataScience»

Тема: Прогнозирование конечных свойств новых материалов
(композиционных материалов)

Слушатель

Мартьянова Александра Евгеньевна

Москва, 2023

Аттестационное задание

Тема: Прогнозирование конечных свойств новых материалов (композиционных материалов).

Описание:

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.

Датасет со свойствами композитов. Объединение делать по индексу тип объединения INNER

https://drive.google.com/file/d/1B1s5gBlvgU81H9GGolLQVw_SOi-vyNf2/view?usp=sharing
Обучить алгоритм машинного обучения, который будет определять значения:

- Модуль упругости при растяжении, ГПа
- Прочность при растяжении, МПа

Написать нейронную сеть, которая будет рекомендовать:

Соотношение матрица-наполнитель

Написать приложение, которое будет выдавать прогноз полученный в задании 2 или 3 (один или два прогноза, на выбор учащегося)

Создать профиль на github.com

Сделать commit приложения на github.com

Сделать commit на веб-хостинг (По желанию учащегося)

Написать пояснительную записку к проекту, которая включает блок-схему и описание процесса подготовки, обучения моделей и инструкцию по установке и запуску приложения.

СОДЕРЖАНИЕ

Введение	4
1 Аналитическая часть	6
1.1 Методы машинного обучения	6
1.2 Постановка задачи	13
2 Практическая часть	15
2.1 Препреподготовка данных	15
2.2 Исследовательский анализ данных	16
2.3 Предобработка данных	33
2.4 Разработка моделей машинного обучения и обучение нейросети	34
2.5 Размещение работы в репозитории GitHub	73
Заключение	74
Библиографический список	75
Приложение 1. Описательная статистика исходного датасета	76
Приложение 2. Значимость коэффициентов корреляции датасета после очистки методом межквартильного расстояния без параметров 'Относительное удлинение при растяжении' и 'Относительное удлинение'	78
Приложение 3. Архитектура нейронных сетей для предсказания параметров по данным исходного датасета без параметров 'Относительное удлинение при растяжении' и 'Относительное удлинение'	79
Приложение 4. Описательная статистика кластеров	87
Приложение 5. Сводные данные по значимости статистики срезов	93

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, то есть 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 работы: созданные прогнозные модели помогут сократить количество проводимых испытаний, а также пополнить базу данных материалов возможными новыми характеристиками материалов, и цифровыми двойниками новых композитов.

Для обработки предлагается датасет в двух таблицах со свойствами композитов, расположенный по адресу:

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

Объединение двух таблиц необходимо сделать по индексу, тип объединения INNER.

Целью работы является создание моделей и обучение алгоритмов машинного обучения, которые будут определять значения:

- Модуль упругости при растяжении, ГПа,
- Прочность при растяжении, МПа,

и написание нейронной сети, которая будет рекомендовать:

- Соотношение матрица-наполнитель.

Этапы, которые необходимо выполнить в работе для достижения поставленной цели:

- изучение теоретического материала,
- выполнение препреподготовки данных,
- выполнение исследовательского анализа данных,
- выполнение предобработки данных,
- выполнение разработки моделей машинного обучения и обучение нейронной сети,
- выполнение размещения работы в репозитории GitHub.

Настоящая выпускная квалификационная работа содержит введение, два раздела основной части, заключение, пять приложений, пятьдесят семь рисунков и восемь таблиц.

Ссылка на размещение работы в репозитории GitHub:

https://github.com/lksnd/bmstu_fqw

1.1 Методы машинного обучения

Предсказание значений вещественной, непрерывной переменной — это задача регрессии. Зависимая переменная должна иметь связь с одной или несколькими независимыми переменными, называемых также предикторами или регрессорами. Регрессионный анализ помогает понять, как типичное значение зависимой переменной изменяется при изменении независимых переменных.

В настоящее время разработано много методов регрессионного анализа. Например, простая и множественная линейная регрессия. Эти модели являются параметрическими в том смысле, что функция регрессии определяется конечным числом неизвестных параметров, которые оцениваются на основе данных.

1.1.1 Регрессия k-ближайших соседей

Метод k-ближайших соседей (kNN — kNearestNeighbours) — метод решения задач классификации и задач регрессии, основанный на поиске ближайших объектов с известными значениями целевой переменной.

Для целевой переменной метод предполагает найти ближайшие к нему объекты $x_1, x_2 \dots x_k$ и построить прогноз по их меткам, то есть определить границы классов и выстроить гиперплоскость регрессии. Метка, назначенная целевой переменной, вычисляется на основе среднего значения меток ее ближайших соседей.

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).

Преимущества данного метода — простая реализация, низкая чувствительность к выбросам, отсутствие необходимости строить модель, настраивать несколько параметров или делать дополнительные допущения, универсальность.

Метод `kNearestNeighbours` — это простой алгоритм машинного обучения с учителем, который можно использовать для решения задач классификации и регрессии. Он прост в реализации и понимании, но имеет существенный недостаток — значительное замедление работы, когда объем данных растет.

На языке Python `k`-ближайших соседей реализован в `sklearn.neighbors.KNeighborsRegressor()`.

1.1.2 Стохастический градиентный спуск

Стохастический градиентный спуск (`SGDRegressor`) — это простой, но очень эффективный подход к подгонке линейных классификаторов и регрессоров под выпуклые функции потерь. Этот подход подразумевает корректировку весов нейронной сети, используя аппроксимацию градиента функционала, вычисленную только на одном случайном обучающем примере из выборки. Достоинства метода: эффективен; прост в реализации; имеет множество возможностей для настройки кода; способен обучаться на избыточно больших выборках. Недостатки метода: требует ряд гиперпараметров; чувствителен к масштабированию функций; может не сходиться или сходиться слишком медленно; функционал многоэкстремален; процесс может «застрять» в одном из локальных минимумов; возможно переобучение.

На языке Python стохастический градиентный спуск реализован в `sklearn.linear_model.SGDRegressor()`.

1.1.3 Линейная регрессия

За базовую модель для прогнозирования всех искомых параметров принята линейная регрессия. `LinearRegression` соответствует линейной модели с коэффициентами $w = (w_1, w_2, \dots, w_p)$, чтобы минимизировать остаточную сумму квадратов между наблюдаемыми целями в наборе данных и целями, предсказанными линейным приближением.

Простая линейная регрессия имеет место, если рассматривается зависимость между одной входной и одной выходной переменными. Для этого опреде-

ляется уравнение регрессии (1) и строится соответствующая прямая, известная как линия регрессии.

$$y = ax + b \quad (1)$$

Коэффициенты a и b , называемые также параметрами модели, определяются таким образом, чтобы сумма квадратов отклонений точек, соответствующих реальным наблюдениям данных, от линии регрессии была бы минимальной. Коэффициенты обычно оцениваются методом наименьших квадратов.

Если ищется зависимость между несколькими входными и одной выходной переменными, то имеет место множественная линейная регрессия. Соответствующее уравнение имеет вид (2).

$$Y = b_0 + b_1 * x_1 + b_2 * x_2 + \dots + b_n * x_n, \quad (2)$$

где n — число входных переменных.

В данном случае модель будет описываться не прямой, а гиперплоскостью. Коэффициенты уравнения множественной линейной регрессии подбираются так, чтобы минимизировать сумму квадратов отклонения реальных точек данных от этой гиперплоскости.

Линейная регрессия — тщательно изученный метод регрессионного анализа. Его главное достоинство — простота. Такую модель можно построить и рассчитать даже без мощных вычислительных средств. Простота является и главным недостатком этого метода. Тем не менее, именно с линейной регрессии целесообразно начать подбор подходящей модели.

На языке Python линейная регрессия реализована в `sklearn.linear_model.LinearRegression()`.

1.1.4 Случайный лес

Случайный лес (RandomForest) — 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:

- Выбирается подвыборка обучающей выборки размера `samplesize` (может быть с возвращением) — по ней строится дерево (для каждого дерева — своя подвыборка).

- Для построения каждого расщепления в дереве просматриваем `max_features` случайных признаков (для каждого нового расщепления — свои случайные признаки).

- 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.

Если точность дерева решений оказалась недостаточной, множество моделей собирается в коллектив. Формула итогового решателя (3) — это усреднение предсказаний отдельных деревьев.

$$a(x) = \frac{1}{N} \sum_{i=1}^N b_i(x) \quad (3)$$

где N — количество деревьев;

i — счетчик для деревьев;

b — решающее дерево;

x — сгенерированная на основе данных выборка.

Для определения входных данных каждому дереву используется метод случайных подпространств. Базовые алгоритмы обучаются на различных подмножествах признаков, которые выделяются случайным образом.

Преимущества случайного леса:

- высокая точность предсказания;
- редко переобучается;
- практически не чувствителен к выбросам в данных;
- одинаково хорошо обрабатывает как непрерывные, так и дискретные признаки, данные с большим числом признаков;

- высокая параллелизуемость и масштабируемость.

Из недостатков можно отметить, что его построение занимает больше времени. Так же теряется интерпретируемость.

На языке Python метод случайного леса реализован в `sklearn.ensemble.RandomForestRegressor`.

1.1.5 Многослойный перцептрон

Нейронная сеть — это последовательность нейронов, соединенных между собой связями. Нейронные сети применяются для решения задач регрессии, классификации, распознавания образов и речи, компьютерного зрения и других. На настоящий момент это самый мощный, гибкий и широко применяемый инструмент в машинном обучении. Вычислительная единица нейронной сети — нейрон или перцептрон (персептрон). Перцептроны пытаются имитировать функциональность человеческого мозга для решения задач.

Многослойный перцептрон является первой и одной из наиболее простых архитектур нейронных сетей, с которых обычно начинается работа [1].

Его можно в какой-то мере сравнить с такой моделью машинного обучения, как линейная регрессия. Многослойный перцептрон уверенно работает с большим количеством входных параметров, но не в любых масштабах (для изображений, где речь идет о десятках миллионов входных значений пикселей, многослойный перцептрон работает очень плохо).

Скрытых слоев может быть достаточно много в многослойном перцептроне. Однако, существует несколько базовых архитектур. Поэтому возможно некоторое количество одинаковых нейронов на скрытых слоях и скрытых слоев тоже может быть некоторое количество (2, 3, 4, 5 и так далее). Аналогию можно провести с полиномиальной регрессией (один скрытый слой — это некоторая почти линейная аппроксимация — линейная регрессия).

Соответственно, количество слоев может быть по аналогии определено порядком полиномиальной регрессии.

При задаче регрессии на выходе будет один выходной слой (так как требуется получить итоговое значение).

На каждом скрытом слое может быть нейрон смещения, необходимый для того, чтобы добиваться нужного смещения для выходного значения.

Всего на многослойном перцептроне может быть четыре типа нейронов:

- 1) входные;
- 2) нейроны скрытых слоев;
- 3) выходные;
- 4) нейроны смещения (bias).

Одной из чуть менее распространенных структур многослойного перцептрона является пирамидка. Есть некоторое число входных скрытых нейронов на первом скрытом слое, затем это количество уменьшается, и за счет умножения нейронов (перемножения связей между ними) выделяются некоторые характерные особенности из данных. Количество связей с каждым скрытым слоем уменьшается. Это чуть менее распространенная архитектура, но в целом ничего не мешает исключить один или даже два скрытых слоя, заменив все небольшим количеством сужающих нейронов, когда через них пытаются выделить основные параметры из всего многообразия факторов и каким-то образом их суммировать. Для задачи регрессии используется обычно либо полулинейная функция активации, либо сигмоида.

Такая архитектура также позволяет учесть все многообразие факторов и высветить ровно те из них, которые могут повлиять на принятие решений.

На языке Python многослойный перцептрон реализован в `sklearn.neural_network.MLPRegressor()`.

1.1.5 Нейронная сеть на Keras

Нейронная сеть — это последовательность нейронов, соединенных между собой связями. Структура нейронной сети пришла в мир программирования из биологии. Вычислительная единица нейронной сети — нейрон или перцептрон.

У каждого нейрона есть определённое количество входов, куда поступают сигналы, которые суммируются с учётом значимости (веса) каждого входа.

Смещение — это дополнительный вход для нейрона, который всегда равен 1 и, следовательно, имеет собственный вес соединения.

Так же у нейрона есть функция активации, которая определяет выходное значение нейрона. Она используется для того, чтобы ввести нелинейность в нейронную сеть. Примеры активационных функций: relu, сигмоида.

У полносвязной нейросети выход каждого нейрона подается на вход всем нейронам следующего слоя. У нейросети имеется:

- входной слой — его размер соответствует входным параметрам;
- скрытые слои — их количество и размерность определяем специалист;
- выходной слой — его размер соответствует выходным параметрам.

Прямое распространение — это процесс передачи входных значений в нейронную сеть и получения выходных данных, которые называются прогнозируемым значением.

Прогнозируемое значение сравниваем с фактическим с помощью функции потерь. В методе обратного распространения ошибки градиенты (производные значений ошибок) вычисляются по значениям весов в направлении, обратном прямому распространению сигналов. Значение градиента вычитают из значения веса, чтобы уменьшить значение ошибки. Таким образом происходит процесс обучения. Обновляются веса каждого соединения, чтобы функция потерь минимизировалась.

Для обновления весов в модели используются различные оптимизаторы.

Количество эпох показывает, сколько раз выполнялся проход для всех примеров обучения.

Нейронные сети применяются для решения задач регрессии, классификации, распознавания образов и речи, компьютерного зрения и других. На настоящий момент это самый мощный, гибкий и широко применяемый инструмент в машинном обучении.

На языке Python нейронная сеть на Keras реализована в `tensorflow.keras.Sequential()`.

1.1.6 Метрики качества моделей

Существует множество различных метрик качества, применимых для регрессии. В этой работе используются:

R^2 (коэффициент детерминации) измеряет долю дисперсии, объясненную моделью, в общей дисперсии целевой переменной. Если он близок к единице, то модель хорошо объясняет данные, если же он близок к нулю, то качество прогноза идентично средней величине целевой переменной (то есть очень низкое). Отрицательные значения коэффициента детерминации означают плохую объясняющую способность модели.

MSE (MeanSquaredError) или средняя квадратичная ошибка принимает значения в тех же единицах, что и целевая переменная. Чем ближе к нулю MSE, тем лучше работают предсказательные качества модели.

1.2 Постановка задачи

В процессе выполнения квалификационной работы необходимо построить модели для прогнозирования некоторых механических характеристик композиционных материалов: модуля упругости при растяжении, прочности при растяжении, и создать нейронную сеть для рекомендации соотношения матрица-наполнитель композита.

Требуется:

1) Изучить теоретические основы и методы решения поставленной задачи.

2) Провести разведочный анализ предложенных данных. Необходимо нарисовать гистограммы распределения каждой из переменной, диаграммы ящика с усами, попарные графики рассеяния точек. Необходимо также для каждой колонки получить среднее, медианное значение, провести анализ и исключение выбросов, проверить наличие пропусков.

3) Провести предобработку данных (удаление шумов, нормализация и т.д.).

4) Обучить нескольких моделей для прогноза модуля упругости при растяжении и прочности при растяжении. При построении модели необходимо 30% данных оставить на тестирование модели, на остальных происходит обучение моделей. При построении моделей провести поиск гиперпараметров модели с помощью поиска по сетке с перекрестной проверкой, количество блоков равно 10.

5) Написать нейронную сеть, которая будет рекомендовать соотношение матрица-наполнитель.

6) Разработать приложение с графическим интерфейсом или интерфейсом командной строки, которое будет выдавать прогноз, полученный в задании 4 или 5 (один или два прогноза, на выбор учащегося).

7) Оценить точность модели на тренировочном и тестовом датасете.

8) Создать репозиторий в GitHub/GitLab и разместить там код исследования. Оформить файл README.

2.1 Препреподготовка данных

Процесс ETL (Extract, Transform, Load — Выбрать, Преобразовать, Загрузить) — процесс препреподготовки данных, когда необходимо понять, все ли готово для того, чтобы начинать работать с моделью [1].

К этапу ETL допустимо возвращаться по несколько раз по ходу оценки точности работы модели.

Процесс состоит из нескольких инструментов:

- получение данных из нескольких источников,
- очистку данных либо заполнение пропусков при существовании таковых,
- объединение данных из нескольких источников.

Исходные датасеты со свойствами композиционных материалов представлены Центром НТИ «Цифровое материаловедение: новые материалы и вещества» МГТУ им. Н.Э. Баумана.

Датасет состоит из двух файлов:

- файл X_bp.xlsx с данными о параметрах базальтопластика,
- файл X_nup.xlsx с данными о нашивках из углепластика.

Количество строк в файле X_bp.xlsx — 1023, столбцов 10. Количество строк в файле X_nup.xlsx — 1040, столбцов 3.

Необходимо было выполнить объединение данных двух файлов X_bp.xlsx и X_nup.xlsx по индексу, используя тип объединения INNER, которое было произведено методом merge библиотеки pandas, см. рисунок 1.

```
Ввод [8]: 1 # соединение объектов DataFrame, метод merge, объединение по индексу, тип объединения INNER
          2 df_s = x_bp.merge(x_nup, how = 'inner', left_index = True, right_index = True)
          3 df_s.head()

Out[8]:
```

Рисунок 1. Объединение исходный файлов датасета методом merge

После объединения получен датасет, который имеет характеристики, представленные на рисунке 2.

```

Ввод [9]: 1 # Вывод сведений о новом объекте DataFrame df_s, метод info()
          2 df_s.info()

<class 'pandas.core.frame.DataFrame'>
Float64Index: 1023 entries, 0.0 to 1022.0
Data columns (total 13 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   Соотношение матрица-наполнитель          1023 non-null   float64
1   Плотность, кг/м3                          1023 non-null   float64
2   модуль упругости, ГПа                     1023 non-null   float64
3   Количество отвердителя, м.%              1023 non-null   float64
4   Содержание эпоксидных групп,%_2          1023 non-null   float64
5   Температура вспышки, C_2                 1023 non-null   float64
6   Поверхностная плотность, г/м2            1023 non-null   float64
7   Модуль упругости при растяжении, ГПа     1023 non-null   float64
8   Прочность при растяжении, МПа            1023 non-null   float64
9   Потребление смолы, г/м2                  1023 non-null   float64
10  Угол нашивки, град                       1023 non-null   float64
11  Шаг нашивки                             1023 non-null   float64
12  Плотность нашивки                        1023 non-null   float64
dtypes: float64(13)
memory usage: 111.9 KB

```

Рисунок 2. Характеристики объединенного датасета

Датасет состоит из 1023 строк и 13 столбцов, который содержит в себе характеристики композиционных материалов. Тип данных всех столбцов — float64 — число с плавающей запятой, столбцы полностью заполнены non-null данными.

2.2 Исследовательский анализ данных

Процесс EDA (Exploratory Data Analysis — Исследовательский анализ данных) — разведка, которая позволяет существенным образом улучшить данные и определиться с наиболее важными параметрами [1].

Процесс исследовательского (разведовательного) анализа данных также состоит из нескольких инструментов:

- выбор некоторой предварительной модели или нахождение некоторой простой взаимосвязи в исходных данных,
- установление наличие пропусков и дубликатов в данных,
- улучшение исходных данных,
- построение ряда графиков.

На этом этапе необходимо определиться с наиболее важными параметрами, в том числе и за счет корреляции между этими параметрами и предсказыва-

емым значением. EDA выполняется единожды и его результаты используются в дальнейшем при построении всей модели обучения.

Было установлено, что пропуски и дубликаты в объединенном датасете отсутствуют, см. рисунок 3.

```

1 # Проверка наличия пропусков типа NULL
2 print(df_s.isnull().sum())

```

Соотношение матрица-наполнитель	0
Плотность, кг/м3	0
модуль упругости, ГПа	0
Количество отвердителя, м.%	0
Содержание эпоксидных групп,%_2	0
Температура вспышки, C_2	0
Поверхностная плотность, г/м2	0
Модуль упругости при растяжении, ГПа	0
Прочность при растяжении, МПа	0
Потребление смолы, г/м2	0
Угол нашивки, град	0
Шаг нашивки	0
Плотность нашивки	0

dtype: int64

```

1 # Проверка наличия дубликатов в объекте
2 df_s.duplicated().sum()

```

0

Рисунок 3. Проверка наличия пропусков и дубликатов в датасете

Следовательно, датасет был предварительно обработан.

```

1 # Проверка количества уникальных значений в
2 df_s.nunique()

```

Соотношение матрица-наполнитель	1014
Плотность, кг/м3	1013
модуль упругости, ГПа	1020
Количество отвердителя, м.%	1005
Содержание эпоксидных групп,%_2	1004
Температура вспышки, C_2	1003
Поверхностная плотность, г/м2	1004
Модуль упругости при растяжении, ГПа	1004
Прочность при растяжении, МПа	1004
Потребление смолы, г/м2	1003
Угол нашивки, град	2
Шаг нашивки	989
Плотность нашивки	988
Относительное удлинение при растяжении	1004
Относительное удлинение	1020

dtype: int64

Рисунок 4. Проверка количества уникальных значений в датасете

В столбце 'Угол нашивки, град' находятся два уникальных значения, в остальных столбцах таковых — от 989 до 1020 значений. Параметр 'Угол нашивки, град' следует рассматривать как категориальный признак и для дальней-

шей обработки он был проранжирован в два ранга: 0 и 1, остальные параметры следует рассматривать как количественные непрерывные.

На этом же этапе было осуществлено преобразование индекса датасета в тип данных `int32` — тип данных целое число, а также в датасет были добавлены два столбца, дополняющие характеристики композитов — 'Относительное удлинение при растяжении' и 'Относительное удлинение', см. рисунок 5.

Датасет представляет собой результаты испытаний на растяжение композитных материалов, тем не менее, данные, которые относятся к результатам замера геометрических параметров образцов, в нем отсутствуют. Восполним этот пробел и создадим соответствующие столбцы:

```
1 # Относительное удлинение при растяжении без нормирования
2 df_s['Относительное удлинение при растяжении'] = df_s['Прочность при растяжении, МПа'] / df_s['Модуль упругости при растяжении, ГПа']

1 # относительное удлинение без нормирования
2 df_s['Относительное удлинение'] = df_s['Прочность при растяжении, МПа'] / df_s['модуль упругости, ГПа']

1 df_s
```

Рисунок 5. Создание новых столбцов в датасете

Методом `sns.pairplot` построены попарные графики рассеяния, которые представлены на рисунке 6.

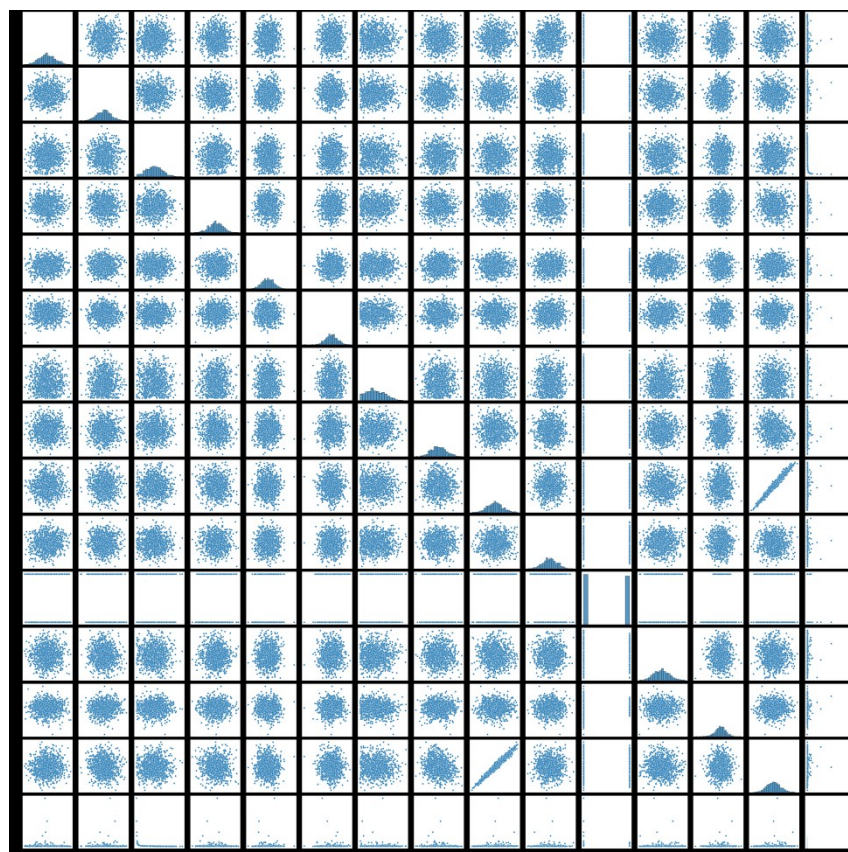


Рисунок 6 . Попарные графики рассеяния параметров

Попарные графики рассеяния в целом не показывают какой-либо зависимости между параметрами, имеют вид случайных данных с нормальным распределением, кроме данных столбца 'Угол нашивки', который является категориальной переменной и имеет два ранжированных уровня: 0 и 1, а также столбцов 'Относительное удлинение при растяжении' и 'Относительное удлинение'. Столбец 'Относительное удлинение при растяжении' показывает качественную линейную связь с некоторым рассеянием со столбцом 'Прочность при растяжении, МПа', что всегда подтверждается соотношениями между физико-механическими характеристиками материалов [2] и приводит к пониманию, что между столбцами 'Прочность при растяжении, МПа' и 'Модуль упругости при растяжении, ГПа' должна тоже существовать связь, что не подтверждается попарными графиками рассеяния параметров. Возможно, что параметры 'Прочность при растяжении, МПа' и 'Модуль упругости при растяжении, ГПа' в значительной степени включают в себя данные, сгенерированные таким образом, что обе компоненты имеют нормальное распределение с увязанными между собой математическим ожиданием и стандартным отклонением [3, 4].

Диаграмма размаха «ящик с усами» (boxplot) показывает одномерное распределение вероятностей параметров, фрагмент диаграммы для параметров 'Соотношение матрица-наполнитель' и 'Плотность, кг/м³', см. рисунок 7.

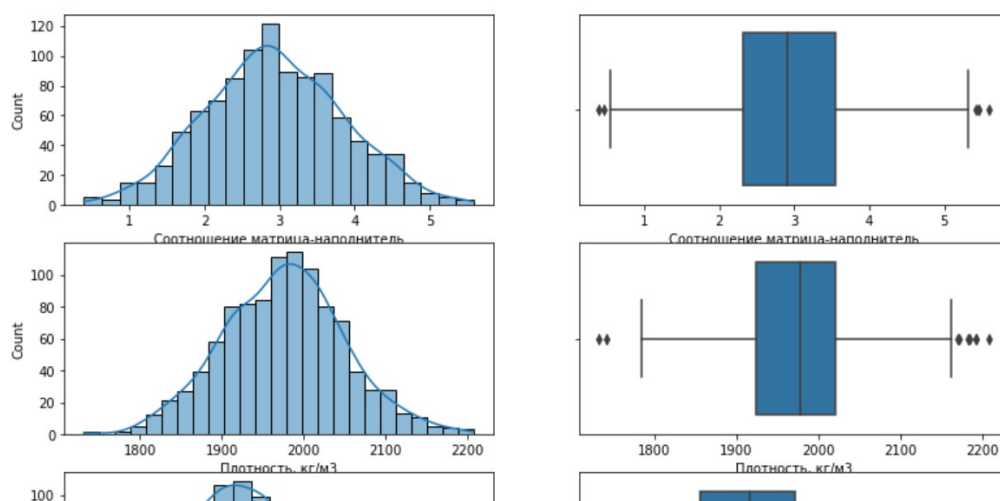


Рисунок 7. Диаграмма размаха «ящик с усами»

Диаграмма «ящик с усами» показывает медиану, нижний и верхний квартили, минимальное и максимальное значение выборки и выбросы по каждому параметру.

Судя по данным графикам, распределения всех параметров, кроме 'модуль упругости, ГПа', 'Поверхностная плотность, г/м²', 'Угол нашивки' и 'Относительное удлинение', близки к нормальным и имеются выбросы по всем параметрам, кроме 'Угол нашивки, град', который является ранговым.

Связь между параметрами следует оценить с помощью коэффициента корреляции и статистической оценкой значимости этого коэффициента.

Предварительно следует оценить нормальность распределения каждого параметра датасета, что осуществляется путем расчета описательной статистики, а также путем проверки статистических гипотез оценка нормальности распределения параметров по статистическому критерию д'Агостино и по статистическому критерию Шапиро-Уилка. Критерий Шапиро-Уилка является статистически более мощным и показывает, что параметры 'модуль упругости, ГПа', 'Поверхностная плотность, г/м²', 'Плотность нашивки', 'Относительное удлинение' имеют распределение статистически отличное от нормального распределения. Следовательно оценка корреляции и оценка значимости такой корреляции параметров с таким распределением, а также рангового параметра 'Угол нашивки, град' должна осуществляться с помощью непараметрического критерия Спирмена. Для упрощения исследования попарная корреляция рассчитывается между всеми параметрами сразу по обоим критериям — Пирсона и Спирмена, затем отбираются необходимые для исследования попарные корреляции с учетом того подчиняется ли распределение параметра нормальному закону распределения.

Описательная статистика показывает, что по всем столбцам, кроме 'Относительное удлинение', средние значения укладываются в доверительные интервалы для средних, мода смещена для некоторых параметров, стандартное отклонение имеет порядок ниже размаха, асимметрия и эксцесс близки к нормальным (близки к 0), см. Приложение 1. Создается впечатление в целом нормального распределения всех параметров. Оценка эксцесса в `scipy.stats.kurtosis` рассчиты-

вается по формуле, из которой уже вычитается 3, чтобы упростить сравнение с нормальным распределением.

Ниже на рисунке 8 представлена визуализация корреляции с помощью тепловой карты `sns.heatmap`. Тепловая карта на рисунке 8 построена без включения столбцов 'Относительное удлинение при растяжении' и 'Относительное удлинение', на рисунке 9 представлена тепловая карта с включением столбцов 'Относительное удлинение при растяжении' и 'Относительное удлинение'.

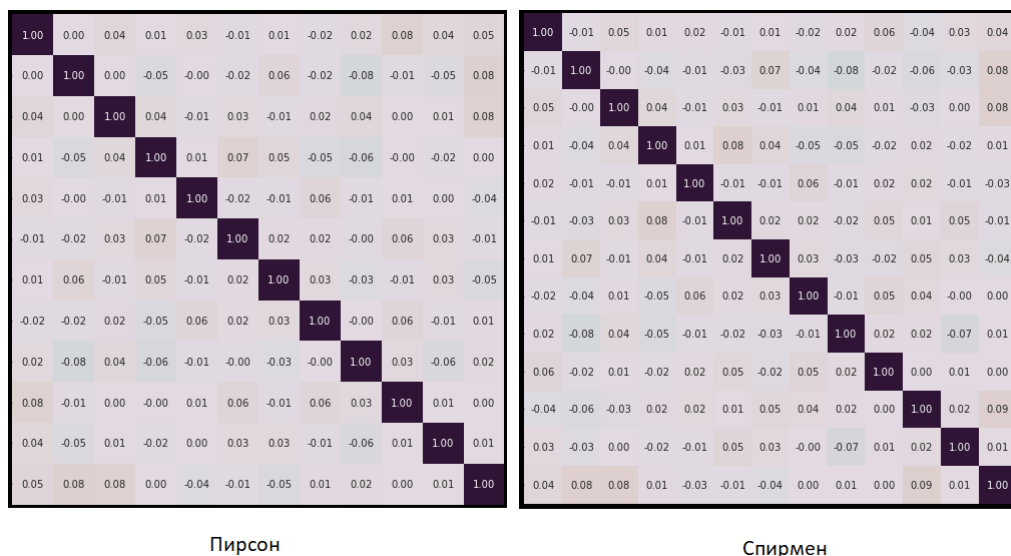


Рисунок 8. Матрицы корреляции по Пирсону и Спирмену без столбцов 'Относительное удлинение при растяжении' и 'Относительное удлинение'

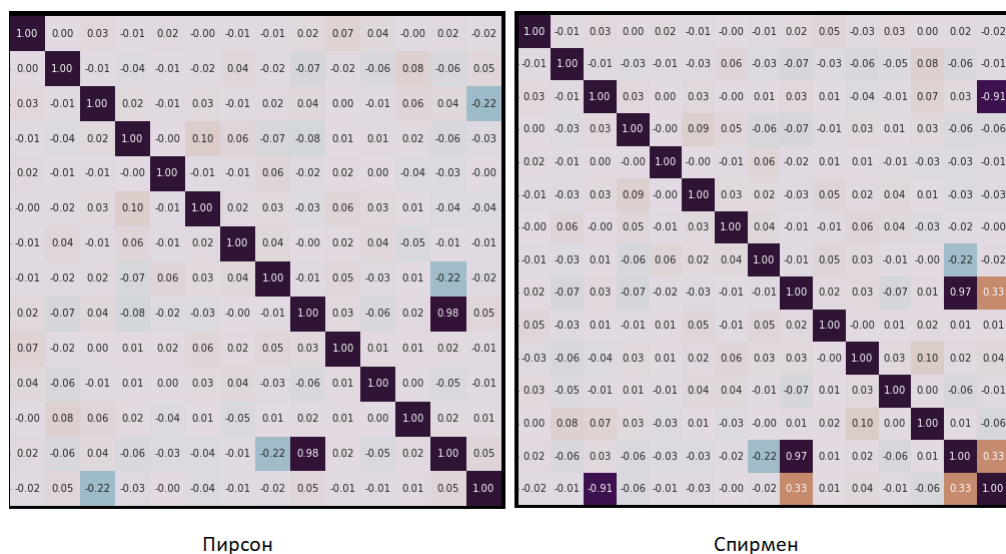


Рисунок 9. Матрицы корреляции по Пирсону и Спирмену с включением столбцов 'Относительное удлинение при растяжении' и 'Относительное удлинение'

Ранговый критерий Спирмена позволяет выявить корреляции между параметрами 'Относительное удлинение при растяжении' и 'Относительное удлинение', а также 'Прочность при растяжении, МПа' и 'Относительное удлинение', хотя между параметрами 'модуль упругости, ГПа' и 'Модуль упругости при растяжении, ГПа' связь не наблюдается. Это приводит к пониманию, что подобная связь может быть выявлена после очистки или другой обработки датасета.

Статистически значимыми по критерию Пирсона оказываются следующие попарные корреляции между количественными параметрами:

Соотношение матрица-наполнитель & Потребление смолы, г/м2 : Statistics=0.073, p-value=0.020
Плотность, кг/м3 & Прочность при растяжении, МПа : Statistics=-0.070, p-value=0.025
Плотность, кг/м3 & Относительное удлинение при растяжении : Statistics=-0.062, p-value=0.046
Количество отвердителя, м.% & Температура вспышки, C_2 : Statistics=0.095, p-value=0.002
Количество отвердителя, м.% & Модуль упругости при растяжении, ГПа : Statistics=-0.066, p-value=0.035
Количество отвердителя, м.% & Прочность при растяжении, МПа : Statistics=-0.075, p-value=0.016
Модуль упругости при растяжении, ГПа & Относительное удлинение при растяжении : Statistics=-0.216, p-value=0.000
Прочность при растяжении, МПа & Относительное удлинение при растяжении : Statistics=0.977, p-value=0.000

Для параметров 'модуль упругости, ГПа', 'Поверхностная плотность', 'Плотность нашивки' и 'Относительное удлинение', распределение которых отличается от нормального, проверка по критерию Спирмена показала, что значимы следующие корреляции:

Плотность, кг/м3 & Плотность нашивки : Statistics=0.080, p-value=0.010
модуль упругости, ГПа & Плотность нашивки : Statistics=0.073, p-value=0.020
модуль упругости, ГПа & Относительное удлинение : Statistics=-0.911, p-value=0.000
Прочность при растяжении, МПа & Относительное удлинение : Statistics=0.332, p-value=0.000
Плотность нашивки & Относительное удлинение : Statistics=-0.063, p-value=0.043
Относительное удлинение при растяжении & Относительное удлинение : Statistics=0.327, p-value=0.000

Также значима связь параметров параметрами: 'Угол нашивки, град' и 'Плотность, кг/м3', а также 'Угол нашивки, град' и 'Плотность нашивки'.

Значимости связи между параметрами 'модуль упругости, ГПа' и 'Модуль упругости при растяжении, ГПа' также не наблюдается, что не соответствует соотношениям между физико-механическими характеристиками материалов [2]. Кроме того, не наблюдается значимой связи между параметрами 'Прочность при растяжении, МПа' и 'Модуль упругости при растяжении, ГПа', что также не соответствует физико-механическим свойствам материалов.

При анализе данных корреляции могут быть обманчивыми, если игнорируются спутывающие переменные [5, 6]. Главной проблемой является то, что

корреляция измеряет связь между двумя переменными при прочих равных условиях. Единственный реальный способ избежать этой неприятности — это знать свои данные и постараться проверить возможные спутывающие факторы.

Не удалось выяснить в общем датасете также заметные связи между данными из двух разных файлов 'X_br.xlsx' и 'X_nur.xlsx'. Соответствие между данными из этих двух файлов, вероятно, является произвольным.

Сделать какое-либо уверенное заключение о корреляции между параметрами рассматриваемого датасета на данном этапе анализа не представляется возможным.

Следует также произвести статистическое сравнение двух групп данных датасета, представленных в столбце 'Угол нашивки, град', который принимает два значения 0 и 90 градусов. Понимание того, что нашивка может располагаться под углом 0 и 90 градусов к образцу должна давать различие значений различных параметров для этих двух групп. Подобная проверка может быть произведена по непараметрическому критерию Колмогорова-Смирнова, предназначенному для сравнения двух групп.

Критерий Колмогорова-Смирнова сравнения двух групп параметра 'Угол нашивки, град' (углы нашивки имеют ранги 0 и 1) показывает, что не отвергается статистически значимое различие этих двух групп для параметров 'модуль упругости, ГПа' и 'Плотность нашивки', см. ниже рисунок 10.

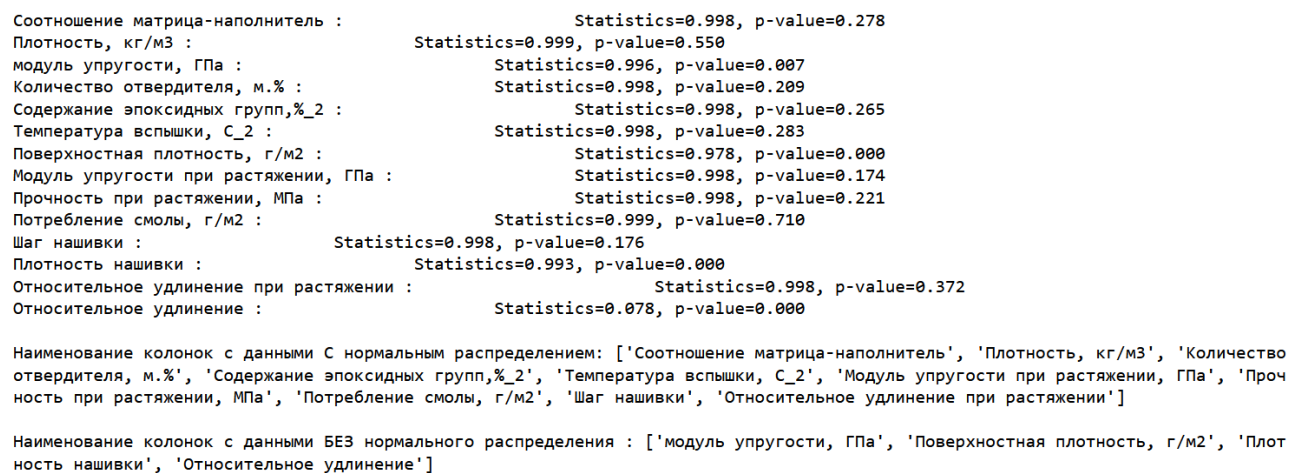


Рисунок 10. Заключение по критерию Колмогорова-Смирнова

Однофакторный анализ для оценки различия средних всех параметров по двум группам угла нашивки (углы нашивки с рангами 0 и 1) позволил выявить значимое различие дисперсий только для параметра 'Плотность, кг/м³', см. ниже рисунок 11.

Модуль упругости при растяжении, ГПа :	Statistics Bartlett=1.025, p-value=0.311
Модуль упругости при растяжении, ГПа :	Statistics F-test=0.541, p-value=0.462
Прочность при растяжении, МПа :	Statistics Bartlett=0.460, p-value=0.497
Прочность при растяжении, МПа :	Statistics F-test=0.559, p-value=0.455
Потребление смолы, г/м ² :	Statistics Bartlett=0.072, p-value=0.789
Потребление смолы, г/м ² :	Statistics F-test=0.240, p-value=0.624
Шаг нашивки :	Statistics Bartlett=15.555, p-value=0.000
Шаг нашивки :	Statistics F-test=0.570, p-value=0.451
Плотность нашивки :	Statistics Bartlett=188.329, p-value=0.000
Плотность нашивки :	Statistics F-test=12.038, p-value=0.001
Относительное удлинение при растяжении :	Statistics Bartlett=0.830, p-value=0.362
Относительное удлинение при растяжении :	Statistics F-test=0.280, p-value=0.597
Относительное удлинение :	Statistics Bartlett=1223.499, p-value=0.000
Относительное удлинение :	Statistics F-test=2.579, p-value=0.109

Вывод: различаются средние для 'Плотность, кг/м³', 'Плотность нашивки'. Но Bartlett-test для 'Плотность нашивки' показывает, что группы имеют не равные дисперсии для плотности нашивки, поэтому F-test не может быть применен для 'Плотность нашивки'

Рисунок 11. Заключение по однофакторному анализу различия средних по двум группам параметра 'Угол нашивки, град'

Создается впечатление малого влияния параметра 'Угол нашивки, град' на датасет в целом, что тоже не соответствует априорным представлениям о физико-механических свойствах образцов.

Предварительное заключение по датасету: данные датасета представляют собой в значительной мере данные с нормальным распределением, кроме параметра 'Угол нашивки, град' принимающего два значения, и параметров 'модуль упругости, ГПа', 'Поверхностная плотность, г/м²', 'Плотность нашивки', 'Относительное удлинение', имеющих распределение статистически отличное от нормального распределения. Не удалось выявить в объединенном датасете также связи между данными из двух разных файлов 'X_br.xlsx' и 'X_nur.xlsx'. Соответствие между данными из этих двух файлов, вероятно, является произвольным. Вероятно также, что последовательности 'Прочность при растяжении, МПа' и 'Модуль упругости при растяжении, ГПа' в значительной степени включают в себя данные сгенерированные по нормальному закону распределения таким образом, что обе компоненты имеют нормальное распределение с увязанными между собой математическим ожиданием и стандартным отклонением параметров, на что указывает попарный график рассеяния [3, 4].

Первый вопрос на который предстоит ответить заключается в том, содержит ли датасет значимые данные?

Диаграмма «ящик с усами», см. рисунок 7 показала наличие выбросов по всем параметрам, кроме 'Угол нашивки, град'. Очистка данных от выбросов производилась методами: 1) трех сигм и 2) межквартильного расстояния.

Было установлено, что имеются следующее количество выбросов: для метода трех сигм — 24 выброса, для метода межквартильных расстояний — 93 выброса.

Оценка нормальности распределения параметров по статистическому критерию Шапиро-Уилка показала, что имеют вид распределения статистически отличный от нормального следующие параметры:

1) для случая очистки методом трех сигм — 'модуль упругости, ГПа', 'Поверхностная плотность, г/м²', 'Угол нашивки, град', 'Относительное удлинение';

2) для случая очистки методом межквартильного расстояния — 'модуль упругости, ГПа', 'Поверхностная плотность, г/м²', 'Модуль упругости при растяжении, ГПа', 'Потребление смолы, г/м²', 'Угол нашивки, град', 'Относительное удлинение';

3) для случая без очистки — 'модуль упругости, ГПа', 'Поверхностная плотность, г/м²', 'Плотность нашивки', 'Угол нашивки, град', 'Относительное удлинение'.

Имеет смысл сравнить тепловые карты корреляции после очистки данных от выбросов. На рисунке 12 представлена тепловая карта с включением столбцов 'Относительное удлинение при растяжении' и 'Относительное удлинение' для случая очистки методом трех сигм, а на рисунке 13 представлена тепловая карта с включением столбцов 'Относительное удлинение при растяжении' и 'Относительное удлинение' для случая очистки методом межквартильного расстояния.

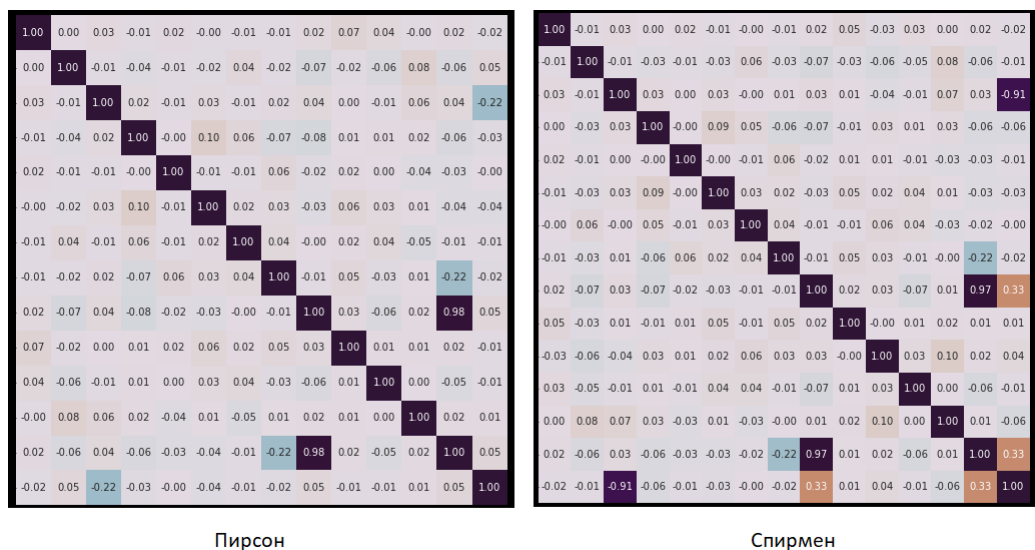


Рисунок 12. Матрицы корреляции по Пирсону и Спирмену с включением столбцов 'Относительное удлинение при растяжении' и 'Относительное удлинение' после очистки методом трех сигм

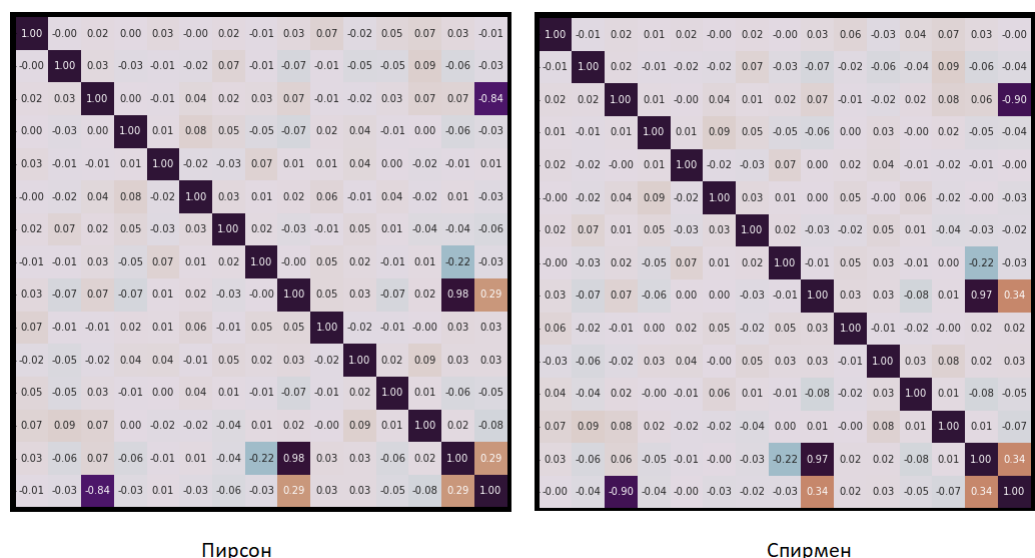


Рисунок 13. Матрицы корреляции по Пирсону и Спирмену с включением столбцов 'Относительное удлинение при растяжении' и 'Относительное удлинение' после очистки методом межквартильного расстояния

По сравнению с тепловой картой датасета для случая до очистки от выбросов картина не особо изменилась, см. также рисунки 8 и 9. Следует оценить значимость корреляций по критериям Пирсона и Спирмена:

1) для случая очистки методом трех сигм статистически значимыми по критерию Пирсона являются следующие попарные корреляции между количественными параметрами:

Плотность, кг/м³ & Прочность при растяжении, МПа : Statistics=-0.076, p-value=0.016
Плотность, кг/м³ & Плотность нашивки : Statistics=0.092, p-value=0.004
Плотность, кг/м³ & Относительное удлинение при растяжении : Statistics=-0.068, p-value=0.031
Количество отвердителя, м.% & Температура вспышки, C_2 : Statistics=0.087, p-value=0.006
Количество отвердителя, м.% & Модуль упругости при растяжении, ГПа : Statistics=-0.077, p-value=0.015
Содержание эпоксидных групп, %_2 & Модуль упругости при растяжении, ГПа : Statistics=0.069, p-value=0.030
Модуль упругости при растяжении, ГПа & Относительное удлинение при растяжении : Statistics=-0.206, p-value=0.000
Прочность при растяжении, МПа & Относительное удлинение при растяжении : Statistics=0.977, p-value=0.000

Для параметров 'модуль упругости, ГПа', 'Поверхностная плотность, г/м²', 'Относительное удлинение' проверка по критерию Спирмена показала, что значимы следующие корреляции:

модуль упругости, ГПа & Плотность нашивки : Statistics=0.080, p-value=0.012
модуль упругости, ГПа & Относительное удлинение : Statistics=-0.912, p-value=0.000
Прочность при растяжении, МПа & Относительное удлинение : Statistics=0.334, p-value=0.000
Плотность нашивки & Относительное удлинение : Statistics=-0.071, p-value=0.024
Относительное удлинение при растяжении & Относительное удлинение : Statistics=0.328, p-value=0.000

Кроме того по Спирмену значима корреляция параметров 'Угол нашивки, град' и 'Плотность, кг/м³', а также 'Угол нашивки, град' и 'Плотность нашивки'.

2) для случая очистки методом межквартильного расстояния — статистически значимыми по критерию Пирсона являются следующие попарные корреляции между количественными параметрами:

Соотношение матрица-наполнитель & Потребление смолы, г/м² : Statistics=0.075, p-value=0.028
Соотношение матрица-наполнитель & Плотность нашивки : Statistics=0.068, p-value=0.047
Плотность, кг/м³ & Прочность при растяжении, МПа : Statistics=-0.069, p-value=0.042
Плотность, кг/м³ & Плотность нашивки : Statistics=0.089, p-value=0.009
Количество отвердителя, м.% & Температура вспышки, C_2 : Statistics=0.079, p-value=0.021
Количество отвердителя, м.% & Прочность при растяжении, МПа : Statistics=-0.071, p-value=0.039
Прочность при растяжении, МПа & Шаг нашивки : Statistics=-0.068, p-value=0.048
Прочность при растяжении, МПа & Относительное удлинение при растяжении : Statistics=0.976, p-value=0.000

Для параметров 'модуль упругости, ГПа', 'Поверхностная плотность, г/м²', 'Относительное удлинение' проверка по критерию Спирмена показала, что значимы следующие корреляции:

Плотность, кг/м³ & Поверхностная плотность, г/м² : Statistics=0.068, p-value=0.045
модуль упругости, ГПа & Прочность при растяжении, МПа : Statistics=0.072, p-value=0.034
модуль упругости, ГПа & Плотность нашивки : Statistics=0.078, p-value=0.023
модуль упругости, ГПа & Относительное удлинение : Statistics=-0.896, p-value=0.000
Прочность при растяжении, МПа & Относительное удлинение : Statistics=0.339, p-value=0.000
Плотность нашивки & Относительное удлинение : Statistics=-0.069, p-value=0.043
Относительное удлинение при растяжении & Относительное удлинение : Statistics=0.338, p-value=0.000

Кроме того по Спирмену значима корреляция параметров 'Угол нашивки, град' и 'Плотность нашивки'.

Вывод по модели корреляции: значимость коэффициентов корреляции после очистки параметров от выбросов меняется мало предсказуемо, кроме как

для параметра 'модуль упругости, ГПа', который нелинейно детерминировано связан с параметром 'Относительное удлинение' и параметра 'Прочность при растяжении, МПа', который линейно детерминировано связан с параметром 'Относительное удлинение при растяжении'. Представляет интерес значимость связи между 'модуль упругости, ГПа' & 'Прочность при растяжении, МПа'. При изучении корреляции параметров без 'Относительное удлинение при растяжении' и 'Относительное удлинение' корреляция и ее значимость между параметрами 'модуль упругости, ГПа' & 'Прочность при растяжении, МПа' не наблюдается.

Результаты расчета значимости корреляций для случая очистки методом межквартильного расстояния датасета без параметров 'Относительное удлинение при растяжении' и 'Относительное удлинение' приведены в Приложении 2.

Также были рассмотрены модели множественной линейной регрессии до очистки и после очистки датасета методами трех сигм и межквартильного расстояния. Для краткости изложения и удобства расчетов названия параметров в датасете были заменены следующими названиями:

'Соотношение матрица-наполнитель'	'col1'
'Плотность, кг/м3'	'col2'
'модуль упругости, ГПа'	'col3'
'Количество отвердителя, м. %'	'col4'
'Содержание эпоксидных групп, %_2'	'col5'
'Температура вспышки, C_2'	'col6'
'Поверхностная плотность, г/м2'	'col7'
'Модуль упругости при растяжении, ГПа'	'col8'
'Прочность при растяжении, МПа'	'col9'
'Потребление смолы, г/м2'	'col10'
'Угол нашивки, град'	'col11'
'Шаг нашивки'	'col12'
'Плотность нашивки'	'col13'
'Относительное удлинение при растяжении'	'col14'
'Относительное удлинение'	'col15'

Параметр 'Угол нашивки, град' является категориальным и был преобразован в порядковый с помощью метода `pd.concat`, аргумент `get_dummies`, см. ниже рисунок 14.

1	# Замена категориальных переменных порядковыми переменными, метод concat,														
2	# Dummy variables, аргумент get_dummies														
3	ff = pd.concat((ff,pd.get_dummies(ff.col11, prefix="col11", prefix_sep=''),),1)														
4	ff.head()														
	col1	col2	col3	col4	col5	col6	col7	col8	col9	col10	col11	col12	col13	col110	col111
1	1.857143	2030.0	738.736842	50.00	23.750000	284.615385	210.0	70.0	3000.0	220.0	0	4.0	60.0	1	0
3	1.857143	2030.0	738.736842	129.00	21.250000	300.000000	210.0	70.0	3000.0	220.0	0	5.0	47.0	1	0
4	2.771331	2030.0	753.000000	111.86	22.267857	284.615385	210.0	70.0	3000.0	220.0	0	5.0	57.0	1	0
5	2.767918	2000.0	748.000000	111.86	22.267857	284.615385	210.0	70.0	3000.0	220.0	0	5.0	60.0	1	0
6	2.569620	1910.0	807.000000	111.86	22.267857	284.615385	210.0	70.0	3000.0	220.0	0	5.0	70.0	1	0

Рисунок 14. Преобразование параметра 'Угол нашивки, град'

Результаты расчетов были сведены в таблицу 1.

Таблица 1. Модели регрессии

Датасет	Адекватность модели	Значимость параметров модели
Неочищенный датасет, линейные модели	col8 — полная модель 'col8 ~ col1 + col2 + col3 + col4 + col5 + col6 + col7 + col9 + col10 + col12 + col13 + col110', неадекватность модели по F-критерию, R-squared = 0,016	col4 — параметр значим, остальные параметры не значимы
	col8 — неполная модель 'col8 ~ col4', адекватность модели по F-критерию, R-squared = 0,004	col4 — параметр значим, остальные параметры не значимы
	col9 — полная модель 'col9 ~ col1 + col2 + col3 + col4 + col5 + col6 + col7 + col8 + col10 + col12 + col13 + col110', адекватность модели по F-критерию, R-squared = 0,021	col2 + col4 + col12 — параметры значимые, остальные параметры не значимы
	col9 — неполная модель 'col9 ~ col2 + col4 + col12', адекватность модели по F-критерию, R-squared = 0,015	col2 + col4 + col12 — параметры значимые
	col9 — полная модель 'col9 ~ col1 + col2 + col4 + col5 + col6 + col7 + col10 + col12 + col13 + col110 + col14 + col15', адекватность модели по F-критерию, R-squared = 0,956	col4, col10, col14 — параметры значимые, остальные параметры не значимы
	col9 — неполная модель 'col9 ~ col4 + col10 + col14', адекватность модели по F-критерию, R-squared = 0,956	col4, col10, col14 — параметры значимые
Очищенный, метод 3-х сигм, линейные модели	col8 — полная модель 'col8 ~ col1 + col2 + col3 + col4 + col5 + col6 + col7 + col9 + col10 + col12 + col13 + col110', неадекватность модели по F-критерию, R-squared = 0,020	col4 + col5 — параметры значимые, остальные параметры не значимы
	col8 — неполная модель 'col8 ~ col4 + col5', адекватность модели по F-критерию, R-squared = 0,011	col4 + col5 — параметры значимы
	col9 — полная модель 'col9 ~ col1 + col2 +	col2, col12 — параметры

Датасет	Адекватность модели	Значимость параметров модели
	col3 + col4 + col5 + col6 + col7 + col8 + col10 + col12 + col13 + col110', неадекватность модели по F-критерию, R-squared = 0,019	значимые, остальные параметры не значимы
	col9 — неполная модель 'col9 ~ col2 + col4 + col12', адекватность модели по F-критерию, R-squared = 0,013	col2 — параметр значим, col4, col12 — параметры не значимые
	col9 — неполная модель 'col9 ~ col2', адекватность модели по F-критерию, R-squared = 0,006	col2 — параметр значим
	col9 — полная модель 'col9 ~ col1 + col2 + col4 + col5 + col6 + col7 + col10 + col12 + col13 + col110 + col14 + col15', адекватность модели по F-критерию, R-squared = 0,956	'col4', 'col5', 'col10' и 'col14' — параметры значимые, остальные параметры не значимы
	col9 — неполная модель 'col9 ~ col4 + col5 + col10 + col14', адекватность модели по F-критерию, R-squared = 0,956	'col4', 'col5', 'col10' и 'col14' — параметры значимые
Очищенный, 1-ая и 4-ая квартиль, линейные модели	col8 — полная модель 'col8 ~ col1 + col2 + col3 + col4 + col5 + col6 + col7 + col9 + col10 + col12 + col13 + col110', неадекватность модели по F-критерию, R-squared = 0,013	все параметры не значимы
	col8 — неполная модель 'col8 ~ col5', неадекватность модели по F-критерию, R-squared = 0,003	col5 — параметр не значим
	col9 — полная модель 'col9 ~ col1 + col2 + col3 + col4 + col5 + col6 + col7 + col8 + col10 + col12 + col13 + col110', неадекватность модели по F-критерию, R-squared = 0,019	col2, col4, col12 — параметры значимы
	col9 — неполная модель 'col9 ~ col2 + col4 + col12', адекватность модели по F-критерию, R-squared = 0,015	col2, col4, col12 — параметры значимы
	col9 — полная модель 'col9 ~ col1 + col2 + col4 + col5 + col6 + col7 + col10 + col12 + col13 + col110 + col14 + col15', адекватность модели по F-критерию, R-squared = 0,953	'col5' и 'col14' — параметры значимые, остальные параметры не значимы
	col9 — неполная модель 'col9 ~ col5 + col14', адекватность модели по F-критерию, R-squared = 0,952	'col5' и 'col14' — параметры значимые
	col3 — неполная модель 'col3 ~ col9', адекватность модели по F-критерию, R-squared = 0,005	'col9' — параметр значимый
	col3 — неполная модель 'col3 ~ col8', неадекватность модели по F-критерию, R-squared = 0,001	'col8' — параметр не значим
Очищенный, 1-ая и 4-ая квартиль, нелинейные степенные модели $\log y = \log a +$	col3 — неполная модель 'col3 ~ col1 + col2 + col4 + col5 + col6 + col7 + col8 + col9 + col10 + col12 + col13 + col15', адекватность модели по F-критерию, R-squared = 1,000	'col9' и 'col15' — параметры значимые, остальные параметры не значимы
	col3 — неполная модель 'col3 ~ col8',	'col8' — параметр не значим

Датасет	Адекватность модели	Значимость параметров модели
$+ b \log x + \log \varepsilon$	неадекватность модели по F-критерию, $R\text{-squared} = 0,001$	значим
	col3 — неполная модель 'col3 ~ col15', адекватность модели по F-критерию, $R\text{-squared} = 0,793$	'col15' — параметр значимый
	col3 — неполная модель 'col3 ~ col9', адекватность модели по F-критерию, $R\text{-squared} = 0,011$	'col9' — параметр значимый
	col3 — неполная модель 'col3 ~ col9 + col15', адекватность модели по F-критерию, $R\text{-squared} = 1,000$	'col9' и 'col15' — параметры значимые

Примечание: следует рассмотреть нелинейные степенные модели типа $\log y = \log a + b \log x + \log \varepsilon$ для параметра зависимости 'модуль упругости, ГПа' от 'Относительное удлинение' поскольку эта зависимость имеет явно нелинейный характер после очистки методом квартилей, см. рисунок 15 ниже.

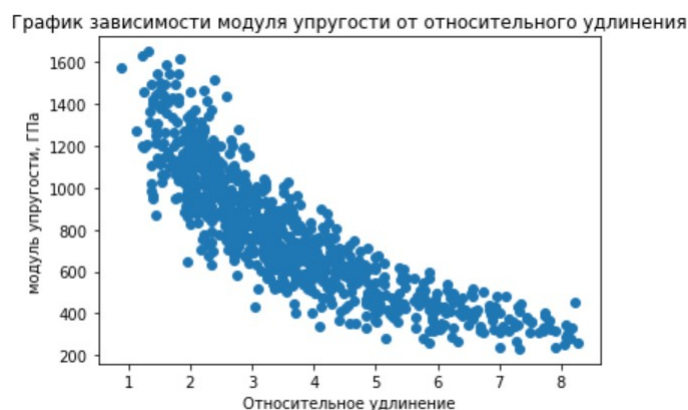


Рисунок 15. Зависимость модуля упругости от относительного удлинения

Например, нелинейная степенная регрессионная модель вида 'col3 ~ col8' представляет собой регрессию

$\log(\text{'модуль упругости, ГПа'}) = \log a + b \log(\text{'Относительное удлинение'}) + \log \varepsilon$,
где $\log a = 7.5947$, $b = -0.8369$.

Вывод по моделям регрессии: параметр 'Модуль упругости при растяжении, ГПа' может иметь линейную регрессионную связь с параметрами 'Количество отвердителя, м.%' и 'Содержание эпоксидных групп, %_2'; параметр 'Прочность при растяжении, МПа' может иметь линейную регрессионную связь с параметрами 'Плотность, кг/м3', 'Количество отвердителя, м.%', 'Содержание

эпоксидных групп, %₂', 'Потребление смолы, г/м²', 'Шаг нашивки' и 'Относительное удлинение при растяжении'; параметр 'модуль упругости, ГПа' имеет линейную и нелинейную степенную регрессионную связь 'Прочность при растяжении, МПа' и 'Относительное удлинение' по условию формирования параметра 'Относительное удлинение'. Параметр 'модуль упругости, ГПа' не имеет регрессионной связи с параметром 'Модуль упругости при растяжении, ГПа', что не соответствует представлениям о физико-механических характеристиках материалов.

Выше при изучении корреляционных связей отмечалось, что существует значимость корреляционной связи между 'модуль упругости, ГПа' & 'Прочность при растяжении, МПа'.

Таким образом, датасет, вероятно, содержит некоторое количество значимых данных, которые зашумлены незначимыми данными. Исходя из этого соображения было принято решение не очищать исходный датасет от выбросов с целью максимального сохранения тех данных, которые в нем находятся.

Второй вопрос заключается в том, в какой части датасета находятся эти значимые данные и сколько их всего?

Дальнейшая обработка датасета может быть проведена согласно следующим стратегиям:

- 1) осуществлением предобработки (нормализации) всего датасета и построением моделей машинного обучения для всего датасета: а) с включением параметров 'Относительное удлинение при растяжении' и 'Относительное удлинение' в модель, б) без включения таковых;

- 2) осуществлением предобработки (нормализации) всего датасета, проведением кластеризации датасета, выделением частей со значимыми корреляционными связями и построением для этих частей моделей машинного обучения;

- 3) делением датасета на части (срезы), выделением частей со значимыми корреляционными связями, выделением значимых данных, оценкой этих данных, дополнением данных с помощью аппроксимационных функций с целью получения необходимого количества данных, проверкой корреляционных связей и построением моделей машинного обучения для нового датасета.

2.3 Предобработка данных

Предобработка (препроцессинг) данных является логическим продолжением процесса ETL, и в большинстве случаев его даже объединяют с этим процессом, однако, принято разделять эти два понятия, потому что в процессе ETL нужно получить чистые данные (заполнить пропуски, найти опечатки, объединить все данные в источники), на этапе подготовки данных эти данные уже соответствующим образом обрабатываются, преобразовываются и переформатируются, чтобы их можно было подать на вход модели машинного обучения. И именно это форматирование сильно зависит от используемой модели машинного обучения [1].

Категориальный параметр 'Угол нашивки, град' был уже ранее преобразован в порядковый (ранговый) тип данных.

На этом этапе обработки нормализация значений параметров осуществляется с помощью метода MinMaxScaler. Соответствующим образом меняются данные в датасете.

Матрицы корреляции по Пирсону для количественных данных и по Спирмену для количественных и категориальных данных представлены на рисунках 16 и 17, сравнить с рисунками 8 и 9.

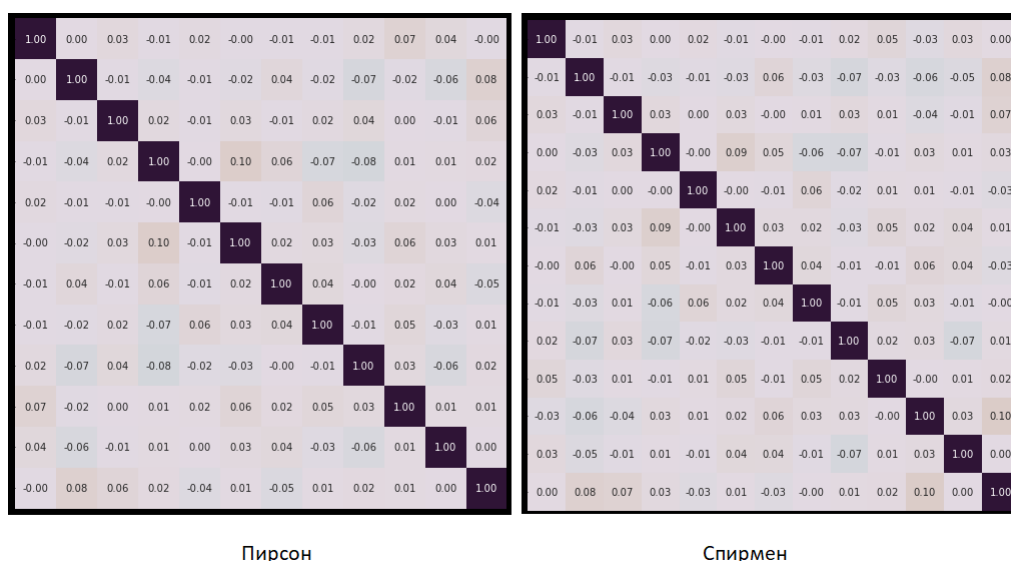


Рисунок 16. Матрицы корреляции по Пирсону и Спирмену без столбцов 'Относительное удлинение при растяжении' и 'Относительное удлинение' после предобработки методом метода MinMaxScaler

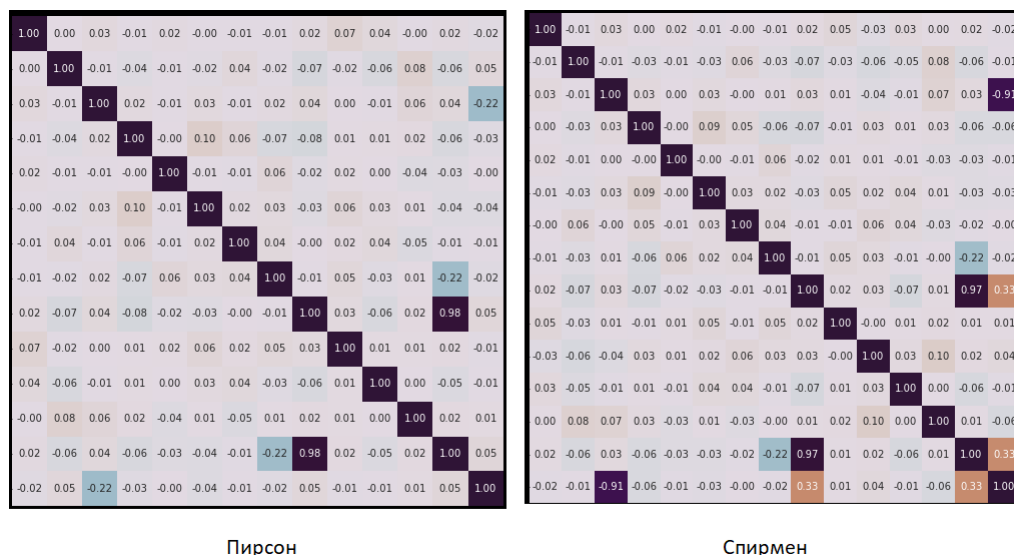


Рисунок 17. Матрицы корреляции по Пирсону и Спирмену с включением столбцов 'Относительное удлинение при растяжении' и 'Относительное удлинение' после предобработки методом метода MinMaxScaler

Данные предварительно не очищались ни по методу трех сигм, ни по методу межквартильного расстояния, см. выше п. 2.2.

Судя по матрицам корреляции по Пирсону для количественных данных и по Спирмену для количественных и категориальных данных, коэффициенты корреляции после нормализации параметров ожидаемо мало изменились, поэтому проверка значимости коэффициентов корреляции на этом этапе не осуществлялась.

2.4 Разработка моделей машинного обучения и обучение нейросети

Разработка моделей машинного обучения и обучение нейросети состоит из следующих основных этапов:

- разбиение данных на обучающую и тестовую выборки,
- модель метода К ближайших соседей,
- модель метода стохастический градиентный спуск,
- модель метода линейная регрессия,
- модель метода случайный лес,
- модель метода многослойный перцептрон,
- построение графиков ошибок моделей,
- построение нейронной сети на Keras.

Для примера на рисунке 18 ниже показано как осуществляется разбиение выборки на обучающую и тестовую выборки:

Разбиение данных на обучающую и тестовую выборки

```

1 x_upr = df.drop(['Модуль упругости при растяжении, ГПа'], axis=1)
2 x_pr = df.drop(['Прочность при растяжении, МПа'], axis=1)
3 y_upr = df[['Модуль упругости при растяжении, ГПа']]
4 y_pr = df[['Прочность при растяжении, МПа']]
5
6 X_train_upr, X_test_upr, y_train_upr, y_test_upr = train_test_split(x_upr, y_upr, test_size=0.3, random_state=5)
7 X_train_pr, X_test_pr, y_train_pr, y_test_pr = train_test_split(x_pr, y_pr, test_size=0.3, random_state=5)

```

Рисунок 18. Разбиение выборки на тестовую и обучающую выборки

Реализация остальных этапов применена в трех стратегиях, описанных ниже в п. 2.4.1 — 2.4.3.

2.4.1 Модели на основе целого исходного датасета

Первая стратегия осуществляется предобработкой (нормализацией) исходного датасета и построением моделей машинного обучения для полного датасета без очистки: а) без включения параметров 'Относительное удлинение при растяжении' и 'Относительное удлинение' в исследуемые модели; б) с включения таковых в модели.

Результаты расчетов датасета без параметров 'Относительное удлинение при растяжении' и 'Относительное удлинение' с помощью различных регрессоров, включая многослойный перцептрон, представлены ниже на рисунке 19.

	Model	MAE	R2 score
Прочность при растяжении	RandomForestRegressor_pr	0.135636	-0.024
Прочность при растяжении	KNeighborsRegressor_pr	0.135928	-0.028
Прочность при растяжении	LinearRegression_pr	0.137698	-0.047
Прочность при растяжении	MLPRegressor_pr	0.138901	-0.061
Прочность при растяжении	SGDRegressor_pr	0.140278	-0.089
Модуль упругости при растяжении	MLPRegressor_upr	0.142147	-0.000
Модуль упругости при растяжении	SGDRegressor_upr	0.142171	-0.000
Модуль упругости при растяжении	RandomForestRegressor_upr	0.142513	-0.006
Модуль упругости при растяжении	KNeighborsRegressor_upr	0.143084	-0.007
Модуль упругости при растяжении	LinearRegression_upr	0.144489	-0.025

Рисунок 19. Результаты расчетов без параметров 'Относительное удлинение при растяжении' и 'Относительное удлинение'

Если коэффициент детерминации равен нулю, это значит, что модель прогнозирует данные с таким же результатом, как если бы бралось среднее значение прогнозируемого параметра; если же коэффициент детерминации отрицательный, это значит, что модель прогнозирует данные с результатом хуже среднего значения прогнозируемого параметра, то есть разработанная модель дает прогноз даже хуже, чем простое усреднение. Средние ошибки моделей примерно одинаковые, и результаты расчета показывают, что все регрессоры одинаково плохо справились, поскольку все коэффициенты детерминации меньше нуля.

Следует рассмотреть построение нейронных сетей на Keras для прогнозирования значений параметров «Модуль упругости при растяжении», «Прочность при растяжении» и «Соотношение матрица-наполнитель».

Результаты расчетов нейронных сетей на tensorflow.keras для прогнозирования значений параметров «Модуль упругости при растяжении», «Прочность при растяжении» и «Соотношение матрица-наполнитель» датасета без параметров 'Относительное удлинение при растяжении' и 'Относительное удлинение' сведены в таблицу 2, а архитектура нейронных сетей приведена в Приложении 3.

Таблица 2. Результаты расчетов моделей нейронных сетей без параметров 'Относительное удлинение при растяжении' и 'Относительное удлинение'

Параметр	Нейронная сеть	MAE	MAE_abs
Модуль упругости при растяжении	Приложение 3, Model: "sequential"	0.1546	0.142124
	Приложение 3, Model: "sequential_1"	0.1500	0.142124
	Приложение 3, Model: "sequential_2"	0.1585	0.142124
Прочность при растяжении	Приложение 3, Model: "sequential_3"	0.1539	0.133637
	Приложение 3, Model: "sequential_4"	0.1681	0.133637
	Приложение 3, Model: "sequential_5"	0.1591	0.133637
Соотношение матрица-наполнитель	Приложение 3, Model: "sequential_6"	0.1780	0.13474
	Приложение 3, Model: "sequential_7"	0.1868	0.13474
	Приложение 3, Model: "sequential_8"	0.2264	0.13474

Показатели MAE моделей нейронных сетей в таблицы 2 плохие, и не позволяют использовать эти модели для прогноза параметров «Модуль упругости

при растяжении», «Прочность при растяжении» и «Соотношение матрица-наполнитель», для которых они создавались.

Графики для параметра «Прочность при растяжении» нейронной сети Model: "sequential_4" приведены ниже на рисунках 20 — 22 в качестве примера.

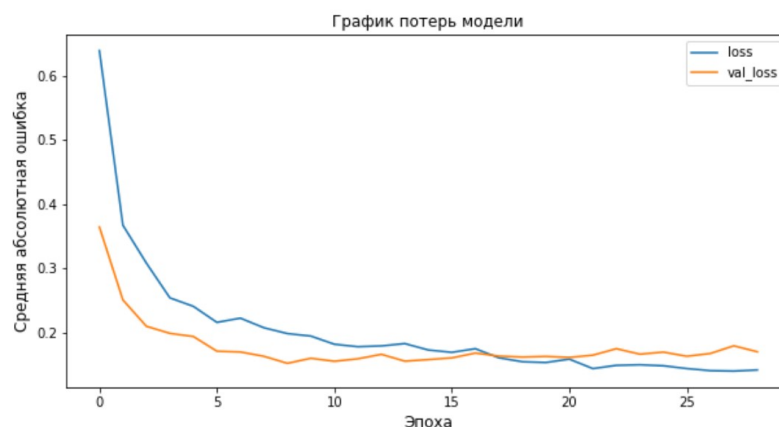


Рисунок 20. График потерь модели сети "sequential_4" для параметра «Прочность при растяжении»

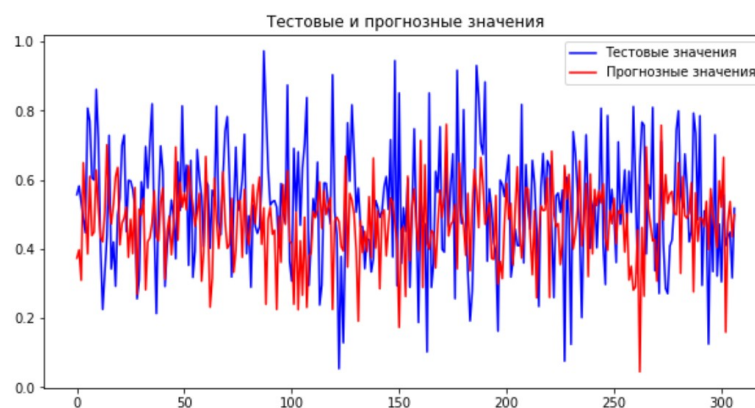


Рисунок 21. Тестовые и прогнозные значения параметра «Прочность при растяжении» модели сети "sequential_4"

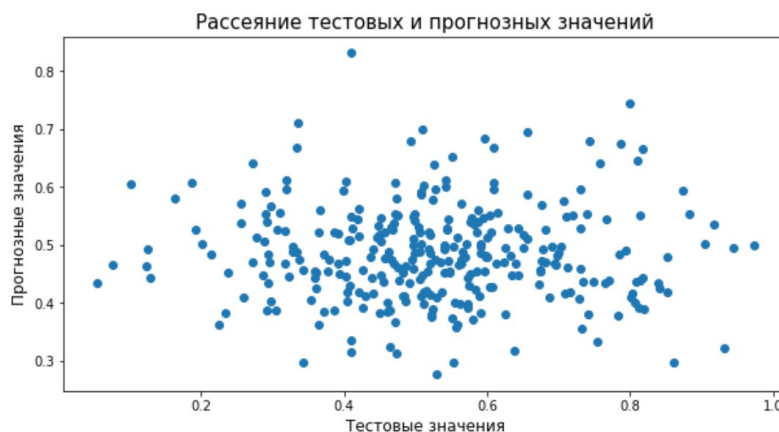


Рисунок 22. Рассеяние тестовых и прогнозных значений параметра «Прочность при растяжении» модели сети "sequential_4"

Графики на рисунках 20 — 22 также показывают, что прогнозирование параметра «Прочность при растяжении» моделью нейронной сети "sequential_4" оставляет желать лучшего, как и всеми остальными моделями сетей таблицы 2.

Результаты расчетов датасета с параметрами 'Относительное удлинение при растяжении' и 'Относительное удлинение' с помощью различных регрессоров, включая многослойный перцептрон, представлены ниже на рисунке 23.

	Model	MAE	R2 score
Прочность при растяжении	LinearRegression_pr	0.005059	0.998
Прочность при растяжении	SGDRegressor_pr	0.005385	0.998
Прочность при растяжении	MLPRegressor_pr	0.010862	0.992
Прочность при растяжении	RandomForestRegressor_pr	0.014409	0.982
Модуль упругости при растяжении	LinearRegression_upr	0.021996	0.961
Модуль упругости при растяжении	SGDRegressor_upr	0.022197	0.963
Модуль упругости при растяжении	MLPRegressor_upr	0.034593	0.935
Прочность при растяжении	KNeighborsRegressor_pr	0.075181	0.697
Модуль упругости при растяжении	RandomForestRegressor_upr	0.087186	0.545
Модуль упругости при растяжении	KNeighborsRegressor_upr	0.141031	0.020

Рисунок 23. Результаты расчетов с параметрами 'Относительное удлинение при растяжении' и 'Относительное удлинение'

Видно, что средние ошибки методов и коэффициенты детерминации варьируются, но результаты расчетов гораздо лучше. При расчете параметра «Прочность при растяжении» датасета с параметрами 'Относительное удлинение при растяжении' и 'Относительное удлинение' в целом все регрессоры показали себя удовлетворительно, за исключением регрессора KneighborsRegressor(), который тоже показал себя намного лучше, чем в предыдущем расчете без параметров 'Относительное удлинение при растяжении' и 'Относительное удлинение'. Для параметра «Модуль прочности при растяжении» тоже все регрессоры показали себя удовлетворительно, кроме того же регрессора KneighborsRegressor(), который имеет коэффициент детерминации чуть больше нуля, равный 0,020, а также регрессора RandomForestRegressor(),

который тоже показал себя намного лучше, чем в предыдущем расчете, но имеет коэффициент детерминации равный 0,697, что все же хуже, чем у других регрессоров.

Результаты расчетов нейронных сетей на tensorflow.keras для прогнозирования значений параметров «Модуль упругости при растяжении», «Прочность при растяжении» и «Соотношение матрица-наполнитель» датасета с параметрами 'Относительное удлинение при растяжении' и 'Относительное удлинение' сведены в таблицу 3, архитектура нейронных сетей также аналогична архитектурам с теми же названиями, приведенным в Приложении 3.

Таблица 3. Результаты расчетов нейронных сетей с параметрами 'Относительное удлинение при растяжении' и 'Относительное удлинение'

Параметр	Нейронная сеть	MAE	MAE_abs
Модуль упругости при растяжении	Приложение 3, Model: "sequential_2"	0.1298	0.142124
Прочность при растяжении	Приложение 3, Model: "sequential_5"	0.0484	0.133637
Соотношение матрица-наполнитель	Приложение 3, Model: "sequential_8"	0.1688	0.13474

Показатели MAE моделей таблицы 3 несколько лучше, чем таблицы 2, особенно для параметра «Прочность при растяжении». Это позволяет использовать эту модель нейронной сети для прогноза параметра «Прочность при растяжении».

Графики для параметра «Прочность при растяжении» нейронной сети Model: "sequential_5" приведены ниже на рисунках 24 — 26.

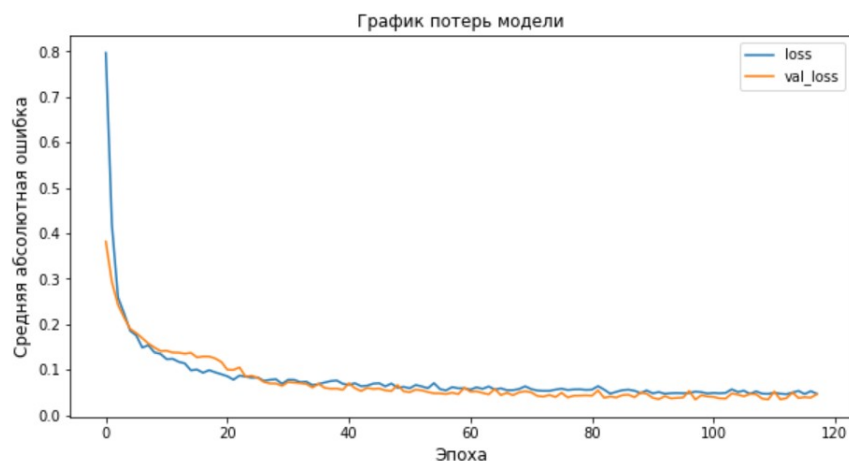


Рисунок 24. График потерь модели сети "sequential_5" для параметра «Прочность при растяжении»

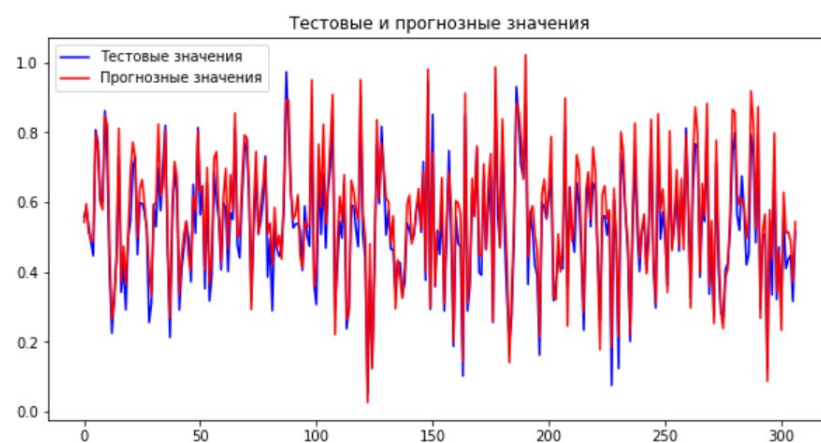


Рисунок 25. Тестовые и прогнозные значения параметра «Прочность при растяжении» модели сети "sequential_5"

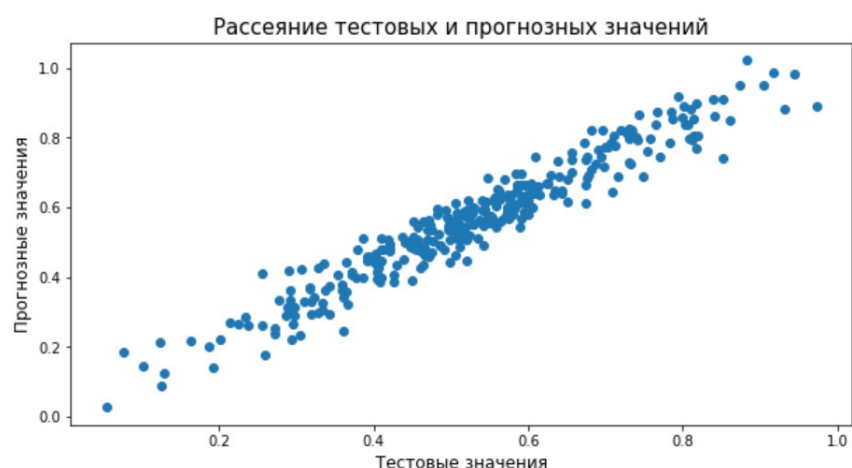


Рисунок 26. Рассеяние тестовых и прогнозных значений параметра «Прочность при растяжении» модели сети "sequential_5"

Графики 24 — 26 показывают, что прогнозирование параметра «Прочность при растяжении» датасета с параметрами 'Относительное удлинение при

растяжении' и 'Относительное удлинение' моделью нейронной сети "sequential_5" лучше, чем для датасета без этих параметров, поскольку нейронная сеть в этом случае использует возможности прогнозирования линейной регрессией, которая работает для этого датасета лучше всего, см. также рисунок 23. График рассеяния тестовых и прогнозных значений тоже выглядит лучше, чем у датасета без параметров 'Относительное удлинение при растяжении' и 'Относительное удлинение'.

Таким образом, улучшение результатов работы регрессоров и моделей нейронной сети здесь достигнуто с помощью введения в датасет расчетных параметров 'Относительное удлинение при растяжении' и 'Относительное удлинение'. Полученные результаты могут привести к лучшему пониманию данных исходного датасета. Поскольку после введения расчетных параметров 'Относительное удлинение при растяжении' и 'Относительное удлинение' достигнута задача прогнозирования значений параметра «Прочность при растяжении» и в какой-то мере улучшены результаты прогнозирования параметра «Модуль упругости при растяжении», а в отношении параметра «Соотношение матрица-наполнитель» поставленная задача не достигнута, можно сделать предположение, что между параметрами 'Модуль упругости при растяжении, ГПа' и 'Прочность при растяжении, МПа' все-таки существует статистически значимая корреляционная связь в какой-то части датасета, что касается параметра 'Соотношение матрица-наполнитель', то подобное предположение не представляется возможным высказать после настоящего этапа исследования.

2.4.2 Модели на основе датасета, подвергнутого кластеризации

Вторая стратегия осуществляется предобработкой (нормализацией) исходного датасета, проведением кластеризации датасета, выделением частей со значимыми корреляционными связями и построением для этих частей моделей машинного обучения.

Предварительно прошедший предобработку датасет следует подвергнуть кластеризации. Кластеризация позволяет решать задачу структурирования дан-

ных по заданным правилам. Доступны только неструктурированные сущности [7].

Хотя регрессор `KneighborsRegressor()` показал себя не очень хорошо, см. п. 2.4.1, метод *k*-ближайших соседей следует использовать для кластеризации в первом приближении, поскольку он является одним из наиболее простых и одним из наиболее мощных. Этот метод является хорошим базовым методом, чтобы получить какую-то первичную точность [1].

Правильное применение этого метода на очень больших выборках данных позволяет достичь наиболее точного решения, но есть проблема большой выборки данных, потому что метод начинает работать менее эффективно. Большой сложностью является проблема расстояния. Поскольку выбираются ближайшие соседи, следует каким-то образом нормировать расстояние между точками. Предварительно все данные должны быть нормированы к одному диапазону, то есть приведение к интервалу от 0 до 1 всех параметров, по которым принимается решение, что и было произведено. Но возникает проблема «Проклятие большой размерности», когда расстояние между удаленными объектами становится примерно одинаковым. В случае разности по большому количеству параметров невозможно установить достоверно, что объект нужного класса находится рядом или не рядом. Приходится выделять главные признаки, выделять второстепенные, и сначала посчитать расстояние по главным, если по главным выходит двусмысленность, то считать уже по второстепенным параметрам расстояние, и по ним уже принимать решение. Для больших выборок классификацию приходится проводить по сетке — параметры разбиваются на диапазоны и после этого принимаются решения в зависимости от этих диапазонов. Однако данный метод становится практически идентичен дереву принятия решений. Метод является хорошим базовым методом в ряде задач, и позволяет достаточно приблизить уже финальное решение, но лучшая его реализация заключается в дереве принятия решений.

Исходя из этих соображений следует произвести предварительную кластеризацию по всем имеющимся параметрам датасета, что и было выполнено. В результате такого подхода было установлено, что в датасете без расчетных пара-

метров 'Относительное удлинение при растяжении' и 'Относительное удлинение' стало возможным методом K-means из sklearn.cluster.KMeans явно выделить кластеры, показанные ниже на рисунках 27 и 28.

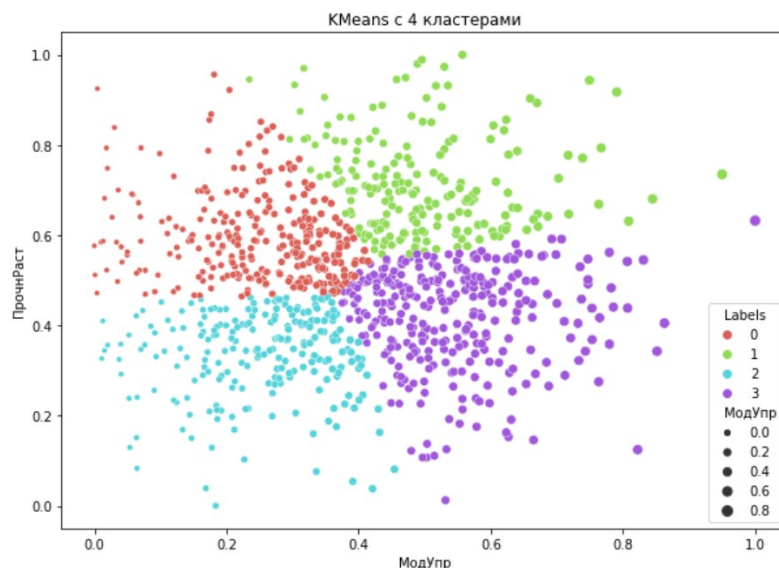


Рисунок 27. Четыре кластера параметров 'Прочность при растяжении, МПа' и 'модуль упругости, ГПа', метод K-means

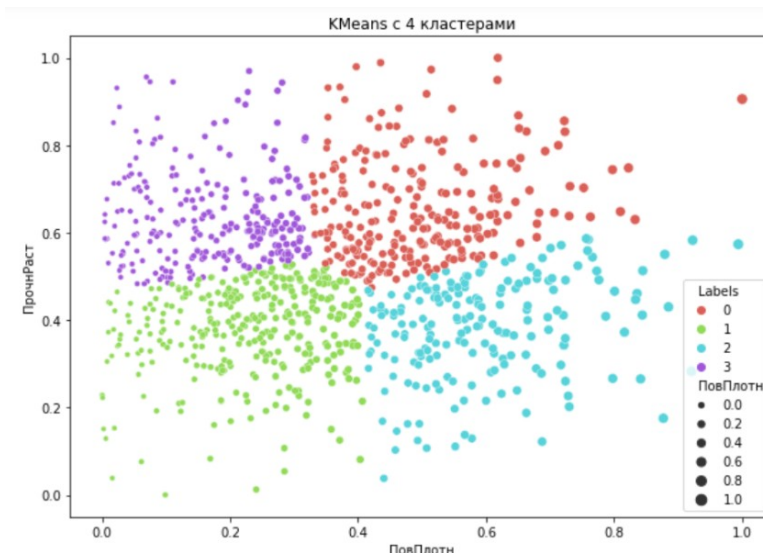


Рисунок 28. Четыре кластера параметров 'Прочность при растяжении, МПа' и 'Поверхностная плотность, г/м2', метод K-means

При выделении кластеров исключались, все параметры, кроме, например параметра 'Прочность при растяжении, МПа' и еще какого одного параметра из датасета. Попытка выделить кластеры из датасета по всем параметрам сразу не привела к успеху. Таким образом был проанализирован весь датасет.

Следующим шагом была кластеризация датасета по параметрам 'Прочность при растяжении, МПа', 'модуль упругости, ГПа' и 'Поверхностная плотность, г/м²' при исключенных остальных параметрах, поскольку границы кластеров на рисунках 27 и 28 кажутся довольно близкими. При включении дополнительных параметров границы кластеров расплываются, поскольку расстояние между объектами становится примерно одинаковым. Результаты представлены на рисунках 29 и 30 ниже.

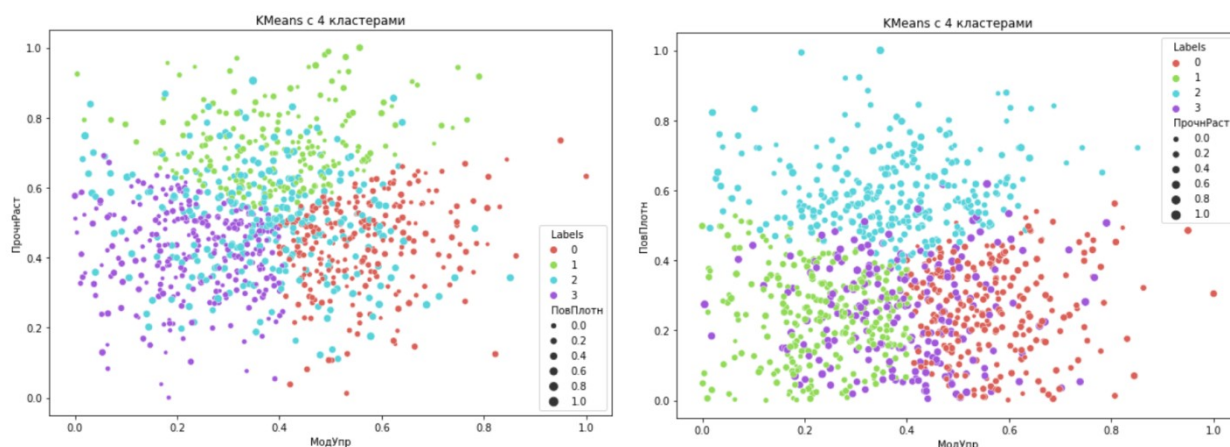


Рисунок 29. Четыре кластера параметров 'Прочность при растяжении, МПа', 'модуль упругости, ГПа' и 'Поверхностная плотность, г/м²', метод K-means

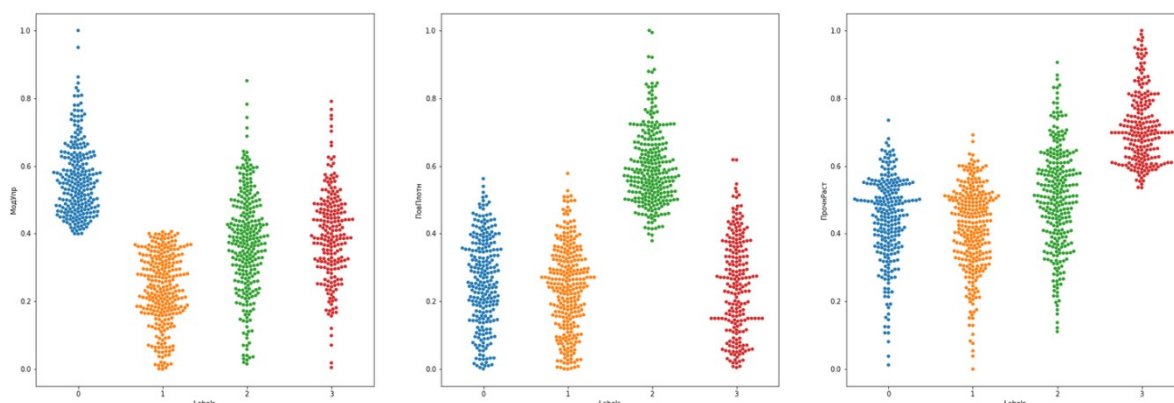


Рисунок 30. Четыре кластера параметров 'Прочность при растяжении, МПа', 'модуль упругости, ГПа' и 'Поверхностная плотность, г/м²' показанные seaborn.swarmplot, метод K-means

Показатель Silhouette score: 0.2422176450274393 для K-means .

На рисунках 29 и 30 кластеры довольно отчетливо выделяются, поэтому следует остановиться на таком разбиении, проверить статистическую значи-

мость различия полученных групп, затем произвести статистический анализ значимости корреляционных связей по каждой части датасета, выделить одну часть, и далее строить модели регрессоров и нейронных сетей для этой части.

Окончательно построение кластеров было осуществлено методом C-means из `fcmeans.FCM()`, представляющим собой развитие и улучшение метода K-means из `sklearn.cluster.Kmeans`. C-means справился примерно также, как и метод K-means (см. рис. 29), что и показано на рисунке 31 ниже.

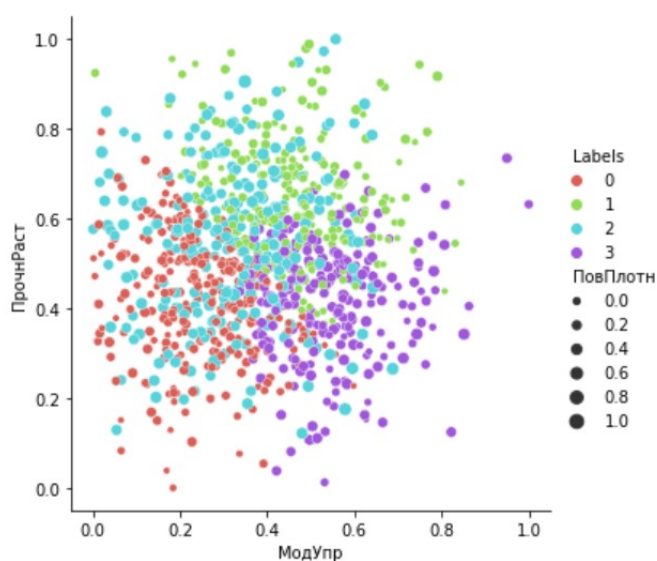


Рисунок 31. Четыре кластера параметров 'Прочность при растяжении, МПа', 'модуль упругости, ГПа' и 'Поверхностная плотность, г/м²', метод C-means

Показатель Silhouette score: 0.22793128028194928 для C-means.

Проверка различия полученных четырех кластеров может быть произведена по непараметрическому критерию Краскала-Уоллиса, предназначенному для сравнения групп более двух. Формулируется статистическая гипотеза для выявления различия групп по критерию Краскала-Уоллиса: H_0 : Статистически значимые отличия в группах отсутствуют с уровнем значимости равным 0,05; H_1 : Присутствуют статистически значимые отличия в группах с тем же уровнем значимости. Результат представлен на рисунке 32 ниже.

Проверка значимости различия по критерию Краскала-Уоллиса

```

1 # индексация кластеров
2 filter_0 = X[X['Labels']==0].index
3 filter_1 = X[X['Labels']==1].index
4 filter_2 = X[X['Labels']==2].index
5 filter_3 = X[X['Labels']==3].index

1 # выделение объектов np.array
2 df_0 = np.array(df_s['ПрочнРаст'][filter_0].reset_index(drop=True))
3 df_1 = np.array(df_s['ПрочнРаст'][filter_1].reset_index(drop=True))
4 df_2 = np.array(df_s['ПрочнРаст'][filter_2].reset_index(drop=True))
5 df_3 = np.array(df_s['ПрочнРаст'][filter_3].reset_index(drop=True))

1 # критерий Краскала-Уоллиса
2 stats.mstats.kruskalwallis(df_0,df_1,df_2,df_3)

KruskalResult(statistic=342.1986251717648, pvalue=7.292148282170502e-74)

```

Рисунок 32. Проверка значимости различия кластеров по Краскалу-Уоллису

Принимается альтернативная гипотеза H1: Присутствуют статистически значимые отличия в полученных кластерах с уровнем значимости 0,05.

Таким образом, получены четыре датасета (247, 216, 279, 281 строк), для которых следует проверить распределение параметров на нормальность, построить матрицы корреляции и установить значимость корреляций для каждого. Ниже представлены описание основных результатов анализа и матрицы корреляции на рисунках 33 — 36 ниже для каждого кластера отдельно.

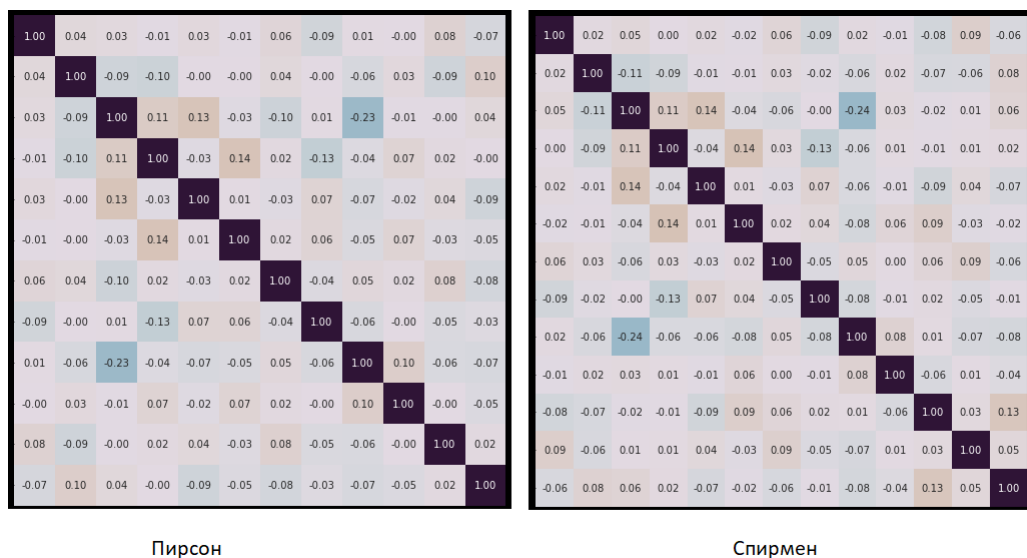


Рисунок 33. Корреляционные матрицы 1-го кластера

Согласно тестам по критериям д'Агостино и Шапиро-Уилка статистически отличаются от нормального (уровень значимости 0,05) распределения следующих параметров: 'Поверхностная плотность, г/м2', 'Плотность нашивки'.

Проверка по критерию Пирсона показала, что статистически значимы следующие корреляции:

модуль упругости, ГПа & Содержание эпоксидных групп,%_2 : Statistics=0.135, p-value=0.023
 модуль упругости, ГПа & Прочность при растяжении, МПа : Statistics=-0.229, p-value=0.000
 Количество отвердителя, м.% & Температура вспышки, С_2 : Statistics=0.139, p-value=0.019
 Количество отвердителя, м.% & Модуль упругости при растяжении, ГПа : Statistics=-0.128, p-value=0.031

Проверка по критерию Спирмена показала, что статистически значимы следующие корреляции:

Угол нашивки, град & Плотность нашивки : Statistics=0.129, p-value=0.030

В 1-ом кластере по критерию Пирсона статистически значима корреляция: 'модуль упругости, ГПа' и 'Прочность при растяжении, МПа' (Statistics=-0.229, p-value=0.000), которая представляет интерес с точки зрения физико-механических свойств материалов.

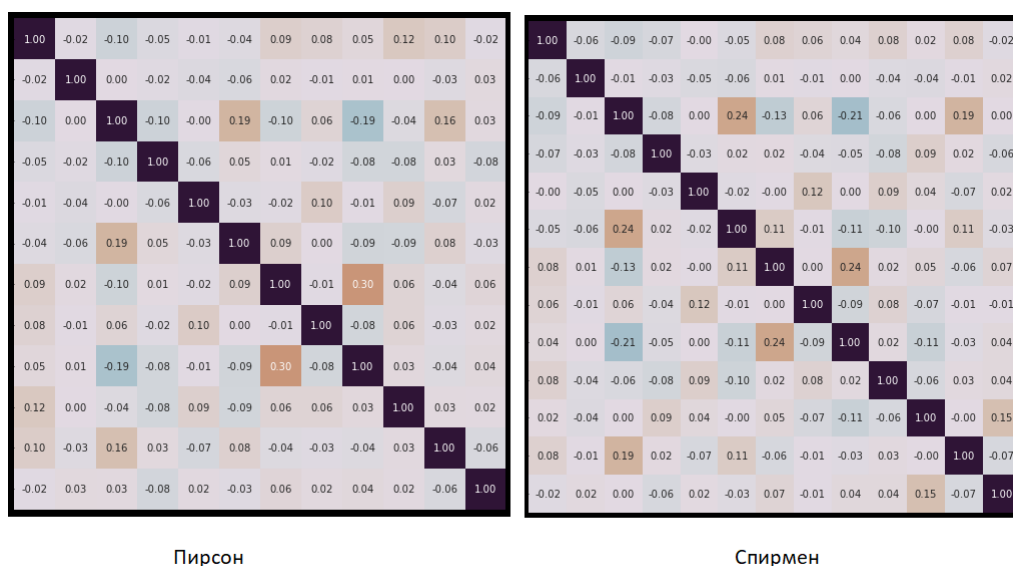


Рисунок 34. Корреляционные матрицы корреляции 2-го кластера

Согласно тестам по критериям д'Агостино и Шапиро-Уилка статистически отличаются от нормального (уровень значимости 0,05) распределения следующих параметров: 'Содержание эпоксидных групп,%_2', 'Температура

вспышки, C_2', 'Поверхностная плотность, г/м2', 'Шаг нашивки', 'Прочность при растяжении, МПа'.

Проверка по критерию Пирсона показала, что статистически значимых корреляций в датасете не имеется.

Проверка по критерию Спирмена показала, что статистически значимы следующие корреляции:

модуль упругости, ГПа & Температура вспышки, C_2 : Statistics=0.242, p-value=0.000

модуль упругости, ГПа & Поверхностная плотность, г/м2 : Statistics=-0.134, p-value=0.034

модуль упругости, ГПа & Прочность при растяжении, МПа : Statistics=-0.208, p-value=0.001

модуль упругости, ГПа & Шаг нашивки : Statistics=0.192, p-value=0.002

Содержание эпоксидных групп, %_2 & Модуль упругости при растяжении, ГПа : Statistics=0.125, p-value=0.048

Поверхностная плотность, г/м2 & Прочность при растяжении, МПа : Statistics=0.242, p-value=0.000

Угол нашивки, град & Плотность нашивки : Statistics=0.155, p-value=0.014

В 2-ом кластере по критерию Спирмена статистически значима корреляция: 'модуль упругости, ГПа' и 'Прочность при растяжении, МПа' (Statistics=-0.208, p-value=0.001), которая также представляет интерес с точки зрения физико-механических свойств материалов.

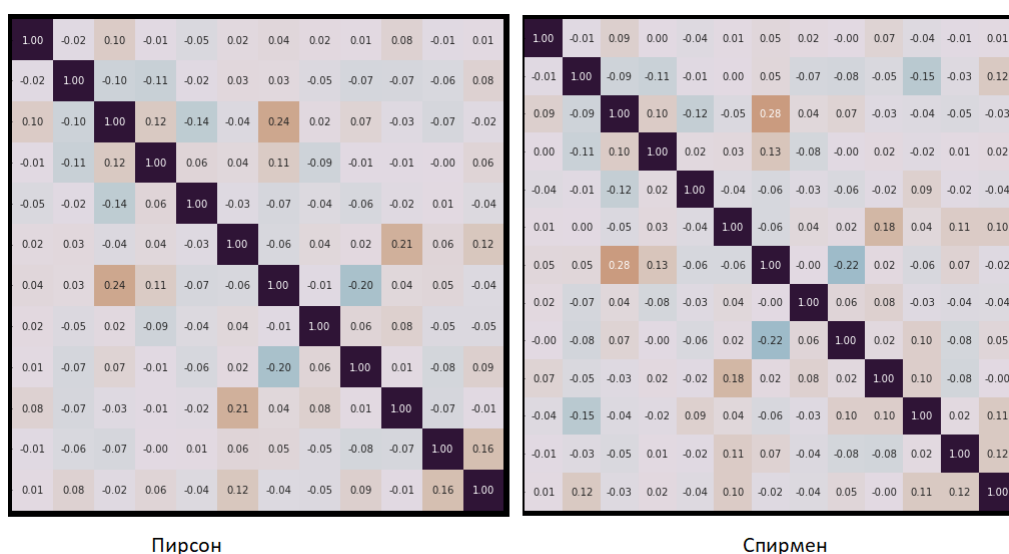


Рисунок 35. Корреляционные матрицы корреляции 3-го кластера

Согласно тестам по критериям д'Агостино и Шапиро-Уилка статистически отличаются от нормального (уровень значимости 0,05) распределения следующих параметров: 'Поверхностная плотность, г/м2', 'Потребление смолы, г/м2', 'Плотность нашивки'.

Проверка по критерию Пирсона показала, что статистически значимы следующие корреляции:

модуль упругости, ГПа & Содержание эпоксидных групп, %₂ : Statistics=-0.138, p-value=0.025

Проверка по критерию Спирмена показала, что статистически значимы следующие корреляции:

Плотность, кг/м³ & Угол нашивки, град : Statistics=-0.152, p-value=0.014

модуль упругости, ГПа & Поверхностная плотность, г/м² : Statistics=0.276, p-value=0.000

Количество отвердителя, м.% & Поверхностная плотность, г/м² : Statistics=0.131, p-value=0.034

Температура вспышки, C₂ & Потребление смолы, г/м² : Statistics=0.181, p-value=0.003

Поверхностная плотность, г/м² & Прочность при растяжении, МПа : Statistics=-0.221, p-value=0.000

В 3-ьем кластере не наблюдается статистически значимых корреляций важных с точки зрения физико-механических свойств материалов.

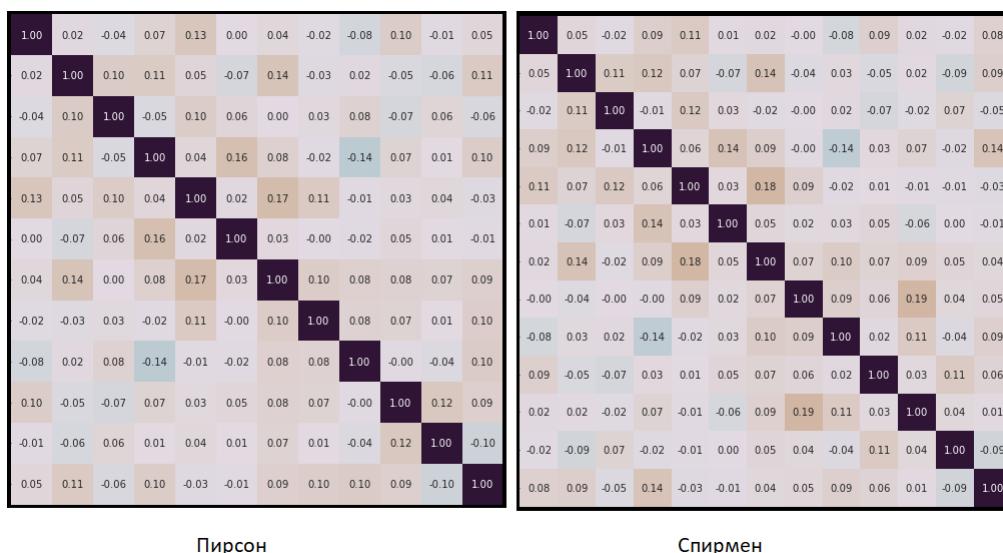


Рисунок 36. Корреляционные матрицы корреляции 4-го кластера

Согласно тестам по критериям д'Агостино и Шапиро-Уилка статистически отличаются от нормального (уровень значимости 0,05) распределения следующих параметров: 'модуль упругости, ГПа', 'Модуль упругости при растяжении, ГПа', 'Прочность при растяжении, МПа'.

Проверка по критерию Пирсона показала, что статистически значимы следующие корреляции:

Плотность, кг/м³ & Поверхностная плотность, г/м² : Statistics=0.135, p-value=0.043

Количество отвердителя, м.% & Температура вспышки, C₂ : Statistics=0.156, p-value=0.019

Содержание эпоксидных групп, %₂ & Поверхностная плотность, г/м² : Statistics=0.166, p-value=0.013

Проверка по критерию Спирмена показала, что статистически значимы следующие корреляции:

Количество отвердителя, м.% & Прочность при растяжении, МПа : Statistics=-0.137, p-value=0.040
 Модуль упругости при растяжении, ГПа & Угол нашивки, град : Statistics=0.186, p-value=0.005

В 4-ом кластере не наблюдается статистически значимых корреляций важных с точки зрения физико-механических свойств материалов.

Полностью вся описательная статистика ненормализованных параметров кластеров представлена в Приложении 4. Выборочная описательная статистика для параметров 'Прочность при растяжении, МПа', 'модуль упругости, ГПа' и 'Поверхностная плотность, г/м²' представлена в таблице 4 ниже.

Таблица 4. Выборочная описательная статистика 1 — 4 кластеров

Параметр	1 кластер	2 кластер	3 кластер	4 кластер
'Прочность при растяжении, МПа'	Среднее: 2217.313804835351 Медиана: 2229.63118003497 Мода : 1800.0 Стандарт : 382.0260134288249 Дисперсия: 145943.8749363207 Асимметрия: - 0.25528976335159875 Эксцесс: 0.027061744863451587 Минимум: 1036.85660535 Максимум: 3267.6609524823 Размах: 2230.8043471323 ДоверитИнтервалСредн: (2172.804807622834, 2261.822802047863)	Среднее: 2869.2488008064674 Медиана: 2826.72324751235 Мода : 3000.0 Стандарт : 369.6378656829604 Дисперсия: 136632.1517466543 Асимметрия: 0.3714971090470974 Эксцесс: - 0.1475612710684735 Минимум: 2066.55369198341 Максимум: 3817.26948410351 Размах: 1750.7157921201 ДоверитИнтервалСредн: (2823.5202241047823, 2914.97737750815)	Среднее: 2572.6542915062655 Медиана: 2576.8116852702897 Мода : 2000.0 Стандарт : 475.15683283961624 Дисперсия: 225774.01579417504 Асимметрия: 0.00174515679434228 Эксцесс: - 28 0.29483236272840685 Минимум: 1379.56851408343 Максимум: 3848.43673187618 Размах: 2468.8682177927503 ДоверитИнтервалСредн: (2515.3373369949263, 2629.971246017603)	Среднее: 2208.0003526732103 Медиана: 2250.21111996466 Мода : 2000.0 Стандарт : 366.17000727115044 Дисперсия: 134080.47422495438 Асимметрия: - 0.5295796554457651 Эксцесс: 0.20647540976278922 Минимум: 1071.12375147656 Максимум: 3103.00073322051 Размах: 2031.87698174395 ДоверитИнтервалСредн: (2160.1550175751945, 2255.845687771224)
'модуль упругости, ГПа'	Среднее: 471.4998946880096 Медиана: 459.047069741573 Мода : 535.0 Стандарт : 216.83408004207092 Дисперсия: 47017.018267691215 Асимметрия: 0.11719582975906488 Эксцесс: - 0.12100394809170645 Минимум:	Среднее: 888.3094354737685 Медиана: 888.191562497692 Мода : 738.736842105263 Стандарт : 251.2917351137937 Дисперсия: 63147.536136501054 Асимметрия: 0.1609419921233092 Эксцесс: 0.401779433425665	Среднее: 624.2568798200306 Медиана: 630.0575352243875 Мода : 933.0 Стандарт : 275.74212091114845 Дисперсия: 76033.71724457841 Асимметрия: - 0.05614084610676067 Эксцесс: - 0.31191215380164694 Минимум:	Среднее: 1047.7222770504497 Медиана: 1027.54525685193 Мода : 889.0 Стандарт : 227.99477332591238 Дисперсия: 51981.61666393416 Асимметрия: 0.7877035339592245 Эксцесс: 0.6456870125502503 Минимум:

Параметр	1 кластер	2 кластер	3 кластер	4 кластер
	4.33915435296057 Максимум: 1144.21605749707 Размах: 1139.8769031441095 ДоверитИнтервалСредн: (446.2370405238422, 496.7627488521773)	Минимум: 11.312942577361 Максимум: 1615.09697286127 Размах: 1603.784030283909 ДоверитИнтервалСредн: (857.2216751264388, 919.3971958210981)	2.4369087535075 Максимум: 1315.30489207828 Размах: 1312.8679833247727 ДоверитИнтервалСредн: (590.9948124165738, 657.5189472234866)	645.814748630493 Максимум: 1911.53647700054 Размах: 1265.721728370047 ДоверитИнтервалСредн: (1017.9315074249721, 1077.513046675927)
'Поверхностная плотность, г/м2'	Среднее: 304.75145752384026 Медиана: 315.234693245364 Мода : 380.0 Стандарт : 160.48410480924824 Дисперсия: 25755.147896425777 Асимметрия: - 0.00635460933881978 Эксцесс: - 0.7327156991719574 Минимум: 0.603739925153945 Максимум: 696.309114521802 Размах: 695.7053745966481 ДоверитИнтервалСредн: (286.0538130637241, 323.44910198395644)	Среднее: 266.7014714025315 Медиана: 267.140817459791 Мода : 210.0 Стандарт : 157.95981779215447 Дисперсия: 24951.304036930636 Асимметрия: 0.2713870068058269 Эксцесс: - 0.6903732268714098 Минимум: 6.77925525877907 Максимум: 709.878674214847 Размах: 703.099418956068 ДоверитИнтервалСредн: (247.15997331941045, 286.24296948565245)	Среднее: 832.2349941358942 Медиана: 806.2051499175379 Мода : 1010.0 Стандарт : 165.39746596635527 Дисперсия: 27356.321748091646 Асимметрия: 0.7358814885990831 Эксцесс: 0.4359588429890584 Минимум: 530.591316744297 Максимум: 1399.54236233989 Размах: 868.9510455955929 ДоверитИнтервалСредн: (812.283520479439, 852.1864677923488)	Среднее: 537.5017553117506 Медиана: 544.203244877878 Мода : 1010.0 Стандарт : 171.41434247499703 Дисперсия: 29382.87680613557 Асимметрия: - 0.03511182540845649 Эксцесс: 0.21338331177744108 Минимум: 27.5365275109749 Максимум: 1010.0 Размах: 982.4634724890251 ДоверитИнтервалСредн: (515.1040261327763, 559.8994844907241)

По данным таблицы 4 и Приложения 4 обнаружено, что представленные в них параметры имеют моды, которые скорее всего имеют неслучайный характер, кроме, может быть, 738.736842105263 параметра 'модуль упругости, ГПа' 2-го кластера, но скорее всего это тоже какая-то значимая величина. Все указывает на то, что кластеры выделяются именно вокруг этих неслучайных значений, что ожидаемо, потому что в качестве выборочной моды дискретной случайной величины используется наиболее часто повторяющийся элемент. Следовательно эти величины встречаются в соответствующих столбцах датасета наиболее часто и является центрами образования кластеров методами K-means и C-means. Можно было бы просто прошерстить датасет на наличие чисел, которые имеют после запятой только нули, и выделить строки с такими данными и, таким образом ответить на первый как минимум вопрос, но здесь

следует и дальше доказывать методами машинного обучения, что в датасете имеются значимые данные.

По полученным результатам именно 1-ый кластер представляет наибольший интерес как кластер с выделяющимся по величине коэффициентом корреляции между параметрами 'Прочность при растяжении, МПа' и 'модуль упругости, ГПа', поэтому далее для него строятся модели машинного обучения.

Результаты расчетов датасета с помощью различных регрессоров, включая многослойный перцептрон, представлены ниже на рисунке 37.

	Model	MAE	R2 score
Модуль упругости при растяжении	KNeighborsRegressor_upr	0.131750	-0.040
Модуль упругости при растяжении	MLPRegressor_upr	0.132132	-0.016
Модуль упругости при растяжении	SGDRegressor_upr	0.132795	-0.037
Модуль упругости при растяжении	RandomForestRegressor_upr	0.133901	-0.045
Модуль упругости при растяжении	LinearRegression_upr	0.135513	-0.094
Прочность при растяжении	LinearRegression_pr	0.157292	-0.048
Прочность при растяжении	MLPRegressor_pr	0.158039	-0.044
Прочность при растяжении	SGDRegressor_pr	0.158357	-0.043
Прочность при растяжении	KNeighborsRegressor_pr	0.159666	-0.061
Прочность при растяжении	RandomForestRegressor_pr	0.162947	-0.091

Рисунок 37. Результаты расчетов по 1-ому кластеру

Если коэффициент детерминации равен нулю, это значит, что модель прогнозирует данные с таким же результатом, как если бы берется среднее значение прогнозируемого параметра; если же коэффициент детерминации отрицательный, это значит, что модель прогнозирует данные с результатом хуже среднего значения прогнозируемого параметра, то есть разработанная модель дает прогноз даже хуже, чем простое усреднение. Средние ошибки моделей примерно одинаковые, и результаты расчета показывают, что все регрессоры одинаково плохо справились, поскольку все коэффициенты детерминации меньше нуля.

Далее следует рассмотреть построение нейронных сетей на Keras для прогнозирования значений параметров «Модуль упругости при растяжении», «Прочность при растяжении» и «Соотношение матрица-наполнитель».

Результаты расчетов нейронных сетей на tensorflow.keras для прогнозирования значений параметров «Модуль упругости при растяжении», «Прочность при растяжении» и «Соотношение матрица-наполнитель» датасета без параметров 'Относительное удлинение при растяжении' и 'Относительное удлинение' сведены в таблицу 5, архитектура нейронных сетей также аналогична архитектурам с теми же названиями, приведенным в Приложении 3.

Таблица 5. Результаты расчетов нейронных сетей

Параметр	Нейронная сеть	MAE	MAE_abs
Модуль упругости при растяжении	Приложение 3, Model: "sequential_1"	0.1749	0.132017
Прочность при растяжении	Приложение 3, Model: "sequential_4"	0.1755	0.152869
Соотношение матрица-наполнитель	Приложение 3, Model: "sequential_7"	0.1901	0.134129

Показатели MAE моделей нейронных сетей в таблицы 5 плохие, и не позволяют использовать эти модели для прогноза параметров «Модуль упругости при растяжении», «Прочность при растяжении» и «Соотношение матрица-наполнитель», для которых они создавались.

Таким образом, результаты расчета оказались плохие, вероятно также потому, что в исходный датасет не были включены параметры 'Относительное удлинение при растяжении' и 'Относительное удлинение', а кластеризация произведена без учета влияния всего многообразия параметров. Но всего вероятнее, что значимые данные оказались в разных кластерах, поскольку они и послужили центрами образования этих кластеров, что не позволило получить хорошие модели.

Исходя из этих соображений следует попытаться произвести кластеризацию датасета с включением расчетных параметров 'Относительное удлинение при растяжении' и 'Относительное удлинение'. Следует также попытаться использовать кластеризатор другого типа, например, агломеративный кластеризатор. На этом этапе при выделении кластеров методом

`sklearn.cluster.AgglomerativeClustering()` исключены все параметры, кроме параметров 'Прочность при растяжении, МПа' и 'Относительное удлинение при растяжении'.

Показатель Silhouette score: 0.41198385885966476 для AgglomerativeClustering.

Кластеризатор отчетливо выделил три кластера. Результаты представлены на рисунке 38 ниже.

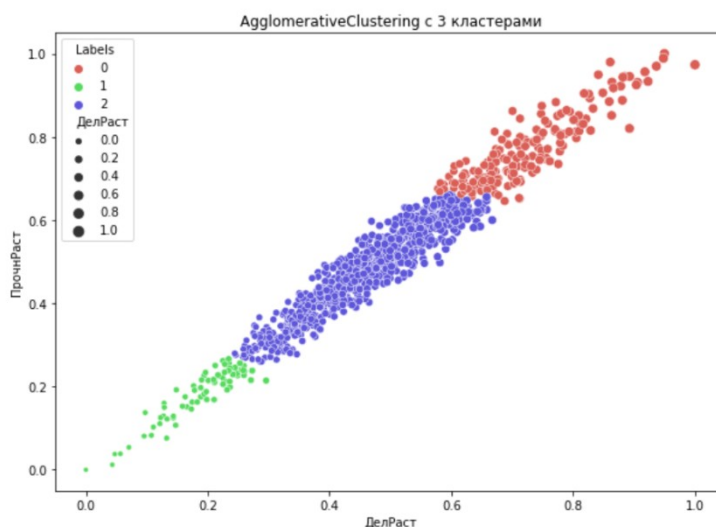


Рисунок 38. Три кластера параметров 'Прочность при растяжении, МПа' и 'Относительное удлинение при растяжении'

Дальнейшие расчеты производятся по датасету без параметров 'Относительное удлинение при растяжении' и 'Относительное удлинение', то есть они участвуют только на этапе кластеризации, в анализе и построение моделей машинного обучения они не участвуют.

Проверка различия этих кластеров произведена по непараметрическому критерию Краскала-Уоллиса, предназначенному для сравнения групп более двух. Формулируется статистическая гипотеза для выявления различия групп по критерию Краскала-Уоллиса: H_0 : Статистически значимые отличия в группах отсутствуют с уровнем значимости равным 0,05; H_1 : Присутствуют статистически значимые отличия в группах с тем же уровнем значимости. Результат `KruskalResult(statistic = 589.1783563436962, pvalue = 1.1522467492437658e-128)` показывает, что принимается альтернативная гипотеза H_1 : Присутствуют статисти-

стически значимые отличия в полученных кластерах с уровнем значимости 0,05.

Таким образом, получены три датасета (188, 71, 764 строки), которые были проанализированы с помощью тепловых карт корреляционных матриц.

По полученным результатам 2-ой кластер явно выделяется по количеству и величинам коэффициентов корреляции, следовательно 2-ой кластер представляет наибольший интерес, поэтому далее для него строятся модели машинного обучения.

Результаты расчетов датасета с помощью различных регрессоров, включая многослойный перцептрон, представлены ниже на рисунке 39.

	Model	MAE	R2 score
Модуль упругости при растяжении	KNeighborsRegressor_upr	0.138317	-0.073
Модуль упругости при растяжении	LinearRegression_upr	0.139230	-0.008
Модуль упругости при растяжении	MLPRegressor_upr	0.142063	-0.167
Модуль упругости при растяжении	RandomForestRegressor_upr	0.148034	-0.152
Прочность при растяжении	RandomForestRegressor_pr	0.148975	-0.172
Модуль упругости при растяжении	SGDRegressor_upr	0.149881	-0.192
Прочность при растяжении	KNeighborsRegressor_pr	0.171437	-0.159
Прочность при растяжении	MLPRegressor_pr	0.173698	-0.275
Прочность при растяжении	SGDRegressor_pr	0.175035	-0.234
Прочность при растяжении	LinearRegression_pr	0.239610	-1.322

Рисунок 39. Результаты расчетов по 2-ому кластеру

Если коэффициент детерминации равен нулю, это значит, что модель прогнозирует данные с таким же результатом, как если бы бралось среднее значение прогнозируемого параметра; если же коэффициент детерминации отрицательный, это значит, что модель прогнозирует данные с результатом хуже среднего значения прогнозируемого параметра, то есть разработанная модель дает прогноз даже хуже, чем простое усреднение. Средние ошибки моделей примерно одинаковые, и результаты расчета показывают, что все регрессоры ожидаемо одинаково плохо справились, поскольку все коэффициенты детерминации меньше нуля.

Далее следует рассмотреть построение нейронных сетей на Keras для прогнозирования значений параметров «Модуль упругости при растяжении», «Прочность при растяжении» и «Соотношение матрица-наполнитель».

Результаты расчетов нейронных сетей на tensorflow.keras для прогнозирования значений параметров «Модуль упругости при растяжении», «Прочность при растяжении» и «Соотношение матрица-наполнитель» датасета без параметров 'Относительное удлинение при растяжении' и 'Относительное удлинение' сведены в таблицу 6, архитектура нейронных сетей также аналогична архитектурам с теми же названиями, приведенным в Приложении 3.

Таблица 6. Результаты расчетов нейронных сетей

Параметр	Нейронная сеть	MAE	MAE_abs
Модуль упругости при растяжении	Приложение 3, Model: "sequential_1"	0.2070	0.143482
Прочность при растяжении	Приложение 3, Model: "sequential_4"	0.6480	0.149162
Соотношение матрица-наполнитель	Приложение 3, Model: "sequential_7"	0.1629	0.135772

Показатели MAE моделей нейронных сетей в таблицы 6 плохие, и не позволяют использовать эти модели для прогноза параметров «Модуль упругости при растяжении», «Прочность при растяжении» и «Соотношение матрица-наполнитель», для которых они создавались.

Таким образом, результаты расчета оказались плохие, может быть и потому, что в датасет при построении моделей машинного обучения не были включены параметры 'Относительное удлинение при растяжении' и 'Относительное удлинение', а кластеризация произведена без учета влияния всего многообразия параметров. Но очевидно, что значимые данные оказались в разных кластерах, но поскольку они послужили центрами образования этих кластеров — это не позволило получить хорошие модели для отдельных кластеров.

Далее производится кластеризация датасета с включением одного расчетного параметра 'Относительное удлинение при растяжении'. Здесь используется

кластеризатор `sklearn.cluster.Kmeans`. Из обработки кластеризатором был исключен с познавательной целью только один параметр 'Угол нашивки, град' как категориальный и имеющий два уровня.

Показатель Silhouette score: 0.08495062879359701 для K-means.

Кластеризатор выделяет три кластера.

Результаты визуализации для параметров 'Прочность при растяжении, МПа' и 'Относительное удлинение при растяжении' представлены на рисунке 40 ниже.

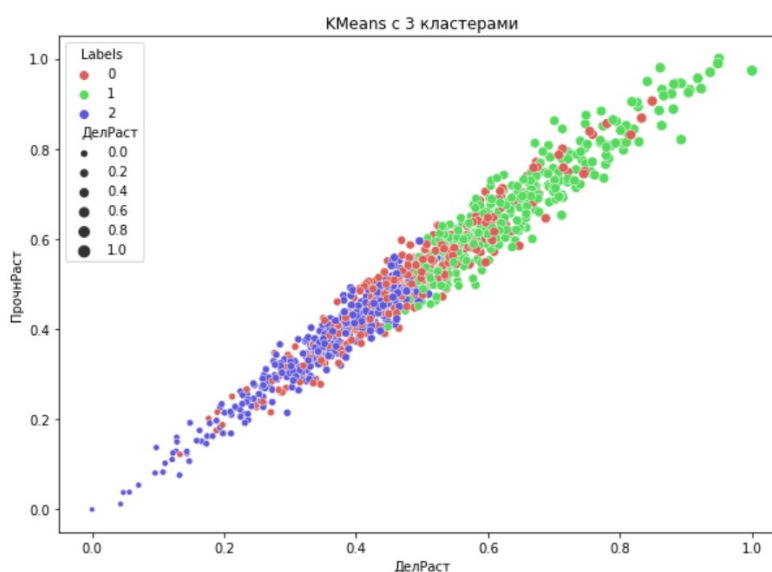


Рисунок 40. Три кластера параметров 'Прочность при растяжении, МПа' и 'Относительное удлинение при растяжении'

Дальнейшие расчеты производятся по датасету без параметра 'Относительное удлинение при растяжении', то есть он участвует только на этапе кластеризации, в анализе и в построение моделей машинного обучения они не участвуют.

Проверка различия этих кластеров произведена по непараметрическому критерию Краскала-Уоллиса, предназначенному для сравнения групп более двух. Формулируется статистическая гипотеза для выявления различия групп по критерию Краскала-Уоллиса: H_0 : Статистически значимые отличия в группах отсутствуют с уровнем значимости равным 0,05; H_1 : Присутствуют статистически значимые отличия в группах с тем же уровнем значимости. Результат

KruskalResult(statistic = 570.1706179879571, pvalue = 1.545337815538502e-124) показывает, что принимается альтернативная гипотеза H1: Присутствуют статистически значимые отличия в полученных кластерах с уровнем значимости 0,05.

Таким образом, получены три датасета (359, 353, 311 строки), которые были проанализированы с помощью тепловых карт корреляционных матриц.

По полученным результатам 1-ый кластер явно выделяется по величине коэффициента корреляции между параметрами 'Прочность при растяжении, МПа' и 'Модуль упругости при растяжении, ГПа', следовательно 1-ый кластер представляет наибольший интерес, поэтому для этого кластера следует построить матрицы корреляции и установить значимость корреляций, далее для него следует создать модели машинного обучения. Ниже представлено описание основных результатов анализа и матрицы корреляции на рисунке 41 для 1-го кластера.

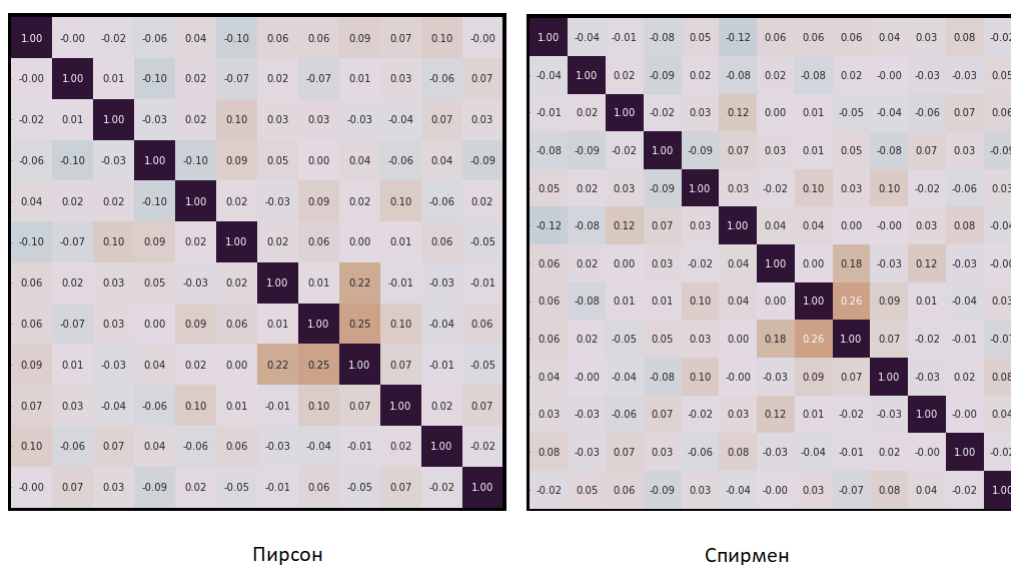


Рисунок 41. Корреляционные матрицы 1-го кластера

Согласно тестам по критерию Шапиро-Уилка статистически отличаются от нормального (уровень значимости 0,05) распределения следующих параметров: 'Поверхностная плотность, г/м2', 'Прочность при растяжении, МПа'.

Проверка по критерию Пирсона показала, что статистически значимы следующие корреляции:

Соотношение матрица-наполнитель & Температура вспышки, C_2 : Statistics=-0.113, p-value=0.033
 Плотность, кг/м3 & Количество отвердителя, м.% : Statistics=-0.106, p-value=0.046
 модуль упругости, ГПа & Температура вспышки, C_2 : Statistics=0.110, p-value=0.036

Проверка по критерию Спирмена показала, что статистически значимы следующие корреляции:

Поверхностная плотность, г/м2 & Прочность при растяжении, МПа : Statistics=0.183, p-value=0.000
 Поверхностная плотность, г/м2 & Угол нашивки, град : Statistics=0.122, p-value=0.021
 Модуль упругости при растяжении, ГПа & Прочность при растяжении, МПа : Statistics=0.262, p-value=0.000

По полученным результатам во 1-ом кластере значимость корреляции по Спирмену 'Модуль упругости при растяжении, ГПа' и 'Прочность при растяжении, МПа': (Statistics=0.262, p-value=0.000) является статистически значимой, она представляет интерес с точки зрения физико-механических свойств материалов.

Эта же корреляция оказывается значимой для 2-го (по Спирмену Statistics=0.158, p-value=0.003) и 3-го (по Спирмену Statistics=0.154, p-value=0.007) кластеров.

Далее уже для 1-го кластера и строятся модели машинного обучения.

Результаты расчетов датасета с помощью различных регрессоров, включая многослойный перцептрон, представлены ниже на рисунке 42.

	Model	MAE	R2 score
Прочность при растяжении	RandomForestRegressor_pr	0.142733	0.073
Прочность при растяжении	KNeighborsRegressor_pr	0.149782	0.036
Модуль упругости при растяжении	RandomForestRegressor_upr	0.150505	0.076
Прочность при растяжении	MLPRegressor_pr	0.154457	-0.117
Модуль упругости при растяжении	MLPRegressor_upr	0.155461	-0.020
Модуль упругости при растяжении	KNeighborsRegressor_upr	0.155531	0.009
Прочность при растяжении	SGDRegressor_pr	0.157788	-0.043
Модуль упругости при растяжении	SGDRegressor_upr	0.157829	-0.027
Модуль упругости при растяжении	LinearRegression_upr	0.159663	-0.058
Прочность при растяжении	LinearRegression_pr	0.160595	-0.150

Рисунок 42. Результаты расчетов по 1-му кластеру

Видно, что средние ошибки методов и коэффициенты детерминации варьируются, и результаты расчетов не намного лучше, чем для двух предыдущих примеров, коэффициент детерминации часто оказывается отрицательным.

Но результаты сильно зависят от исходной работы кластеризатора `sklearn.cluster.Kmeans`. Вероятно, при каждом новом варианте кластеризации часть значимых данных датасета попадает то в одни, то в другие кластеры, что довольно сильно влияет на результаты расчета, и это указывает, еще и на то, что их, этих значимых данных не так уж и много в кластерах и, что они сдвинуты к границам кластеров (отражается это даже в описательной статистике). Хотя параметры 'Относительное удлинение при растяжении' и 'Относительное удлинение' не входят в датасет, на котором производится построение моделей машинного обучения, тем не менее качество работы регрессоров иногда несколько лучше, чем предыдущих двух случаях. Вероятно, это можно объяснить и тем, что в построении модели кластеризации используется набор всех параметров исходного датасета, поэтому значимые данные не так сильно раскидываются по разным кластерам.

Результаты расчетов нейронных сетей на `tensorflow.keras` для прогнозирования значений параметров «Модуль упругости при растяжении», «Прочность при растяжении» и «Соотношение матрица-наполнитель» датасета с параметрами 'Относительное удлинение при растяжении' и 'Относительное удлинение' сведены в таблицу 7, архитектура нейронных сетей также аналогична архитектурам с теми же названиями, приведенным в Приложении 3.

Таблица 7. Результаты расчетов нейронных сетей

Параметр	Нейронная сеть	MAE	MAE_abs
Модуль упругости при растяжении	Приложение 3, Model: "sequential_1"	0.1592	0.157204
Прочность при растяжении	Приложение 3, Model: "sequential_4"	0.1612	0.151243
Соотношение матрица-наполнитель	Приложение 3, Model: "sequential_7"	0.1593	0.147031

Показатели MAE моделей нейронных сетей в таблицы 7 плохие, и не позволяют использовать эти модели для прогноза параметров «Модуль упругости при растяжении», «Прочность при растяжении» и «Соотношение матриц-наполнитель», для которых они создавались.

Таким образом, результаты расчета оказались плохие, но иногда чуть лучше, чем в предыдущих двух примерах, вероятно также потому, что кластеризация производится с учетом влияния всего многообразия параметров. Но, поскольку значимые данные, по-видимому, попадают при кластеризации в разные кластеры, то результаты оставляют желать лучшего.

Метод k-ближайших соседей действительно оказывается довольно простым и мощным методом, и который позволяет лучше понять данные, в частности позволяет выявить наличие повторяющихся значений параметров, вокруг которых сгруппированы остальные данные исходного датасета.

Поскольку значимые данные при кластеризации оказываются в разных кластерах, представляется возможным использовать третью стратегию, которая, возможно, позволит выделить значимые данные в один срез.

2.4.3 Модели на основе датасета, подвергнутого разделению на срезы

Третья стратегия осуществляется делением исходного датасета на части (срезы), выделением частей со значимыми корреляционными связями, выделением значимых данных и оценкой этих данных, дополнением данных с помощью аппроксимационных функций с целью получения необходимого количества данных, проверкой корреляционных связей и построением моделей машинного обучения для нового датасета.

Исходный датасет был дополнен расчетными параметрами 'Относительное удлинение при растяжении' и 'Относительное удлинение', см. п. 2.2, и был подвергнут разделению на срезы согласно схеме, представленной на схеме рисунка 43, на которой показаны индексы датасета и обозначения срезов.

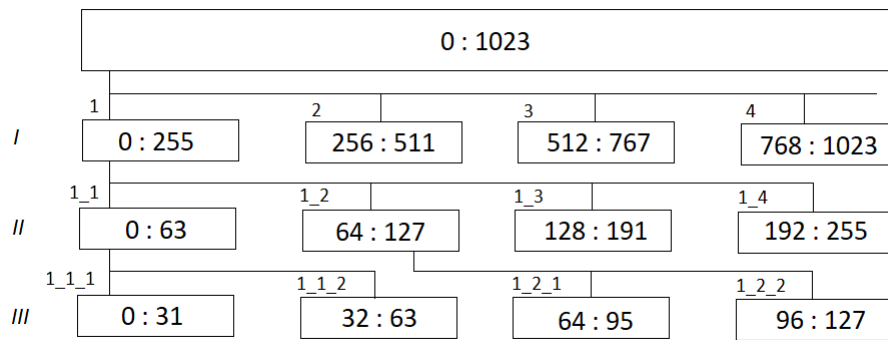


Рисунок 43. Схема получения срезов датасета

Выполнение срезов осуществлялось по мере того, как получались значимые статистические результаты. На каждом уровне по каждому срезу выполнялся следующий статистический анализ: определялось соответствие распределения данных нормальному, осуществлялось построение корреляционных матриц и определение значимых корреляционных связей. Причем для следующего уровня отбирался тот срез датасета, в котором наблюдалось больше значимых корреляционных связей, и в котором корреляционные связи соответствовали существующим известным связям между физико-механическими свойствами материалов.

Кроме того, выполнялась проверка различия групп в нормализованном датасете по критериям ANOVA (параметрический, множественная), Колмогорова-Смирнова (независимые выборки, парная) и Краскала-Уоллиса (независимые выборки, множественная, различие средних), Фридмана (зависимые выборки, множественная, различие средних) на предмет различия между срезами на каждом уровне. На I-ом уровне наблюдаются различия между группами по критерию Колмогорова-Смирнова в срезах `df_dnorm[0 : 255]` и `df_dnorm[256 :]`, `df_dnorm[512 : 767]` и `df_dnorm[0:511].append(df_dnorm[768:])`, `df_dnorm[0 : 767]` и `df_dnorm[768 :]` — только по параметру 'Плотность нашивки', в срезах `df_dnorm[256 : 511]` и `df_dnorm[0:255].append(df_dnorm[512:])` — 'Температура вспышки, C_2', 'Плотность нашивки'.

На II-ом уровне наблюдаются различия между группами по критерию Колмогорова-Смирнова в срезах `df_dnorm[0 : 63]` и `df_dnorm[64 : 255]` — 'Коли-

чество отвердителя, м.%, 'Потребление смолы, г/м²', 'Плотность нашивки', df_dnorm[64 : 127] и df_dnorm[0 : 63].append(df_dnorm[128 :]) — 'Плотность нашивки', df_dnorm[128:191] и df_dnorm[0:127].append(df_dnorm[192:]) — 'Шаг нашивки', 'Плотность нашивки', df_dnorm[0:191] и df_dnorm[192:] — 'Температура вспышки, С_2', 'Плотность нашивки'.

На III-ьем уровне наблюдаются различия между группами по критерию Колмогорова-Смирнова в срезах между группами по критерию Краска-Уоллиса по параметру 'Шаг нашивки', по критерию Колмогорова-Смирнова в срезах df_dnorm[0:31] и df_dnorm[32:63] — 'Содержание эпоксидных групп,%_2', 'Шаг нашивки', df_dnorm[32:63] и df_dnorm[64:95] — не наблюдаются, df_dnorm[64:95] и df_dnorm[96:127] — не наблюдаются.

Таким образом, эти различия касаются параметров 'Температура вспышки, С_2', 'Плотность нашивки', 'Температура вспышки, С_2', 'Содержание эпоксидных групп,%_2', 'Шаг нашивки', что связано, вероятно, с группами (кластерами), по которым производилось формирование данных для исходного датасета по этим параметрам.

В Приложение 5 сведены все имеющиеся данные по значимости статистики срезов, отбор наиболее значимых коэффициентов корреляции для Приложения 5 не производился, поскольку было важным внимательно рассмотреть всю имеющуюся информацию и сделать необходимые заключения.

В частности выделены следующие значимые коэффициенты корреляции:

- срез 1_1_1, по Спирмену (параметр 'Модуль упругости при растяжении, ГПа' имеет вид распределения статистически отличающийся от нормального распределения согласно критерию Шапиро-Уилка):

модуль упругости, ГПа & Модуль упругости при растяжении, ГПа : Statistics=0.455, p-value=0.010

и

Модуль упругости при растяжении, ГПа & Прочность при растяжении, МПа : Statistics=-0.682, p-value=0.000;

- срез 1_2_2, по Пирсону (параметр 'Модуль упругости при растяжении, ГПа' имеет вид распределения статистически не отличающийся от нормального распределения согласно критерию Шапиро-Уилка):

модуль упругости, ГПа & Модуль упругости при растяжении, ГПа : Statistics=-0.382, p-value=0.034.

Теперь представляется возможным сделать заключение, что в срезе 1_1_1 [0 : 32] сосредоточены значимые данные, поскольку именно здесь наблюдаются значимые корреляционные связи с точки зрения физико-механических свойств материалов. Срез 1_2_2 [96 : 127] не содержит таких данных, поскольку параметры 'модуль упругости, ГПа' и 'Модуль упругости при растяжении, ГПа' не могут быть связаны обратной зависимостью с точки зрения физико-механических свойств материалов.

Необходимо построить графики зависимостей между параметрами среза 1_1_1 [0 : 32], некоторые из которых представлены на рисунках 44 — 47 ниже.

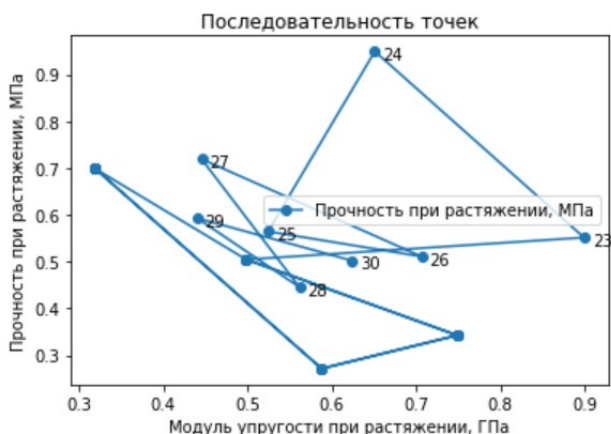


Рисунок 44. Зависимость между параметрами 'Прочность при растяжении, МПа' и 'Модуль упругости при растяжении, ГПа' среза 1_1_1 [0 : 32]

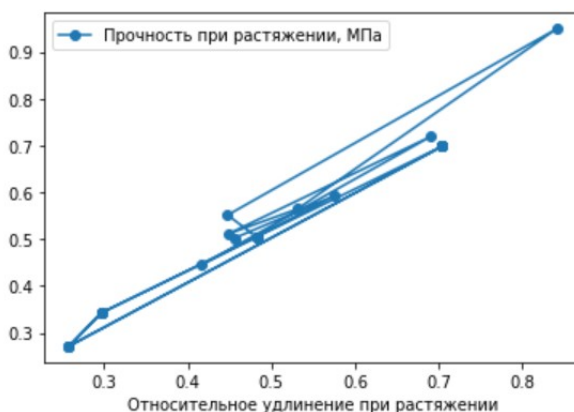


Рисунок 45. Зависимость между параметрами 'Прочность при растяжении, МПа' и 'Относительное удлинение при растяжении' среза 1_1_1 [0 : 32]

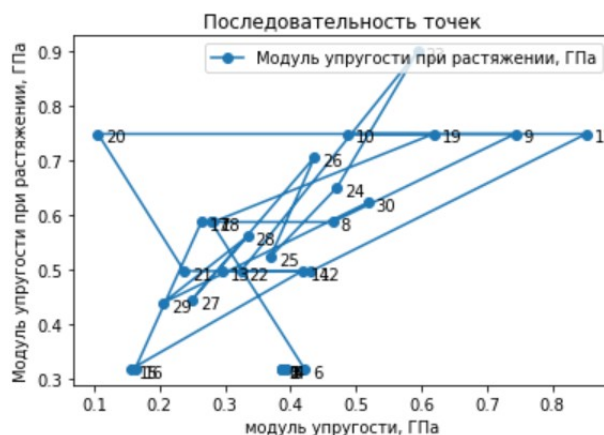


Рисунок 46. Зависимость между параметрами 'Прочность при растяжении, МПа' и 'Модуль упругости при растяжении, ГПа' среза 1_1_1 [0 : 32]

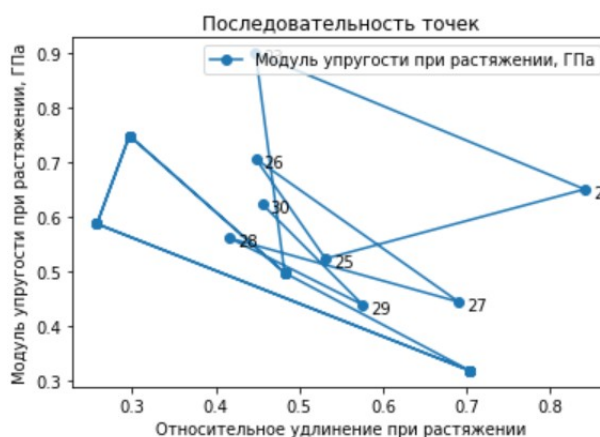


Рисунок 47. Зависимость между параметрами 'Модуль упругости при растяжении, ГПа' и 'Относительное удлинение при растяжении' среза 1_1_1 [0 : 32]

Анализируя зависимости представленные на графиках рисунков 44 — 47, оказывается возможным сделать заключение о том, что значимые данные располагаются в диапазоне [0 : 22] исходного датасета. Величины параметров в этом диапазоне представляют собой те моды вокруг которых происходила кластеризация.

Таким образом, удалось ответить на второй вопрос, касавшийся того, в какой части датасета находятся значимые данные и сколько их всего.

Как только недостоверные данные идентифицированы и отброшены, самым важным шагом подготовки данных становится решение проблемы неполных данных. Для этого необходимо прежде всего знать структуру недостающих данных и тип информации, которую они содержат. Внесение значений взамен недостающих должно осуществляться с особой осторожностью, чтобы не нару-

шить закономерности, скрытые в существующих данных. Данные, нарушающие или искажающие закономерности, называют шумом [7].

Для того, чтобы найти способы заполнения недостающих значений, причиняющих наименьший ущерб структуре набора данных, необходимо найти контекст для новых значений, рассмотреть набор данных в целом и удостовериться, что все методы добавления недостающих значений являются компромиссными с уже имеющимися данными.

Всю доступную информацию о причине отсутствия данных необходимо включать в наборы данных.

Методы аппроксимации данных относятся более к вычислительной математике и не совсем к машинному обучению [1].

Однако, само машинное обучение является результатом работы, решением задачи по экстраполяции данных, то есть, когда у нас есть некоторый набор данных, он ограничен некоторыми пределами (временными пределами, пределами значений и другими) и нужно предсказать, какие могут быть данные за пределами этих ограничений. Выделяют два вида аппроксимации данных — экстраполяция и интерполяция.

Одним из наиболее простых методов интерполяции данных является линейная интерполяция данных. Линейная интерполяция строится ровно по двум точкам, соединяя эти две точки в один общий отрезок. В большинстве случаев, когда пропусков становится очень много, то использовать линейную интерполяцию становится «невыгодно» в том плане, что она даст менее точное значение. Поэтому чуть более улучшенной версией линейной интерполяции является кубическая интерполяция, которая выполняется сплайном.

Диапазон $[0 : 22]$ исходного датасета недостаточен создания моделей машинного обучения. Вероятно, исходный датасет тоже был создан по причине недостаточности экспериментальных данных путем дополнения его искусственными данными. Но, как, установлено в настоящей работе он не подходит для достижения поставленной цели, а именно для создания моделей и обучения алгоритмов машинного обучения, которые будут определять значения:

- Модуль упругости при растяжении, ГПа,
 - Прочность при растяжении, МПа,
- и написание нейронной сети, которая будет рекомендовать:
- Соотношение матрица-наполнитель.

Поэтому в настоящей работе выполнена интерполяция недостающих данных по тем данным, которые находятся в срезе [0: 22] исходного датасета. Предварительно этот срез был превращен в самостоятельный датасет 'X_br&X_nur.xlsx' из 1101 строк, информация о котором приведена на рисунке 48 ниже.

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1101 entries, 0 to 1100
Data columns (total 13 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   Соотношение матрица-наполнитель          23 non-null     float64
1   Плотность, кг/м3                         23 non-null     float64
2   модуль упругости, ГПа                    23 non-null     float64
3   Количество отвердителя, м.%              23 non-null     float64
4   Содержание эпоксидных групп,%_2          23 non-null     float64
5   Температура вспышки, C_2                 23 non-null     float64
6   Поверхностная плотность, г/м2            23 non-null     float64
7   Модуль упругости при растяжении, ГПа     23 non-null     float64
8   Прочность при растяжении, МПа            23 non-null     float64
9   Потребление смолы, г/м2                  23 non-null     float64
10  Угол нашивки, град                       1101 non-null   int64
11  Шаг нашивки                              23 non-null     float64
12  Плотность нашивки                        23 non-null     float64
dtypes: float64(12), int64(1)
memory usage: 120.4 KB
```

Рисунок 48. Информация о датасете подготовленном для интерполяции

Процедура интерполирования подготовленного датасета 'X_br&X_nur.xlsx' методом `scipy.interpolate` представлена на рисунке 49 ниже.

```
1 # интерполяция недостающими данными, метод interpolate библиотеки scipy.interpolate
2 interpolate_columns = ['Соотношение матрица-наполнитель', 'Плотность, кг/м3',
3                       'модуль упругости, ГПа', 'Количество отвердителя, м.%',
4                       'Содержание эпоксидных групп,%_2', 'Температура вспышки, C_2',
5                       'Поверхностная плотность, г/м2', 'Модуль упругости при растяжении, ГПа',
6                       'Прочность при растяжении, МПа', 'Потребление смолы, г/м2', 'Шаг нашивки',
7                       'Плотность нашивки']
8 for col in interpolate_columns:
9     df_s[col] = df_s[col].interpolate(limit_direction = 'both', kind = 'cubic')
```

Рисунок 49. Процедура интерполяции недостающими данными методом `scipy.interpolate.interp1d`

Возникает третий вопрос, как идентифицировать и отбросить недостоверные данные?

Как определить, что полученный датасет с интерполяцией значимых данных не статистически не отличается от тех значимых данных, которые находятся в срезе $[0 : 22]$ исходного датасета?

Следует сначала выполнить построение корреляционных матриц для среза $[0 : 22]$ исходного датасета и для датасета с интерполяцией значимых данных, которые представлены в виде тепловых карт на рисунке 50 ниже.

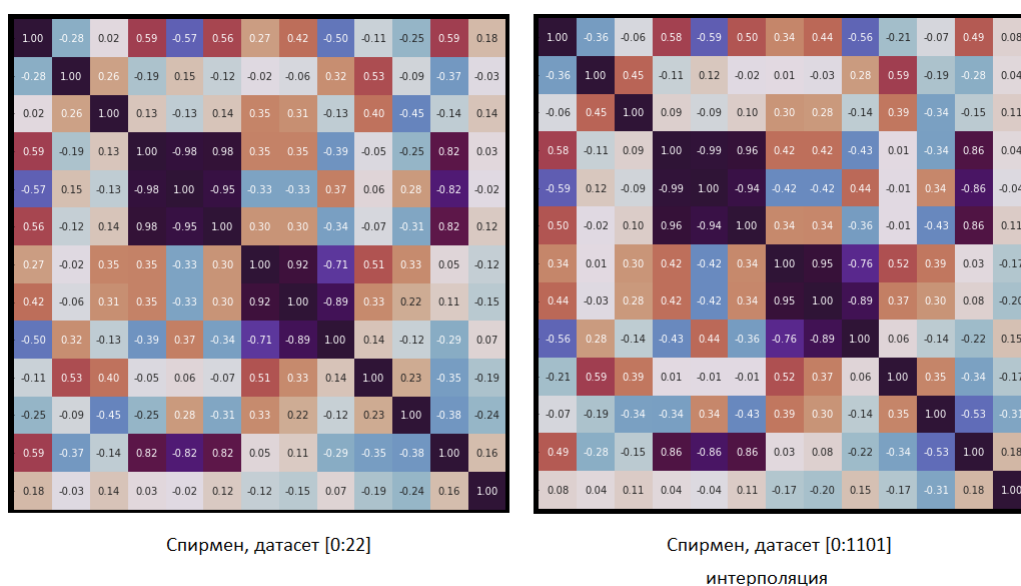


Рисунок 50. Тепловые карты корреляционных матриц среза $[0 : 22]$ исходного датасета и датасета с интерполяцией

Рисунок 50 показывает, что имеющиеся тепловые карты корреляционных матриц обладают большим сходством.

Также проверка статистической гипотезы по критерию Колмогорова-Смирнова показала, что отсутствуют статистически значимые отличия попарно в одноименных столбцах для этих групп (срез $[0 : 22]$ исходного датасета и для датасет с интерполяцией значимых данных) с уровнем значимости 0,05.

Рассматривая полный исходный датасет и датасет с интерполяцией данных как группы, проверка статистической гипотезы по критерию Колмогорова-Смирнова показала, что присутствуют статистические значимые отличия попарно в одноименных столбцах для этих групп.

Для ответа на третий вопрос как идентифицировать принадлежность к недостоверным данным следует выполнить дискриминационный анализ, который относится к методам распознавания образов с обучением (классификации) [8].

Обучение производилось на датасете с интерполяцией. Расчет среднего R2-score показал значение 0.9707961972667855.

Метод показал достаточно уверенное различие данных исходного датасета и датасета с интерполяцией.

Результаты расчетов датасета с интерполяцией с помощью различных регрессоров, включая многослойный перцептрон, представлены ниже на рисунке 51.

	Model	MAE	R2 score
Прочность при растяжении	LinearRegression_pr	2.977435e-16	1.000
Модуль упругости при растяжении	LinearRegression_upr	5.495352e-16	1.000
Модуль упругости при растяжении	SGDRegressor_upr	6.593637e-04	1.000
Прочность при растяжении	SGDRegressor_pr	1.156676e-03	1.000
Прочность при растяжении	KNeighborsRegressor_pr	2.068717e-03	1.000
Модуль упругости при растяжении	KNeighborsRegressor_upr	2.207152e-03	1.000
Прочность при растяжении	MLPRegressor_pr	9.005498e-03	0.999
Модуль упругости при растяжении	RandomForestRegressor_upr	1.304903e-02	0.995
Прочность при растяжении	RandomForestRegressor_pr	1.362307e-02	0.995
Модуль упругости при растяжении	MLPRegressor_upr	1.473170e-02	0.997

Рисунок 51. Результаты расчетов датасета с интерполяцией

Рисунок 51 показывает, что средние ошибки методов и коэффициенты детерминации показывают хорошие результаты расчетов. При расчете параметров «Прочность при растяжении», «Модуль прочности при растяжении» все регрессоры показали себя хорошо, что обусловлено качеством датасета с интерполяцией, коэффициент детерминации равен или близок к единице.

Следует рассмотреть построение нейронных сетей на Keras для прогнозирования значений параметров «Модуль упругости при растяжении», «Прочность при растяжении» и «Соотношение матрица-наполнитель».

Результаты расчетов нейронных сетей на tensorflow.keras для прогнозирования значений параметров «Модуль упругости при растяжении», «Прочность при растяжении» и «Соотношение матрица-наполнитель» датасета сведены в таблицу 8, архитектура нейронных сетей также аналогична архитектурам с теми же названиями, приведенным в Приложении 3.

Таблица 8. Результаты расчетов нейронных сетей

Параметр	Нейронная сеть	MAE	MAE_abs
Модуль упругости при растяжении	Приложение 3, Model: "sequential_1"	0.1263	0.322344
Прочность при растяжении	Приложение 3, Model: "sequential_4"	0.0284	0.332669
Соотношение матрица-наполнитель	Приложение 3, Model: "sequential_7"	0.0464	0.181713

Показатели MAE моделей таблицы 8 даже лучше, чем таблицы 3, особенно для параметров «Прочность при растяжении» и «Соотношение матрица-наполнитель». Это позволяет использовать эти модели нейронной сети для прогноза параметра «Прочность при растяжении» и «Соотношение матрица-наполнитель».

Графики для параметра «Соотношение матрица-наполнитель» нейронной сети Model: "sequential_7" приведены ниже на рисунках 52 — 54.

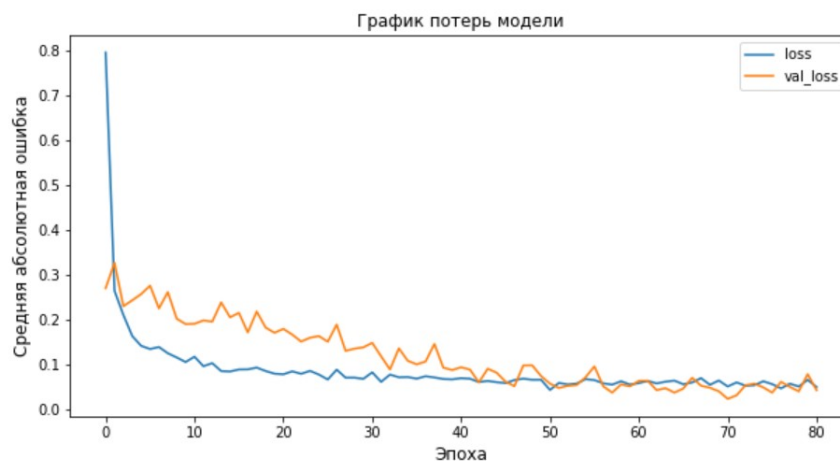


Рисунок 52. График потерь модели сети "sequential_7" параметра «Соотношение матрица-наполнитель»



Рисунок 53. Тестовые и прогнозные значения параметра «Соотношение матрица-наполнитель» sequential_7"

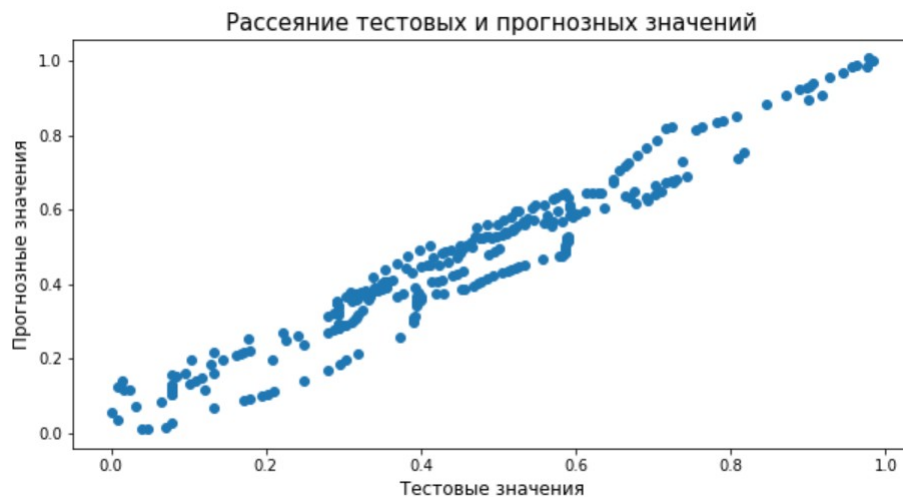


Рисунок 54. Рассеяние тестовых и прогнозных значений параметра «Соотношение матрица-наполнитель» модели сети "sequential_7"

Графики 52 — 54 показывают, что прогнозирование параметра «Соотношение матрица-наполнитель» моделью нейронной сети "sequential_7" достаточно хорошее, что также показывает соответствующее значение MAE таблицы 8.

Графики рассеяния тестовых и прогнозных значений параметра «Соотношение матрица-наполнитель» тоже выглядит достаточно хорошо и показывает, что обучение нейросети возможно.

Графики для параметра «Прочность при растяжении» нейронной сети Model: "sequential_7" приведены ниже на рисунках 55 — 57.



Рисунок 55. График потерь модели сети "sequential_7" параметра «Прочность при растяжении»

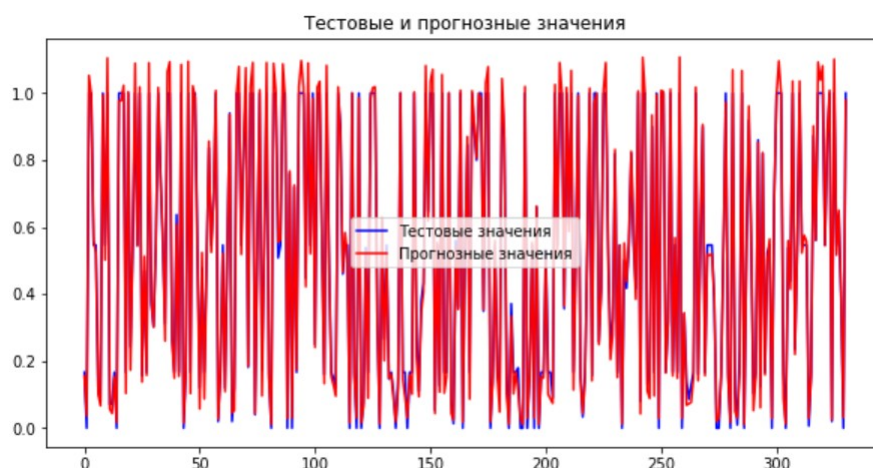


Рисунок 56. Тестовые и прогнозные значения параметра «Прочность при растяжении» sequential_7"

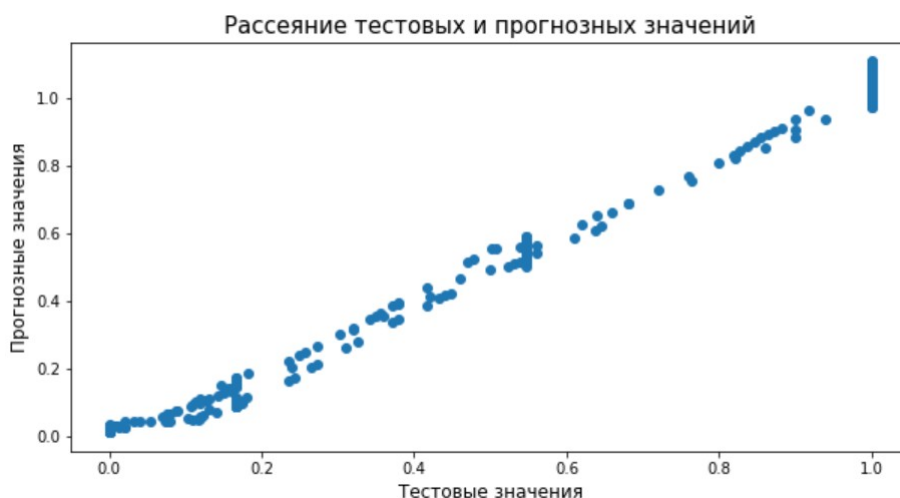


Рисунок 57. Рассеяние тестовых и прогнозных значений параметра «Прочность при растяжении» модели сети "sequential_7"

Графики 55 — 57 показывают, что прогнозирование параметра «Прочность при растяжении» моделью нейронной сети "sequential_7" хорошее, что также показывает соответствующее значение MAE таблицы 8.

Графики рассеяния тестовых и прогнозных значений параметра «Прочность при растяжении» тоже выглядят достаточно и показывает, что обучение нейросети возможно.

Таким образом, улучшение результатов работы регрессоров и моделей нейронной сети в третьей стратегии достигнуто выделением значимых данных из исходного датасета и дополнение недостающих данных путем интерполяции данных датасета до необходимого объема данных.

2.5 Размещение работы в репозитории GitHub

Ссылка на размещение работы в репозитории GitHub:

https://github.com/lksnd/bmstu_fqw

Заключение

Выполнено исследование датасета предоставленного Центром НТИ «Цифровое материаловедение: новые материалы и вещества» (структурное подразделение МГТУ им. Н.Э. Баумана с целью создания моделей и обучения алгоритмов машинного обучения, которые должны определять значения:

- Модуль упругости при растяжении, ГПа,
- Прочность при растяжении, МПа,

и написание нейронной сети, которая будет рекомендовать:

- Соотношение матрица-наполнитель.

Поставленная цель достигнута, разработаны соответствующие модели и обучены алгоритмы машинного обучения.

В процессе изучения исходного датасета были выделены значимые данные и установлены значимые связи между параметрами, которые позволили достаточно эффективно выполнить поставленную цель.

В частности при построении нейронных сетей на Keras для прогнозирования параметров «Модуль упругости при растяжении», «Прочность при растяжении» и «Соотношение матрица-наполнитель» исходного датасета получены показатели MAE 0.1263, 0.0284 и 0.0464 соответственно.

Библиографический список

1. Протодяконов А.В. Алгоритмы Data Science и их практическая реализация на Python : Учеб. пособие / А.В. Протодяконов, П.А. Пылов, В.Е. Садовников. — Москва ; Вологда : Инфра-Инженерия, 2022. — 392 с. : ил.
2. Степин П.А. Сопротивление материалов : Учеб. для вузов / П.А. Степин. — 8-е изд. — М. : Высш. шк., 1988. — 367 с. : ил.
3. Вентцель Е.С. Теория вероятностей : Учеб. для вузов / Е.С. Вентцель. — 6-е изд. стер. — М. : Высш. шк., 1999. — 576 с. : ил.
4. Вадзинский Р.Н. Статистические вычисления в среде Excel. Библиотека пользователя / Р.Н. Вадзинский. — СПб. : Питер, 2008. — 608 с. : ил.
5. Грас Дж. Data Science. Наука о данных с нуля : Пер. с англ. / Дж. Грас. — СПб. : БХВ-Петербург, 2018. — 336 с. : ил.
6. Макшанов А.В. Большие данные. Big Data : учебник для вузов / А.В. Макшанов, А.Е. Журавлев, Л.Н. Тындыкарь. — СПб. : Лань, 2021. — 188 с. : ил.
7. Мыльников Л.А. Статистические методы интеллектуального анализа данных / Л.А. Мыльников. — СПб. : БХВ-Петербург, 2021. — 240 с. : ил.
8. Гайдышев И. Анализ и обработка данных : специальный справочник / И. Гайдышев. — СПб. : Питер, 2001. — 752 с. : ил.
9. https://github.com/Wazzup1/BMSTU_VKR_Pikunov
10. <https://keras.io/api/>
11. <https://matplotlib.org/stable/users/index.html>
12. <https://numpy.org/doc/1.22/user/index.html#user>
13. https://pandas.pydata.org/docs/user_guide/index.html#user-guide
14. https://scikit-learn.org/stable/user_guide.html
15. <https://seaborn.pydata.org/tutorial.html>
16. <https://www.tensorflow.org/overview>
17. <https://docs.python.org/3.8/index.html>
18. <https://flaskrussian-docs.readthedocs.io/ru/latest/quickstart.html>

Приложение 1. Описательная статистика исходного датасета

Соотношение матрица-наполнитель :

Среднее: 2.9303657734325483 Медиана: 2.90687765033521 Мода : 1.85714285714285
Стандарт : 0.913222362148388 Дисперсия: 0.8339748527172307
Асимметрия: 0.08215047390733503 Экссесс: -0.20422037310846086
Минимум: 0.389402605178414 Максимум: 5.59174159869754 Размах: 5.202338993519127
ДоверитИнтервалСредн: (2.8744046078183585, 2.9863269390467426)

Плотность, кг/м³ :

Среднее: 1975.7348881101545 Медиана: 1977.62165679058 Мода : 2030.0
Стандарт : 73.72923055065388 Дисперсия: 5435.999437591473
Асимметрия: 0.05854323312878896 Экссесс: 0.05745882123102097
Минимум: 1731.764635096 Максимум: 2207.77348061119 Размах: 476.00884551518993
ДоверитИнтервалСредн: (1971.2168491073446, 1980.252927112965)

модуль упругости, ГПа :

Среднее: 739.9232327560723 Медиана: 739.664327697792 Мода : 738.736842105263
Стандарт : 330.23158056102693 Дисперсия: 109052.89679983404
Асимметрия: 0.11999727729086174 Экссесс: -0.21118318382570234
Минимум: 2.4369087535075 Максимум: 1911.53647700054 Размах: 1909.0995682470325
ДоверитИнтервалСредн: (719.6870367481765, 760.1594287639673)

Количество отвердителя, м.% :

Среднее: 110.57076864736254 Медиана: 110.564839894065 Мода : 129.0
Стандарт : 28.295911288788815 Дисперсия: 800.6585956630063
Асимметрия: -0.08607032706818164 Экссесс: 0.13919209027132995
Минимум: 17.7402745562519 Максимум: 198.953207190451 Размах: 181.2129326341991
ДоверитИнтервалСредн: (108.83682907050817, 112.304708224217)

Содержание эпоксидных групп, %_2 :

Среднее: 22.24438954776773 Медиана: 22.2307437560244 Мода : 21.25
Стандарт : 2.40630129152943 Дисперсия: 5.790285905616203
Асимметрия: 0.05417837966920955 Экссесс: 0.03656688330785274
Минимум: 14.2549854977161 Максимум: 33.0 Размах: 18.7450145022839
ДоверитИнтервалСредн: (22.096934284686235, 22.391844810849207)

Температура вспышки, C_2 :

Среднее: 285.88215135162187 Медиана: 285.896812331237 Мода : 300.0
Стандарт : 40.943259952923306 Дисперсия: 1676.3505355726531
Асимметрия: -0.05644348004065189 Экссесс: 0.18114755485580591
Минимум: 100.0 Максимум: 413.273418243566 Размах: 313.273418243566
ДоверитИнтервалСредн: (283.37319738603856, 288.391105317206)

Поверхностная плотность, г/м² :

Среднее: 482.73183303841853 Медиана: 451.86436518306 Мода : 210.0
Стандарт : 281.314690236661 Дисперсия: 79137.95494294855
Асимметрия: 0.394885855345264 Экссесс: -0.44281878519445605
Минимум: 0.603739925153945 Максимум: 1399.54236233989 Размах: 1398.938622414736
ДоверитИнтервалСредн: (465.4932055217581, 499.97046055507803)

Модуль упругости при растяжении, ГПа :

Среднее: 73.32857125009068 Медиана: 73.2688045943481 Мода : 70.0
Стандарт : 3.118982889469303 Дисперсия: 9.728054264802283
Асимметрия: 0.12709647236217236 Экссесс: -0.16646020713883214
Минимум: 64.0540605597917 Максимум: 82.682051035271 Размах: 18.627990475479308
ДоверитИнтервалСредн: (73.13744371168467, 73.5196987884968)

Прочность при растяжении, МПа :

Среднее: 2466.9228426979025 Медиана: 2459.52452600309 Мода : 3000.0
Стандарт : 485.62800627853596 Дисперсия: 235834.56048206578
Асимметрия: 0.0784493597849788 Экссесс: 0.016178734221863333
Минимум: 1036.85660535 Максимум: 3848.43673187618 Размах: 2811.58012652618
ДоверитИнтервалСредн: (2437.1641397135286, 2496.6815456822765)

Потребление смолы, г/м² :

Среднее: 218.42314367654282 Медиана: 219.198882195134 Мода : 220.0
Стандарт : 59.735930873323504 Дисперсия: 3568.3814373024843
Асимметрия: 0.00628504890441259 Эксцесс: -0.1294260111723915
Минимум: 33.8030255329625 Максимум: 414.590628361534 Размах: 380.7876028285715
ДоверитИнтервалСредн: (214.76259732960335, 222.08369002348195)

Шаг нашивки :

Среднее: 6.899222077675017 Медиана: 6.9161438559491 Мода : 5.0
Стандарт : 2.563467072833882 Дисперсия: 6.571363433503512
Асимметрия: 0.10184777153596856 Эксцесс: -0.018606675940957107
Минимум: 0.0 Максимум: 14.4405218753969 Размах: 14.4405218753969
ДоверитИнтервалСредн: (6.742135883624404, 7.056308271725643)

Плотность нашивки :

Среднее: 57.153929432857645 Медиана: 57.3419198469929 Мода : 57.0
Стандарт : 12.350968798651323 Дисперсия: 152.54643026525852
Асимметрия: -0.22919818106745457 Эксцесс: 0.850635754756762
Минимум: 0.0 Максимум: 103.988901301494 Размах: 103.988901301494
ДоверитИнтервалСредн: (56.39707684755282, 57.91078201816244)

Относительное удлинение при растяжении :

Среднее: 33.704329418866315 Медиана: 33.602706017803406 Мода : 42.857142857142854
Стандарт : 6.793861265250671 Дисперсия: 46.15655089147345
Асимметрия: 0.11486107313698887 Эксцесс: 0.04876076303291166
Минимум: 13.0980051022361 Максимум: 55.351970133490354 Размах: 42.25396503125425
ДоверитИнтервалСредн: (33.28800973199538, 34.12064910573725)

Относительное удлинение :

Среднее: 7.4276126069408 Медиана: 3.4186145457822947 Мода : 4.060986035907667
Стандарт : 41.15500454516032 Дисперсия: 1693.7343991121663
Асимметрия: 21.03443776763639 Эксцесс: 505.6915221825756
Минимум: 0.8819938067646648 Максимум: 1090.5805536748592 Размах: 1089.6985598680944
ДоверитИнтервалСредн: (4.905683186178503, 9.949542027703092)

Приложение 2. Значимость коэффициентов корреляции датасета после очистки методом межквартильного расстояния без параметров 'Относительное удлинение при растяжении' и 'Относительное удлинение'

Статистически значимыми по критерию Пирсона стали следующие попарные корреляции между количественными параметрами:

Плотность, кг/м³ & Прочность при растяжении, МПа : Statistics=-0.076, p-value=0.020

Плотность, кг/м³ & Плотность нашивки : Statistics=0.084, p-value=0.010

Для параметров 'модуль упругости, ГПа', 'Количество отвердителя, м.%', 'Поверхностная плотность, г/м²', 'Модуль упругости при растяжении, ГПа' и 'Потребление смолы, г/м²' проверка по критерию Спирмена показала, что значимы следующие корреляции:

модуль упругости, ГПа & Плотность нашивки : Statistics=0.073, p-value=0.020

Количество отвердителя, м.% & Температура вспышки, С₂ : Statistics=0.087, p-value=0.005

Количество отвердителя, м.% & Модуль упругости при растяжении, ГПа : Statistics=-0.064, p-value=0.040

Количество отвердителя, м.% & Прочность при растяжении, МПа : Statistics=-0.068, p-value=0.030

Содержание эпоксидных групп, %₂ & Модуль упругости при растяжении, ГПа : Statistics=0.062, p-value=0.048

Согласно критерию Спирмена также после очистки статистически значима корреляция между 'Угол нашивки, град' и 'Плотность, кг/м³' и между 'Угол нашивки, град' и 'Плотность нашивки', корреляция между 'Угол нашивки, град' и 'Плотность, кг/м³'

Приложение 3. Архитектура нейронных сетей для предсказания параметров по данным исходного датасета без параметров 'Относительное удлинение при растяжении' и 'Относительное удлинение'

1-ая сеть, прогнозируемый параметр «Модуль упругости при растяжении»

Model: "sequential"

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 128)	1664
batch_normalization (Batch Normalization)	(None, 128)	512
leaky_re_lu (LeakyReLU)	(None, 128)	0
dense_1 (Dense)	(None, 64)	8256
batch_normalization_1 (Batch Normalization)	(None, 64)	256
leaky_re_lu_1 (LeakyReLU)	(None, 64)	0
dense_2 (Dense)	(None, 32)	2080
batch_normalization_2 (Batch Normalization)	(None, 32)	128
leaky_re_lu_2 (LeakyReLU)	(None, 32)	0
dense_3 (Dense)	(None, 1)	33
activation (Activation)	(None, 1)	0
Total params: 12,929		
Trainable params: 12,481		
Non-trainable params: 448		

2-ая сеть, прогнозируемый параметр «Модуль упругости при растяжении»

Model: "sequential_1"

Layer (type)	Output Shape	Param #
dense_4 (Dense)	(None, 128)	1664
batch_normalization_3 (Batch Normalization)	(None, 128)	512
leaky_re_lu_3 (LeakyReLU)	(None, 128)	0
dense_5 (Dense)	(None, 64)	8256

batch_normalization_4 (Batch Normalization)	(None, 64)	256
leaky_re_lu_4 (LeakyReLU)	(None, 64)	0
dense_6 (Dense)	(None, 64)	4160
batch_normalization_5 (Batch Normalization)	(None, 64)	256
leaky_re_lu_5 (LeakyReLU)	(None, 64)	0
dense_7 (Dense)	(None, 32)	2080
batch_normalization_6 (Batch Normalization)	(None, 32)	128
leaky_re_lu_6 (LeakyReLU)	(None, 32)	0
dense_8 (Dense)	(None, 32)	1056
batch_normalization_7 (Batch Normalization)	(None, 32)	128
leaky_re_lu_7 (LeakyReLU)	(None, 32)	0
dense_9 (Dense)	(None, 1)	33
activation_1 (Activation)	(None, 1)	0

Total params: 18,529

Trainable params: 17,889

Non-trainable params: 640

3-ья сеть, прогнозируемый параметр «Модуль упругости при растяжении»

Model: "sequential_2"

Layer (type)	Output Shape	Param #
dense_10 (Dense)	(None, 256)	3328
batch_normalization_8 (Batch Normalization)	(None, 256)	1024
leaky_re_lu_8 (LeakyReLU)	(None, 256)	0
dense_11 (Dense)	(None, 128)	32896
batch_normalization_9 (Batch Normalization)	(None, 128)	512
leaky_re_lu_9 (LeakyReLU)	(None, 128)	0
dense_12 (Dense)	(None, 128)	16512
batch_normalization_10 (Batch Normalization)	(None, 128)	512

leaky_re_lu_10 (LeakyReLU) (None, 128)	0
dense_13 (Dense) (None, 128)	16512
batch_normalization_11 (Batch Normalization) (None, 128)	512
leaky_re_lu_11 (LeakyReLU) (None, 128)	0
dense_14 (Dense) (None, 64)	8256
batch_normalization_12 (Batch Normalization) (None, 64)	256
leaky_re_lu_12 (LeakyReLU) (None, 64)	0
dense_15 (Dense) (None, 64)	4160
batch_normalization_13 (Batch Normalization) (None, 64)	256
leaky_re_lu_13 (LeakyReLU) (None, 64)	0
dense_16 (Dense) (None, 64)	4160
batch_normalization_14 (Batch Normalization) (None, 64)	256
leaky_re_lu_14 (LeakyReLU) (None, 64)	0
dense_17 (Dense) (None, 64)	4160
batch_normalization_15 (Batch Normalization) (None, 64)	256
leaky_re_lu_15 (LeakyReLU) (None, 64)	0
dense_18 (Dense) (None, 32)	2080
batch_normalization_16 (Batch Normalization) (None, 32)	128
leaky_re_lu_16 (LeakyReLU) (None, 32)	0
dense_19 (Dense) (None, 32)	1056
batch_normalization_17 (Batch Normalization) (None, 32)	128
leaky_re_lu_17 (LeakyReLU) (None, 32)	0
dense_20 (Dense) (None, 32)	1056
batch_normalization_18 (Batch Normalization) (None, 32)	128
leaky_re_lu_18 (LeakyReLU) (None, 32)	0
dense_21 (Dense) (None, 1)	33
activation_2 (Activation) (None, 1)	0

Total params: 98,177

Trainable params: 96,193
Non-trainable params: 1,984

1-ая сеть, прогнозируемый параметр «Прочность при растяжении»

Model: "sequential_3"

Layer (type)	Output Shape	Param #
dense_22 (Dense)	(None, 128)	1664
batch_normalization_19 (Batch Normalization)	(None, 128)	512
leaky_re_lu_19 (LeakyReLU)	(None, 128)	0
dense_23 (Dense)	(None, 64)	8256
batch_normalization_20 (Batch Normalization)	(None, 64)	256
leaky_re_lu_20 (LeakyReLU)	(None, 64)	0
dense_24 (Dense)	(None, 32)	2080
batch_normalization_21 (Batch Normalization)	(None, 32)	128
leaky_re_lu_21 (LeakyReLU)	(None, 32)	0
dense_25 (Dense)	(None, 1)	33
activation_3 (Activation)	(None, 1)	0

Total params: 12,929
Trainable params: 12,481
Non-trainable params: 448

2-ая сеть, прогнозируемый параметр «Прочность при растяжении»

Model: "sequential_4"

Layer (type)	Output Shape	Param #
dense_26 (Dense)	(None, 128)	1664
batch_normalization_22 (Batch Normalization)	(None, 128)	512
leaky_re_lu_22 (LeakyReLU)	(None, 128)	0
dense_27 (Dense)	(None, 64)	8256
batch_normalization_23 (Batch Normalization)	(None, 64)	256

leaky_re_lu_23 (LeakyReLU) (None, 64)	0
dense_28 (Dense) (None, 64)	4160
batch_normalization_24 (Batch Normalization) (None, 64)	256
leaky_re_lu_24 (LeakyReLU) (None, 64)	0
dense_29 (Dense) (None, 32)	2080
batch_normalization_25 (Batch Normalization) (None, 32)	128
leaky_re_lu_25 (LeakyReLU) (None, 32)	0
dense_30 (Dense) (None, 1)	33
activation_4 (Activation) (None, 1)	0

Total params: 17,345
Trainable params: 16,769
Non-trainable params: 576

3-ья сеть, прогнозируемый параметр «Прочность при растяжении»

Model: "sequential_5"

Layer (type)	Output Shape	Param #
dense_31 (Dense)	(None, 128)	1664
batch_normalization_26 (Batch Normalization)	(None, 128)	512
leaky_re_lu_26 (LeakyReLU) (None, 128)	0	
dense_32 (Dense)	(None, 64)	8256
batch_normalization_27 (Batch Normalization)	(None, 64)	256
leaky_re_lu_27 (LeakyReLU) (None, 64)	0	
dense_33 (Dense)	(None, 64)	4160
batch_normalization_28 (Batch Normalization)	(None, 64)	256
leaky_re_lu_28 (LeakyReLU) (None, 64)	0	
dense_34 (Dense)	(None, 64)	4160
batch_normalization_29 (Batch Normalization)	(None, 64)	256
leaky_re_lu_29 (LeakyReLU) (None, 64)	0	

dense_35 (Dense)	(None, 32)	2080
batch_normalization_30 (Batch Normalization)	(None, 32)	128
leaky_re_lu_30 (LeakyReLU)	(None, 32)	0
dense_36 (Dense)	(None, 32)	1056
batch_normalization_31 (Batch Normalization)	(None, 32)	128
leaky_re_lu_31 (LeakyReLU)	(None, 32)	0
dense_37 (Dense)	(None, 32)	1056
batch_normalization_32 (Batch Normalization)	(None, 32)	128
leaky_re_lu_32 (LeakyReLU)	(None, 32)	0
dense_38 (Dense)	(None, 1)	33
activation_5 (Activation)	(None, 1)	0

=====
Total params: 24,129

Trainable params: 23,297

Non-trainable params: 832
=====

1-ая сеть, прогнозируемый параметр «Соотношение матрица-наполнитель»

Model: "sequential_6"

Layer (type)	Output Shape	Param #
dense_39 (Dense)	(None, 128)	1664
batch_normalization_33 (Batch Normalization)	(None, 128)	512
leaky_re_lu_33 (LeakyReLU)	(None, 128)	0
dense_40 (Dense)	(None, 64)	8256
batch_normalization_34 (Batch Normalization)	(None, 64)	256
dense_41 (Dense)	(None, 32)	2080
batch_normalization_35 (Batch Normalization)	(None, 32)	128
leaky_re_lu_34 (LeakyReLU)	(None, 32)	0
dense_42 (Dense)	(None, 1)	33
activation_6 (Activation)	(None, 1)	0

Total params: 12,929
Trainable params: 12,481
Non-trainable params: 448

2-ая сеть, прогнозируемый параметр «Соотношение матрица-наполнитель»

Model: "sequential_7"

Layer (type)	Output Shape	Param #
dense_43 (Dense)	(None, 128)	1664
batch_normalization_36 (Batch Normalization)	(None, 128)	512
leaky_re_lu_35 (LeakyReLU)	(None, 128)	0
dense_44 (Dense)	(None, 128)	16512
batch_normalization_37 (Batch Normalization)	(None, 128)	512
dense_45 (Dense)	(None, 64)	8256
batch_normalization_38 (Batch Normalization)	(None, 64)	256
dense_46 (Dense)	(None, 32)	2080
batch_normalization_39 (Batch Normalization)	(None, 32)	128
dense_47 (Dense)	(None, 16)	528
batch_normalization_40 (Batch Normalization)	(None, 16)	64
leaky_re_lu_36 (LeakyReLU)	(None, 16)	0
dense_48 (Dense)	(None, 1)	17
activation_7 (Activation)	(None, 1)	0

Total params: 30,529
Trainable params: 29,793
Non-trainable params: 736

3-ья сеть, прогнозируемый параметр «Соотношение матрица-наполнитель»

Model: "sequential_8"

Layer (type)	Output Shape	Param #
dense_49 (Dense)	(None, 128)	1664

batch_normalization_41 (Batch Normalization)	(None, 128)	512
leaky_re_lu_37 (LeakyReLU)	(None, 128)	0
dense_50 (Dense)	(None, 128)	16512
batch_normalization_42 (Batch Normalization)	(None, 128)	512
dense_51 (Dense)	(None, 64)	8256
batch_normalization_43 (Batch Normalization)	(None, 64)	256
dense_52 (Dense)	(None, 64)	4160
batch_normalization_44 (Batch Normalization)	(None, 64)	256
dense_53 (Dense)	(None, 32)	2080
batch_normalization_45 (Batch Normalization)	(None, 32)	128
dense_54 (Dense)	(None, 32)	1056
batch_normalization_46 (Batch Normalization)	(None, 32)	128
dense_55 (Dense)	(None, 16)	528
batch_normalization_47 (Batch Normalization)	(None, 16)	64
dense_56 (Dense)	(None, 16)	272
batch_normalization_48 (Batch Normalization)	(None, 16)	64
leaky_re_lu_38 (LeakyReLU)	(None, 16)	0
dense_57 (Dense)	(None, 1)	17
activation_8 (Activation)	(None, 1)	0

Total params: 36,465

Trainable params: 35,505

Non-trainable params: 960

Приложение 4. Описательная статистика кластеров

кластер 1, датасет ddf_0d

Соотношение матрица-наполнитель :

Среднее: 2.8951776773922355 Медиана: 2.89683591198512 Мода : 2.56147540983606
Стандарт : 0.9121910070217526 Дисперсия: 0.8320924332913591
Асимметрия: 0.03135001947097484 Экссесс: -0.016329463482998463
Минимум: 0.389402605178414 Максимум: 5.42513947440363 Размах: 5.035736869225216
ДоверитИнтервалСредн: (2.788900341395123, 3.001455013389345)

Плотность, кг/м3 :

Среднее: 1977.3866696683915 Медиана: 1977.68827695531 Мода : 1880.0
Стандарт : 71.37916775492207 Дисперсия: 5094.9855893853055
Асимметрия: 0.18654838561817982 Экссесс: -0.07566047394531239
Минимум: 1807.59591013457 Максимум: 2192.29763744164 Размах: 384.70172730706986
ДоверитИнтервалСредн: (1969.0704423253378, 1985.7028970114497)

модуль упругости, ГПа :

Среднее: 471.4998946880096 Медиана: 459.047069741573 Мода : 535.0
Стандарт : 216.83408004207092 Дисперсия: 47017.018267691215
Асимметрия: 0.11719582975906488 Экссесс: -0.12100394809170645
Минимум: 4.33915435296057 Максимум: 1144.21605749707 Размах: 1139.8769031441095
ДоверитИнтервалСредн: (446.2370405238422, 496.7627488521773)

Количество отвердителя, м.% :

Среднее: 110.65806263319693 Медиана: 111.86 Мода : 129.0
Стандарт : 28.903369141863834 Дисперсия: 835.4047477508466
Асимметрия: -0.08262245203213545 Экссесс: 0.05311420898331409
Минимум: 32.0192215581845 Максимум: 191.053003718703 Размах: 159.0337821605185
ДоверитИнтервалСредн: (107.29059567716159, 114.02552958923236)

Содержание эпоксидных групп, %_2 :

Среднее: 22.308614180911214 Медиана: 22.2594062112449 Мода : 21.25
Стандарт : 2.351950543765543 Дисперсия: 5.531671360319032
Асимметрия: 0.05330259410159357 Экссесс: -0.5315514848654717
Минимум: 16.3915947268609 Максимум: 28.3249680236444 Размах: 11.933373296783497
ДоверитИнтервалСредн: (22.034593678336375, 22.582634683486074)

Температура вспышки, С_2 :

Среднее: 285.6095374766887 Медиана: 284.615384615384 Мода : 300.0
Стандарт : 43.0329033936117 Дисперсия: 1851.830774483917
Асимметрия: 0.07051711400004537 Экссесс: -0.3986048970471221
Минимум: 186.508612778258 Максимум: 413.273418243566 Размах: 226.76480546530803
ДоверитИнтервалСредн: (280.5958700564999, 290.6232048968773)

Поверхностная плотность, г/м2 :

Среднее: 304.75145752384026 Медиана: 315.234693245364 Мода : 380.0
Стандарт : 160.48410480924824 Дисперсия: 25755.147896425777
Асимметрия: -0.006354609338819783 Экссесс: -0.7327156991719574
Минимум: 0.603739925153945 Максимум: 696.309114521802 Размах: 695.7053745966481
ДоверитИнтервалСредн: (286.0538130637241, 323.44910198395644)

Модуль упругости при растяжении, ГПа :

Среднее: 73.2499801622055 Медиана: 73.33333333333333 Мода : 75.0
Стандарт : 3.0778397381666833 Дисперсия: 9.473097453837958
Асимметрия: -0.06389487448550285 Экссесс: -0.1200710501211419
Минимум: 64.0540605597917 Максимум: 80.4198624060801 Размах: 16.365801846288406

ДоверитИнтервалСредн: (72.8913879315428, 73.60857239286824)

Прочность при растяжении, МПа :

Среднее: 2217.313804835351 Медиана: 2229.63118003497 Мода : 1800.0
Стандарт : 382.0260134288249 Дисперсия: 145943.8749363207
Асимметрия: -0.25528976335159875 Эксцесс: 0.027061744863451587
Минимум: 1036.85660535 Максимум: 3267.6609524823 Размах: 2230.8043471323
ДоверитИнтервалСредн: (2172.804807622834, 2261.822802047863)

Потребление смолы, г/м2 :

Среднее: 216.6969400146487 Медиана: 220.0 Мода : 220.0
Стандарт : 60.32450612258577 Дисперсия: 3639.0460389338873
Асимметрия: -0.22334694767266158 Эксцесс: 0.14694782950863017
Минимум: 41.0482779512307 Максимум: 414.590628361534 Размах: 373.5423504103033
ДоверитИнтервалСредн: (209.66866659745978, 223.72521343183774)

Шаг нашивки :

Среднее: 7.041612487393783 Медиана: 7.07313399156919 Мода : 10.0
Стандарт : 2.5166425879792884 Дисперсия: 6.333489915631089
Асимметрия: -0.09690588556201998 Эксцесс: -0.11592756366219925
Минимум: 0.239660317801428 Максимум: 14.0332153037961 Размах: 13.793554985994671
ДоверитИнтервалСредн: (6.748404082741162, 7.3348208920464)

Плотность нашивки :

Среднее: 56.85383326864423 Медиана: 57.2863442298907 Мода : 47.0
Стандарт : 12.050723749789078 Дисперсия: 145.21994289373055
Асимметрия: 0.04526850885385473 Эксцесс: 0.8816562975324285
Минимум: 15.4160755126792 Максимум: 103.988901301494 Размах: 88.5728257888148
ДоверитИнтервалСредн: (55.44983037110836, 58.2578361661801)

кластер 2, датасет ddf_1d

Соотношение матрица-наполнитель :

Среднее: 3.037817227832622 Медиана: 2.99238771225055 Мода : 1.85714285714285
Стандарт : 0.9161949400001719 Дисперсия: 0.8394131680819185
Асимметрия: 0.13420445961755567 Эксцесс: -0.27191697102756285
Минимум: 0.875457630981136 Максимум: 5.59174159869754 Размах: 4.716283967716405
ДоверитИнтервалСредн: (2.924473075414461, 3.1511613802507785)

Плотность, кг/м3 :

Среднее: 1967.4715647602072 Медиана: 1967.43409316299 Мода : 2030.0
Стандарт : 79.61993058197625 Дисперсия: 6339.3333458787165
Асимметрия: 0.1864142294341815 Эксцесс: 0.13780683063282195
Минимум: 1740.65749569112 Максимум: 2207.77348061119 Размах: 467.1159849200699
ДоверитИнтервалСредн: (1957.621637464817, 1977.3214920555988)

модуль упругости, ГПа :

Среднее: 888.3094354737685 Медиана: 888.191562497692 Мода : 738.736842105263
Стандарт : 251.2917351137937 Дисперсия: 63147.536136501054
Асимметрия: 0.1609419921233092 Эксцесс: 0.401779433425665
Минимум: 11.312942577361 Максимум: 1615.09697286127 Размах: 1603.784030283909
ДоверитИнтервалСредн: (857.2216751264388, 919.3971958210981)

Количество отвердителя, м.% :

Среднее: 108.04738620336254 Медиана: 108.571611076147 Мода : 111.86
Стандарт : 29.507673217090208 Дисперсия: 870.7027786865829
Асимметрия: -0.21883754792260607 Эксцесс: 0.12142165184021092
Минимум: 17.7402745562519 Максимум: 192.705385447053 Размах: 174.9651108908011

ДоверитИнтервалСредн: (104.39693795855732, 111.69783444816784)

Содержание эпоксидных групп, % 2 :

Среднее: 22.316335127952094 Медиана: 22.2822424667293 Мода : 22.2678571428571
Стандарт : 2.4600949125347755 Дисперсия: 6.052066978679485
Асимметрия: 0.20072029040778575 Экссесс: 0.764998639102989
Минимум: 15.8816675122159 Максимум: 33.0 Размах: 17.1183324877841
ДоверитИнтервалСредн: (22.011992285009278, 22.620677970894917)

Температура вспышки, С 2 :

Среднее: 285.0347215081841 Медиана: 284.615384615384 Мода : 284.615384615384
Стандарт : 40.40360859291332 Дисперсия: 1632.4515873293394
Асимметрия: -0.31752467226422304 Экссесс: 1.249030961634344
Минимум: 100.0 Максимум: 397.151290856661 Размах: 297.151290856661
ДоверитИнтервалСредн: (280.0363171613997, 290.03312585496866)

Поверхностная плотность, г/м2 :

Среднее: 266.7014714025315 Медиана: 267.140817459791 Мода : 210.0
Стандарт : 157.95981779215447 Дисперсия: 24951.304036930636
Асимметрия: 0.2713870068058269 Экссесс: -0.6903732268714098
Минимум: 6.77925525877907 Максимум: 709.878674214847 Размах: 703.099418956068
ДоверитИнтервалСредн: (247.15997331941045, 286.24296948565245)

Модуль упругости при растяжении, ГПа :

Среднее: 73.152878152635 Медиана: 72.8596510465653 Мода : 70.0
Стандарт : 3.0096514821535285 Дисперсия: 9.058002044028932
Асимметрия: 0.2554380568802352 Экссесс: -0.07323490877909355
Минимум: 65.7938449666054 Максимум: 82.2375998591107 Размах: 16.443754892505297
ДоверитИнтервалСредн: (72.7805486609237, 73.52520764434632)

Прочность при растяжении, МПа :

Среднее: 2869.2488008064674 Медиана: 2826.72324751235 Мода : 3000.0
Стандарт : 369.6378656829604 Дисперсия: 136632.1517466543
Асимметрия: 0.3714971090470974 Экссесс: -0.1475612710684735
Минимум: 2066.55369198341 Максимум: 3817.26948410351 Размах: 1750.7157921201
ДоверитИнтервалСредн: (2823.5202241047823, 2914.97737750815)

Потребление смолы, г/м2 :

Среднее: 219.97181887305214 Медиана: 220.0 Мода : 220.0
Стандарт : 62.47825471284031 Дисперсия: 3903.5323119625523
Асимметрия: 0.06559294215543778 Экссесс: -0.1343308653378461
Минимум: 33.8030255329625 Максимум: 402.163809219598 Размах: 368.3607836866355
ДоверитИнтервалСредн: (212.24251966409614, 227.70111808200792)

Шаг нашивки :

Среднее: 6.730112499158953 Медиана: 6.63290218991181 Мода : 5.0
Стандарт : 2.5144659490856163 Дисперсия: 6.32253900911103
Асимметрия: 0.4037795017101823 Экссесс: -0.0911159091248912
Минимум: 1.28005723493344 Максимум: 14.3764512763632 Размах: 13.096394041429761
ДоверитИнтервалСредн: (6.419043315800028, 7.0411816825178715)

Плотность нашивки :

Среднее: 57.83713542844637 Медиана: 58.0990914151626 Мода : 60.0
Стандарт : 11.870881016345011 Дисперсия: 140.91781610422035
Асимметрия: -0.2746243463286393 Экссесс: 0.17512895474199075
Минимум: 19.2505331367488 Максимум: 86.0124270098611 Размах: 66.7618938731123
ДоверитИнтервалСредн: (56.36856701738946, 59.30570383950329)

кластер 3, датасет ddf_2d

Соотношение матрица-наполнитель :

Среднее: 2.86183396289099 Медиана: 2.83812171672495 Мода : 2.91967783631954
Стандарт : 0.8985726111604062 Дисперсия: 0.8074327375276307
Асимметрия: -0.00015252762702759959 Экссесс: -0.4498702755506825
Минимум: 0.463342035924085 Максимум: 5.04572461837477 Размах: 4.582382582450685
ДоверитИнтервалСредн: (2.7534414450538307, 2.9702264807281455)

Плотность, кг/м3 :

Среднее: 1973.5479034147322 Медиана: 1976.6062266141298 Мода : 2160.0
Стандарт : 71.7853894459957 Дисперсия: 5153.142137913272
Асимметрия: -0.22271650634040602 Экссесс: 0.1062396834635373
Минимум: 1731.764635096 Максимум: 2160.0 Размах: 428.2353649040001
ДоверитИнтервалСредн: (1964.8886153290166, 1982.2071915004492)

модуль упругости, ГПа :

Среднее: 624.2568798200306 Медиана: 630.0575352243875 Мода : 933.0
Стандарт : 275.74212091114845 Дисперсия: 76033.71724457841
Асимметрия: -0.05614084610676067 Экссесс: -0.31191215380164694
Минимум: 2.4369087535075 Максимум: 1315.30489207828 Размах: 1312.8679833247727
ДоверитИнтервалСредн: (590.9948124165738, 657.5189472234866)

Количество отвердителя, м.% :

Среднее: 110.8629660702628 Медиана: 110.4388711279655 Мода : 111.86
Стандарт : 28.386946574738 Дисперсия: 805.8187358370295
Асимметрия: 0.009774934106983602 Экссесс: 0.007988158469943496
Минимум: 33.6241865036291 Максимум: 198.953207190451 Размах: 165.3290206868219
ДоверитИнтервалСредн: (107.43872124605072, 114.2872108944748)

Содержание эпоксидных групп, %_2 :

Среднее: 22.2327034010069 Медиана: 22.1714503901681 Мода : 22.2678571428571
Стандарт : 2.5097751027127875 Дисперсия: 6.298971066196984
Асимметрия: -0.027404026909761143 Экссесс: -0.10837497901732052
Минимум: 14.2549854977161 Максимум: 28.9074700337636 Размах: 14.652484536047501
ДоверитИнтервалСредн: (21.92052258303658, 22.526018097164787)

Температура вспышки, C_2 :

Среднее: 286.3816742434732 Медиана: 285.22822006188005 Мода : 284.615384615384
Стандарт : 39.72271640000952 Дисперсия: 1577.894198195585
Асимметрия: 0.039334399479369085 Экссесс: 0.12694383609243287
Минимум: 160.255842892284 Максимум: 403.652860876814 Размах: 243.39701798452998
ДоверитИнтервалСредн: (281.5900244731022, 291.1733240138441)

Поверхностная плотность, г/м2 :

Среднее: 832.2349941358942 Медиана: 806.2051499175379 Мода : 1010.0
Стандарт : 165.39746596635527 Дисперсия: 27356.321748091646
Асимметрия: 0.7358814885990831 Экссесс: 0.4359588429890584
Минимум: 530.591316744297 Максимум: 1399.54236233989 Размах: 868.9510455955929
ДоверитИнтервалСредн: (812.283520479439, 852.1864677923488)

Модуль упругости при растяжении, ГПа :

Среднее: 73.48582508603586 Медиана: 73.49599161781765 Мода : 78.0
Стандарт : 2.993975852490928 Дисперсия: 8.963891405298778
Асимметрия: -0.08900844827861973 Экссесс: -0.3936889402919679
Минимум: 65.5533359100584 Максимум: 80.2396088698371 Размах: 14.686272959778705
ДоверитИнтервалСредн: (73.12466942987133, 73.84698074220047)

Прочность при растяжении, МПа :

Среднее: 2572.6542915062655 Медиана: 2576.8116852702897 Мода : 2000.0

Стандарт : 475.15683283961624 Дисперсия: 225774.01579417504
Асимметрия: 0.0017451567943422828 Экссесс: -0.29483236272840685
Минимум: 1379.56851408343 Максимум: 3848.43673187618 Размах: 2468.8682177927503
ДоверитИнтервалСредн: (2515.3373369949263, 2629.971246017603)

Потребление смолы, г/м² :

Среднее: 215.48821197005037 Медиана: 210.4846181379645 Мода : 300.0
Стандарт : 59.98324329402571 Дисперсия: 3597.9894760702805
Асимметрия: 0.16726556972171355 Экссесс: -0.6299351130148474
Минимум: 74.0978656867135 Максимум: 359.052219789673 Размах: 284.95435410295954
ДоверитИнтервалСредн: (208.25258661498648, 222.72383732511432)

Шаг нашивки :

Среднее: 7.008150611869242 Медиана: 7.017358073234686 Мода : 7.0
Стандарт : 2.487610512220022 Дисперсия: 6.18820606050756
Асимметрия: 0.2511134210141128 Экссесс: 0.233507357094473
Минимум: 0.0 Максимум: 14.4405218753969 Размах: 14.4405218753969
ДоверитИнтервалСредн: (6.708076512716023, 7.308224711022449)

Плотность нашивки :

Среднее: 55.1530712353149 Медиана: 55.646861585445095 Мода : 47.0
Стандарт : 12.855055615081866 Дисперсия: 165.25245486684784
Асимметрия: -0.5260934338716593 Экссесс: 1.4976808501522303
Минимум: 0.0 Максимум: 92.9634919477473 Размах: 92.9634919477473
ДоверитИнтервалСредн: (53.602398726772726, 56.70374374385711)

кластер 4, датасет ddf_3d

Соотношение матрица-наполнитель :

Среднее: 2.9351671694679324 Медиана: 2.9277966027972 Мода : 3.55701754385964
Стандарт : 0.9233349606921687 Дисперсия: 0.8525474496364086
Асимметрия: 0.1665828967800721 Экссесс: -0.1770459894373686
Минимум: 0.596782895663544 Максимум: 5.45556617827335 Размах: 4.858783282609806
ДоверитИнтервалСредн: (2.814520284893043, 3.055814054042824)

Плотность, кг/м³ :

Среднее: 1985.4415722854758 Медиана: 1986.9445606882 Мода : 1930.0
Стандарт : 71.28672627223071 Дисперсия: 5081.797342611947
Асимметрия: 0.1342623125955305 Экссесс: -0.12049618902331849
Минимум: 1804.8402302942 Максимум: 2184.49319952038 Размах: 379.6529692261802
ДоверитИнтервалСредн: (1976.12694454752, 1994.7562000234316)

модуль упругости, ГПа :

Среднее: 1047.7222770504497 Медиана: 1027.54525685193 Мода : 889.0
Стандарт : 227.99477332591238 Дисперсия: 51981.61666393416
Асимметрия: 0.7877035339592245 Экссесс: 0.6456870125502503
Минимум: 645.814748630493 Максимум: 1911.53647700054 Размах: 1265.721728370047
ДоверитИнтервалСредн: (1017.9315074249721, 1077.513046675927)

Количество отвердителя, м.% :

Среднее: 112.93310053983897 Медиана: 112.821548651682 Мода : 129.0
Стандарт : 25.895079953696683 Дисперсия: 670.5551658083438
Асимметрия: 0.06461450468792149 Экссесс: 0.24133403156706157
Минимум: 45.1221753550373 Максимум: 192.85170183308 Размах: 147.72952647804271
ДоверитИнтервалСредн: (109.54953893410362, 116.31666214557438)

Содержание эпоксидных групп, %_2 :

Среднее: 22.108129810070924 Медиана: 22.1789141860262 Мода : 21.25

Стандарт : 2.2964772508018174 Дисперсия: 5.273807763450273
Асимметрия: -0.03504560184548409 Экссесс: -0.134152773889594
Минимум: 16.0489783345136 Максимум: 28.8489016191684 Размах: 12.799923284654803
ДоверитИнтервалСредн: (21.808062296545106, 22.40819732359672)

Температура вспышки, С₂ :

Среднее: 286.58428723544523 Медиана: 288.247731867025 Мода : 300.0
Стандарт : 40.50665815538564 Дисперсия: 1640.7893549172695
Асимметрия: -0.060079961122955536 Экссесс: -0.09736959100894627
Минимум: 173.973907031866 Максимум: 396.898222222153 Размах: 222.924315190287
ДоверитИнтервалСредн: (281.2915144942031, 291.87705997668735)

Поверхностная плотность, г/м² :

Среднее: 537.5017553117506 Медиана: 544.203244877878 Мода : 1010.0
Стандарт : 171.41434247499703 Дисперсия: 29382.87680613557
Асимметрия: -0.03511182540845649 Экссесс: 0.21338331177744108
Минимум: 27.5365275109749 Максимум: 1010.0 Размах: 982.4634724890251
ДоверитИнтервалСредн: (515.1040261327763, 559.8994844907241)

Модуль упругости при растяжении, ГПа :

Среднее: 73.43890561739445 Медиана: 73.2842558523516 Мода : 78.0
Стандарт : 3.42645348740439 Дисперсия: 11.740583501345707
Асимметрия: 0.3552591572593527 Экссесс: -0.2192856088483106
Минимум: 66.4823827565976 Максимум: 82.682051035271 Размах: 16.1996682786734
ДоверитИнтервалСредн: (72.99119058872678, 73.88662064606201)

Прочность при растяжении, МПа :

Среднее: 2208.0003526732103 Медиана: 2250.21111996466 Мода : 2000.0
Стандарт : 366.17000727115044 Дисперсия: 134080.47422495438
Асимметрия: -0.5295796554457651 Экссесс: 0.20647540976278922
Минимум: 1071.12375147656 Максимум: 3103.00073322051 Размах: 2031.87698174395
ДоверитИнтервалСредн: (2160.1550175751945, 2255.845687771224)

Потребление смолы, г/м² :

Среднее: 222.31034426545932 Медиана: 220.732220147039 Мода : 300.0
Стандарт : 55.55729779170072 Дисперсия: 3086.6133379157136
Асимметрия: 0.07165616251466589 Экссесс: 0.060279234481498634
Минимум: 64.5241796378105 Максимум: 386.903431430897 Размах: 322.3792517930865
ДоверитИнтервалСредн: (215.05099074878598, 229.56969778213266)

Шаг нашивки :

Среднее: 6.780968234252107 Медиана: 6.93053823619134 Мода : 9.0
Стандарт : 2.7569405250401995 Дисперсия: 7.600721058608931
Асимметрия: -0.06777487599384259 Экссесс: -0.08658000864533832
Минимум: 0.0376389366987437 Максимум: 13.732404403383 Размах: 13.694765466684258
ДоверитИнтервалСредн: (6.420734625145596, 7.141201843358627)

Плотность нашивки :

Среднее: 59.11690309388391 Медиана: 58.9133232111871 Мода : 57.0
Стандарт : 12.355059919841826 Дисперсия: 152.6475056228819
Асимметрия: -0.07331504225854457 Экссесс: 0.2477414481470075
Минимум: 23.8981957702763 Максимум: 98.2026027528147 Размах: 74.3044069825384
ДоверитИнтервалСредн: (57.502538262568976, 60.73126792519888)

Приложение 5. Сводные данные по значимости статистики срезов

Таблица. Статистика по срезам исходного датасета

Статистика по срезам исходного датасета
Срез 1 [0 : 255], уровень I
<p>Наименование колонок с данными БЕЗ нормального распределения : ['Количество отвердителя, м.%', 'Содержание эпоксидных групп,%_2', 'Температура вспышки, C_2', 'Поверхностная плотность, г/м2', 'Потребление смолы, г/м2', 'Относительное удлинение']</p> <p>В 1-ой части датасета по критерию Пирсона значимыми являются корреляции между следующими параметрами:</p> <p>Соотношение матрица-наполнитель & Потребление смолы, г/м2 : Statistics=0.129, p-value=0.040</p> <p>Соотношение матрица-наполнитель & Шаг нашивки : Statistics=0.147, p-value=0.019</p> <p>Плотность, кг/м3 & Поверхностная плотность, г/м2 : Statistics=0.134, p-value=0.033</p> <p>Плотность, кг/м3 & Прочность при растяжении, МПа : Statistics=-0.145, p-value=0.020</p> <p>Плотность, кг/м3 & Шаг нашивки : Statistics=-0.127, p-value=0.042</p> <p>Плотность, кг/м3 & Относительное удлинение при растяжении : Statistics=-0.136, p-value=0.030</p> <p>модуль упругости, ГПа & Количество отвердителя, м.% : Statistics=0.147, p-value=0.019</p> <p>модуль упругости, ГПа & Относительное удлинение : Statistics=-0.366, p-value=0.000</p> <p>Количество отвердителя, м.% & Содержание эпоксидных групп,%_2 : Statistics=-0.131, p-value=0.037</p> <p>Количество отвердителя, м.% & Температура вспышки, C_2 : Statistics=0.168, p-value=0.007</p> <p>Количество отвердителя, м.% & Поверхностная плотность, г/м2 : Statistics=0.139, p-value=0.026</p> <p>Количество отвердителя, м.% & Прочность при растяжении, МПа : Statistics=-0.220, p-value=0.000</p> <p>Количество отвердителя, м.% & Относительное удлинение при растяжении : Statistics=-0.222, p-value=0.000</p> <p>Количество отвердителя, м.% & Относительное удлинение : Statistics=-0.127, p-value=0.043</p> <p>Поверхностная плотность, г/м2 & Модуль упругости при растяжении, ГПа : Statistics=0.141, p-value=0.025</p> <p>Поверхностная плотность, г/м2 & Прочность при растяжении, МПа : Statistics=-0.190, p-value=0.002</p> <p>Поверхностная плотность, г/м2 & Относительное удлинение при растяжении : Statistics=-0.211, p-value=0.001</p> <p>Модуль упругости при растяжении, ГПа & Относительное удлинение при растяжении : Statistics=-0.286, p-value=0.000</p> <p>Прочность при растяжении, МПа & Относительное удлинение при растяжении : Statistics=0.976, p-value=0.000</p> <p>В 1-ой части датасета по критерию Спирмена значимыми являются корреляции между следующими параметрами:</p> <p>Соотношение матрица-наполнитель & Плотность, кг/м3 : Statistics=-0.124, p-value=0.048</p> <p>Плотность, кг/м3 & Шаг нашивки : Statistics=-0.126, p-value=0.044</p> <p>модуль упругости, ГПа & Количество отвердителя, м.% : Statistics=0.182, p-value=0.004</p> <p>модуль упругости, ГПа & Плотность нашивки : Statistics=0.144, p-value=0.022</p> <p>модуль упругости, ГПа & Относительное удлинение : Statistics=-0.925, p-value=0.000</p> <p>Количество отвердителя, м.% & Температура вспышки, C_2 : Statistics=0.131, p-value=0.037</p> <p>Количество отвердителя, м.% & Поверхностная плотность, г/м2 : Statistics=0.143, p-value=0.023</p> <p>Количество отвердителя, м.% & Прочность при растяжении, МПа : Statistics=-0.194, p-value=0.002</p> <p>Количество отвердителя, м.% & Относительное удлинение при растяжении : Statistics=-0.196, p-value=0.002</p> <p>Количество отвердителя, м.% & Относительное удлинение : Statistics=-0.249, p-value=0.000</p> <p>Содержание эпоксидных групп,%_2 & Угол нашивки, град : Statistics=0.126, p-value=0.044</p> <p>Поверхностная плотность, г/м2 & Модуль упругости при растяжении, ГПа : Statistics=0.124, p-value=0.048</p> <p>Поверхностная плотность, г/м2 & Прочность при растяжении, МПа : Statistics=-0.204, p-value=0.001</p> <p>Поверхностная плотность, г/м2 & Относительное удлинение при растяжении : Statistics=-0.213, p-value=0.001</p>

Статистика по срезам исходного датасета
<p>Модуль упругости при растяжении, ГПа & Относительное удлинение при растяжении : Statistics=-0.293, p-value=0.000</p> <p>Прочность при растяжении, МПа & Относительное удлинение при растяжении : Statistics=0.973, p-value=0.000</p> <p>Прочность при растяжении, МПа & Относительное удлинение : Statistics=0.272, p-value=0.000</p> <p>Относительное удлинение при растяжении & Относительное удлинение : Statistics=0.265, p-value=0.000</p>
Срез 2 [256 : 511], уровень I
<p>Наименование колонок с данными БЕЗ нормального распределения : ['Поверхностная плотность, г/м2', 'Относительное удлинение']</p> <p>В 2-ой части датасета по критерию Пирсона значимыми являются корреляции между следующими параметрами:</p> <p>модуль упругости, ГПа & Относительное удлинение : Statistics=-0.283, p-value=0.000</p> <p>Температура вспышки, C_2 & Модуль упругости при растяжении, ГПа : Statistics=-0.149, p-value=0.018</p> <p>Температура вспышки, C_2 & Потребление смолы, г/м2 : Statistics=0.149, p-value=0.018</p> <p>Модуль упругости при растяжении, ГПа & Относительное удлинение при растяжении : Statistics=-0.207, p-value=0.001</p> <p>Прочность при растяжении, МПа & Относительное удлинение при растяжении : Statistics=0.978, p-value=0.000</p> <p>В 2-ой части датасета по критерию Спирмена значимыми являются корреляции между следующими параметрами:</p> <p>модуль упругости, ГПа & Относительное удлинение : Statistics=-0.889, p-value=0.000</p> <p>Температура вспышки, C_2 & Модуль упругости при растяжении, ГПа : Statistics=-0.135, p-value=0.031</p> <p>Температура вспышки, C_2 & Относительное удлинение : Statistics=-0.133, p-value=0.034</p> <p>Модуль упругости при растяжении, ГПа & Относительное удлинение при растяжении : Statistics=-0.177, p-value=0.005</p> <p>Прочность при растяжении, МПа & Относительное удлинение при растяжении : Statistics=0.974, p-value=0.000</p> <p>Прочность при растяжении, МПа & Относительное удлинение : Statistics=0.402, p-value=0.000</p> <p>Относительное удлинение при растяжении & Относительное удлинение : Statistics=0.409, p-value=0.000</p>
Срез 3 [512 : 767], уровень I
<p>Наименование колонок с данными БЕЗ нормального распределения : ['Поверхностная плотность, г/м2', 'Относительное удлинение']</p> <p>В 3-ей части датасета по критерию Пирсона значимыми являются корреляции между следующими параметрами:</p> <p>Соотношение матрица-наполнитель & Прочность при растяжении, МПа : Statistics=0.129, p-value=0.040</p> <p>Соотношение матрица-наполнитель & Шаг нашивки : Statistics=-0.129, p-value=0.040</p> <p>Соотношение матрица-наполнитель & Относительное удлинение при растяжении : Statistics=0.147, p-value=0.019</p> <p>модуль упругости, ГПа & Относительное удлинение : Statistics=-0.219, p-value=0.000</p> <p>Температура вспышки, C_2 & Модуль упругости при растяжении, ГПа : Statistics=0.153, p-value=0.014</p> <p>Поверхностная плотность, г/м2 & Плотность нашивки : Statistics=-0.169, p-value=0.007</p> <p>Модуль упругости при растяжении, ГПа & Относительное удлинение при растяжении : Statistics=-0.234, p-value=0.000</p> <p>Прочность при растяжении, МПа & Шаг нашивки : Statistics=-0.150, p-value=0.016</p> <p>Прочность при растяжении, МПа & Относительное удлинение при растяжении : Statistics=0.980, p-value=0.000</p>

Статистика по срезам исходного датасета	
Шаг нашивки & Относительное удлинение при растяжении : Statistics=-0.153, p-value=0.015	
В 3-ей части датасета по критерию Спирмена значимыми являются корреляции между следующими параметрами:	
Соотношение матрица-наполнитель & модуль упругости, ГПа : Statistics=0.146, p-value=0.019	
Соотношение матрица-наполнитель & Шаг нашивки : Statistics=-0.150, p-value=0.016	
Плотность, кг/м3 & Температура вспышки, C_2 : Statistics=-0.133, p-value=0.034	
модуль упругости, ГПа & Относительное удлинение : Statistics=-0.914, p-value=0.000	
Температура вспышки, C_2 & Модуль упругости при растяжении, ГПа : Statistics=0.150, p-value=0.017	
Поверхностная плотность, г/м2 & Плотность нашивки : Statistics=-0.175, p-value=0.005	
Модуль упругости при растяжении, ГПа & Относительное удлинение при растяжении : Statistics=-0.246, p-value=0.000	
Прочность при растяжении, МПа & Шаг нашивки : Statistics=-0.153, p-value=0.015	
Прочность при растяжении, МПа & Относительное удлинение при растяжении : Statistics=0.974, p-value=0.000	
Прочность при растяжении, МПа & Относительное удлинение : Statistics=0.338, p-value=0.000	
Угол нашивки, град & Плотность нашивки : Statistics=0.160, p-value=0.011	
Шаг нашивки & Относительное удлинение при растяжении : Statistics=-0.159, p-value=0.011	
Относительное удлинение при растяжении & Относительное удлинение : Statistics=0.335, p-value=0.000	
Срез 4 [768 : 1023], уровень I	
Наименование колонок с данными БЕЗ нормального распределения : ['Поверхностная плотность, г/м2', 'Относительное удлинение']	
В 4-ой части датасета по критерию Пирсона значимыми являются корреляции между следующими параметрами:	
модуль упругости, ГПа & Относительное удлинение : Statistics=-0.555, p-value=0.000	
Поверхностная плотность, г/м2 & Прочность при растяжении, МПа : Statistics=0.135, p-value=0.031	
Поверхностная плотность, г/м2 & Потребление смолы, г/м2 : Statistics=0.147, p-value=0.019	
Прочность при растяжении, МПа & Относительное удлинение при растяжении : Statistics=0.976, p-value=0.000	
Прочность при растяжении, МПа & Относительное удлинение : Statistics=0.163, p-value=0.009	
Относительное удлинение при растяжении & Относительное удлинение : Statistics=0.145, p-value=0.021	
В 4-ой части датасета по критерию Спирмена значимыми являются корреляции между следующими параметрами:	
модуль упругости, ГПа & Относительное удлинение : Statistics=-0.911, p-value=0.000	
Поверхностная плотность, г/м2 & Прочность при растяжении, МПа : Statistics=0.124, p-value=0.048	
Модуль упругости при растяжении, ГПа & Относительное удлинение при растяжении : Statistics=-0.137, p-value=0.028	
Прочность при растяжении, МПа & Относительное удлинение при растяжении : Statistics=0.970, p-value=0.000	
Прочность при растяжении, МПа & Относительное удлинение : Statistics=0.314, p-value=0.000	
Относительное удлинение при растяжении & Относительное удлинение : Statistics=0.297, p-value=0.000	
Срез 1_1 [0 : 63], уровень II	
Наименование колонок с данными БЕЗ нормального распределения : ['модуль упругости, ГПа', 'Количество отвердителя, м.%', 'Содержание эпоксидных групп,%_2', 'Температура вспышки, C_2', 'Поверхностная плотность, г/м2', 'Потребление смолы, г/м2', 'Шаг нашивки', 'Плотность нашивки', 'Относительное удлинение']	
В 1_1-ой части датасета по критерию Пирсона значимыми являются корреляции между следующими	

Статистика по срезам исходного датасета

параметрами:

Соотношение матрица-наполнитель & Прочность при растяжении, МПа : Statistics=-0.353, p-value=0.005

Соотношение матрица-наполнитель & Шаг нашивки : Statistics=0.270, p-value=0.032

Соотношение матрица-наполнитель & Относительное удлинение при растяжении : Statistics=-0.355, p-value=0.004

Плотность, кг/м3 & Потребление смолы, г/м2 : Statistics=0.334, p-value=0.008

модуль упругости, ГПа & Количество отвердителя, м.% : Statistics=0.307, p-value=0.014

модуль упругости, ГПа & Поверхностная плотность, г/м2 : Statistics=0.329, p-value=0.009

модуль упругости, ГПа & Модуль упругости при растяжении, ГПа : Statistics=0.305, p-value=0.015

модуль упругости, ГПа & Относительное удлинение : Statistics=-0.832, p-value=0.000

Количество отвердителя, м.% & Температура вспышки, C_2 : Statistics=0.430, p-value=0.000

Количество отвердителя, м.% & Модуль упругости при растяжении, ГПа : Statistics=0.334, p-value=0.007

Количество отвердителя, м.% & Относительное удлинение при растяжении : Statistics=-0.262, p-value=0.038

Поверхностная плотность, г/м2 & Модуль упругости при растяжении, ГПа : Statistics=0.475, p-value=0.000

Поверхностная плотность, г/м2 & Относительное удлинение при растяжении : Statistics=-0.308, p-value=0.014

Модуль упругости при растяжении, ГПа & Прочность при растяжении, МПа : Statistics=-0.342, p-value=0.006

Модуль упругости при растяжении, ГПа & Относительное удлинение при растяжении : Statistics=-0.519, p-value=0.000

Модуль упругости при растяжении, ГПа & Относительное удлинение : Statistics=-0.299, p-value=0.017

Прочность при растяжении, МПа & Относительное удлинение при растяжении : Statistics=0.980, p-value=0.000

Прочность при растяжении, МПа & Относительное удлинение : Statistics=0.288, p-value=0.022

Относительное удлинение при растяжении & Относительное удлинение : Statistics=0.323, p-value=0.010

В 1_1-ой части датасета по критерию Спирмена значимыми являются корреляции между следующими параметрами:

Соотношение матрица-наполнитель & Прочность при растяжении, МПа : Statistics=-0.294, p-value=0.020

Соотношение матрица-наполнитель & Шаг нашивки : Statistics=0.318, p-value=0.011

Соотношение матрица-наполнитель & Относительное удлинение при растяжении : Statistics=-0.310, p-value=0.014

Плотность, кг/м3 & Потребление смолы, г/м2 : Statistics=0.262, p-value=0.038

модуль упругости, ГПа & Количество отвердителя, м.% : Statistics=0.394, p-value=0.001

модуль упругости, ГПа & Модуль упругости при растяжении, ГПа : Statistics=0.294, p-value=0.020

модуль упругости, ГПа & Относительное удлинение : Statistics=-0.905, p-value=0.000

Количество отвердителя, м.% & Температура вспышки, C_2 : Statistics=0.374, p-value=0.003

Количество отвердителя, м.% & Модуль упругости при растяжении, ГПа : Statistics=0.323, p-value=0.010

Количество отвердителя, м.% & Относительное удлинение : Statistics=-0.426, p-value=0.001

Содержание эпоксидных групп, %_2 & Угол нашивки, град : Statistics=0.362, p-value=0.004

Поверхностная плотность, г/м2 & Модуль упругости при растяжении, ГПа : Statistics=0.390, p-value=0.002

Поверхностная плотность, г/м2 & Угол нашивки, град : Statistics=0.303, p-value=0.016

Поверхностная плотность, г/м2 & Относительное удлинение при растяжении : Statistics=-0.267, p-value=0.035

Модуль упругости при растяжении, ГПа & Прочность при растяжении, МПа : Statistics=-0.385, p-value=0.002

Модуль упругости при растяжении, ГПа & Относительное удлинение при растяжении : Statistics=-0.535, p-value=0.000

Статистика по срезам исходного датасета
<p>Модуль упругости при растяжении, ГПа & Относительное удлинение : Statistics=-0.425, p-value=0.001</p> <p>Прочность при растяжении, МПа & Относительное удлинение при растяжении : Statistics=0.967, p-value=0.000</p> <p>Прочность при растяжении, МПа & Относительное удлинение : Statistics=0.371, p-value=0.003</p> <p>Потребление смолы, г/м2 & Относительное удлинение : Statistics=0.249, p-value=0.049</p> <p>Относительное удлинение при растяжении & Относительное удлинение : Statistics=0.408, p-value=0.001</p>
Срез 1_2 [64 : 127], уровень II
<p>Наименование колонок с данными БЕЗ нормального распределения : ['модуль упругости, ГПа', 'Поверхностная плотность, г/м2', 'Потребление смолы, г/м2', 'Относительное удлинение']</p> <p>В 1_2-ой части датасета по критерию Пирсона значимыми являются корреляции между следующими параметрами:</p> <p>Соотношение матрица-наполнитель & Температура вспышки, C_2 : Statistics=-0.314, p-value=0.012</p> <p>Плотность, кг/м3 & Потребление смолы, г/м2 : Statistics=0.332, p-value=0.008</p> <p>модуль упругости, ГПа & Модуль упругости при растяжении, ГПа : Statistics=-0.362, p-value=0.004</p> <p>модуль упругости, ГПа & Плотность нашивки : Statistics=0.253, p-value=0.046</p> <p>модуль упругости, ГПа & Относительное удлинение при растяжении : Statistics=0.270, p-value=0.033</p> <p>модуль упругости, ГПа & Относительное удлинение : Statistics=-0.385, p-value=0.002</p> <p>Количество отвердителя, м.% & Относительное удлинение : Statistics=-0.275, p-value=0.029</p> <p>Поверхностная плотность, г/м2 & Прочность при растяжении, МПа : Statistics=-0.271, p-value=0.032</p> <p>Поверхностная плотность, г/м2 & Шаг нашивки : Statistics=0.248, p-value=0.050</p> <p>Поверхностная плотность, г/м2 & Относительное удлинение при растяжении : Statistics=-0.265, p-value=0.036</p> <p>Модуль упругости при растяжении, ГПа & Относительное удлинение при растяжении : Statistics=-0.297, p-value=0.018</p> <p>Прочность при растяжении, МПа & Относительное удлинение при растяжении : Statistics=0.984, p-value=0.000</p> <p>Шаг нашивки & Относительное удлинение : Statistics=-0.305, p-value=0.015</p> <p>В 1_2-ой части датасета по критерию Спирмена значимыми являются корреляции между следующими параметрами:</p> <p>Соотношение матрица-наполнитель & Температура вспышки, C_2 : Statistics=-0.305, p-value=0.015</p> <p>Соотношение матрица-наполнитель & Модуль упругости при растяжении, ГПа : Statistics=0.250, p-value=0.048</p> <p>Плотность, кг/м3 & Потребление смолы, г/м2 : Statistics=0.264, p-value=0.037</p> <p>модуль упругости, ГПа & Количество отвердителя, м.% : Statistics=0.272, p-value=0.031</p> <p>модуль упругости, ГПа & Модуль упругости при растяжении, ГПа : Statistics=-0.354, p-value=0.004</p> <p>модуль упругости, ГПа & Плотность нашивки : Statistics=0.293, p-value=0.020</p> <p>модуль упругости, ГПа & Относительное удлинение при растяжении : Statistics=0.258, p-value=0.041</p> <p>модуль упругости, ГПа & Относительное удлинение : Statistics=-0.934, p-value=0.000</p> <p>Количество отвердителя, м.% & Относительное удлинение : Statistics=-0.328, p-value=0.009</p> <p>Поверхностная плотность, г/м2 & Прочность при растяжении, МПа : Statistics=-0.258, p-value=0.041</p> <p>Поверхностная плотность, г/м2 & Относительное удлинение при растяжении : Statistics=-0.265, p-value=0.036</p> <p>Модуль упругости при растяжении, ГПа & Относительное удлинение при растяжении : Statistics=-0.287, p-value=0.023</p> <p>Модуль упругости при растяжении, ГПа & Относительное удлинение : Statistics=0.329, p-value=0.008</p> <p>Прочность при растяжении, МПа & Относительное удлинение при растяжении : Statistics=0.977, p-value=0.000</p> <p>Плотность нашивки & Относительное удлинение : Statistics=-0.252, p-value=0.046</p>

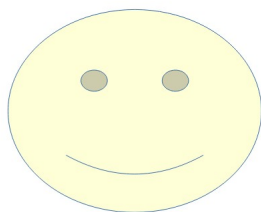
Статистика по срезам исходного датасета
Срез 1_3 [128 : 191], уровень II
<p>Наименование колонок с данными БЕЗ нормального распределения : ['Относительное удлинение']</p> <p>В 1_3-ой части датасета по критерию Пирсона значимыми являются корреляции между следующими параметрами:</p> <p>Соотношение матрица-наполнитель & Потребление смолы, г/м2 : Statistics=0.270, p-value=0.033</p> <p>Соотношение матрица-наполнитель & Относительное удлинение : Statistics=0.300, p-value=0.017</p> <p>Плотность, кг/м3 & Количество отвердителя, м.% : Statistics=-0.251, p-value=0.047</p> <p>Плотность, кг/м3 & Шаг нашивки : Statistics=-0.350, p-value=0.005</p> <p>Плотность, кг/м3 & Плотность нашивки : Statistics=0.412, p-value=0.001</p> <p>модуль упругости, ГПа & Относительное удлинение : Statistics=-0.434, p-value=0.000</p> <p>Прочность при растяжении, МПа & Относительное удлинение при растяжении : Statistics=0.978, p-value=0.000</p> <p>В 1_3-ой части датасета по критерию Спирмена значимыми являются корреляции между следующими параметрами:</p> <p>Соотношение матрица-наполнитель & Потребление смолы, г/м2 : Statistics=0.278, p-value=0.028</p> <p>Плотность, кг/м3 & Шаг нашивки : Statistics=-0.346, p-value=0.005</p> <p>Плотность, кг/м3 & Плотность нашивки : Statistics=0.408, p-value=0.001</p> <p>модуль упругости, ГПа & Относительное удлинение : Statistics=-0.894, p-value=0.000</p> <p>Поверхностная плотность, г/м2 & Потребление смолы, г/м2 : Statistics=-0.248, p-value=0.050</p> <p>Прочность при растяжении, МПа & Относительное удлинение при растяжении : Statistics=0.974, p-value=0.000</p> <p>Прочность при растяжении, МПа & Относительное удлинение : Statistics=0.292, p-value=0.020</p> <p>Относительное удлинение при растяжении & Относительное удлинение : Statistics=0.265, p-value=0.036</p>
Срез 1_4 [192 : 255], уровень II
<p>Наименование колонок с данными БЕЗ нормального распределения : ['Относительное удлинение']</p> <p>В 1_4-ой части датасета по критерию Пирсона значимыми являются корреляции между следующими параметрами:</p> <p>Плотность, кг/м3 & Поверхностная плотность, г/м2 : Statistics=0.402, p-value=0.001</p> <p>Плотность, кг/м3 & Прочность при растяжении, МПа : Statistics=-0.329, p-value=0.009</p> <p>Плотность, кг/м3 & Относительное удлинение при растяжении : Statistics=-0.350, p-value=0.005</p> <p>модуль упругости, ГПа & Шаг нашивки : Statistics=-0.254, p-value=0.045</p> <p>модуль упругости, ГПа & Относительное удлинение : Statistics=-0.549, p-value=0.000</p> <p>Количество отвердителя, м.% & Прочность при растяжении, МПа : Statistics=-0.318, p-value=0.011</p> <p>Количество отвердителя, м.% & Относительное удлинение при растяжении : Statistics=-0.330, p-value=0.008</p> <p>Температура вспышки, C_2 & Потребление смолы, г/м2 : Statistics=0.333, p-value=0.008</p> <p>Модуль упругости при растяжении, ГПа & Относительное удлинение при растяжении : Statistics=-0.260, p-value=0.040</p> <p>Прочность при растяжении, МПа & Относительное удлинение при растяжении : Statistics=0.962, p-value=0.000</p> <p>Потребление смолы, г/м2 & Относительное удлинение : Statistics=0.396, p-value=0.001</p> <p>В 1_4-ой части датасета по критерию Спирмена значимыми являются корреляции между следующими параметрами:</p> <p>Плотность, кг/м3 & Поверхностная плотность, г/м2 : Statistics=0.403, p-value=0.001</p> <p>Плотность, кг/м3 & Прочность при растяжении, МПа : Statistics=-0.326, p-value=0.009</p> <p>Плотность, кг/м3 & Относительное удлинение при растяжении : Statistics=-0.327, p-value=0.009</p> <p>модуль упругости, ГПа & Шаг нашивки : Statistics=-0.296, p-value=0.018</p> <p>модуль упругости, ГПа & Относительное удлинение : Statistics=-0.943, p-value=0.000</p> <p>Количество отвердителя, м.% & Температура вспышки, C_2 : Statistics=0.265, p-value=0.036</p> <p>Количество отвердителя, м.% & Прочность при растяжении, МПа : Statistics=-0.252, p-value=0.046</p>

Статистика по срезам исходного датасета
<p>Количество отвердителя, м.% & Относительное удлинение при растяжении : Statistics=-0.256, p-value=0.043</p> <p>Температура вспышки, C_2 & Потребление смолы, г/м2 : Statistics=0.334, p-value=0.007</p> <p>Прочность при растяжении, МПа & Относительное удлинение при растяжении : Statistics=0.961, p-value=0.000</p> <p>Прочность при растяжении, МПа & Относительное удлинение : Statistics=0.401, p-value=0.001</p> <p>Шаг нашивки & Относительное удлинение : Statistics=0.296, p-value=0.018</p> <p>Относительное удлинение при растяжении & Относительное удлинение : Statistics=0.391, p-value=0.002</p>
Срез 1_1_1 [0 : 31], уровень III
<p>Наименование колонок с данными БЕЗ нормального распределения : ['Количество отвердителя, м.%', 'Содержание эпоксидных групп,%_2', 'Температура вспышки, C_2', 'Поверхностная плотность, г/м2', 'Модуль упругости при растяжении, ГПа', 'Прочность при растяжении, МПа', 'Потребление смолы, г/м2', 'Шаг нашивки', 'Плотность нашивки', 'Относительное удлинение при растяжении', 'Относительное удлинение']</p> <p>В 1_1_1-ой части датасета по критерию Пирсона значимыми являются корреляции между следующими параметрами:</p> <p>Соотношение матрица-наполнитель & Содержание эпоксидных групп,%_2 : Statistics=-0.357, p-value=0.049</p> <p>Соотношение матрица-наполнитель & Прочность при растяжении, МПа : Statistics=-0.441, p-value=0.013</p> <p>Соотношение матрица-наполнитель & Шаг нашивки : Statistics=0.368, p-value=0.042</p> <p>Соотношение матрица-наполнитель & Относительное удлинение при растяжении : Statistics=-0.425, p-value=0.017</p> <p>Плотность, кг/м3 & Потребление смолы, г/м2 : Statistics=0.433, p-value=0.015</p> <p>модуль упругости, ГПа & Поверхностная плотность, г/м2 : Statistics=0.427, p-value=0.017</p> <p>модуль упругости, ГПа & Модуль упругости при растяжении, ГПа : Statistics=0.488, p-value=0.005</p> <p>модуль упругости, ГПа & Относительное удлинение : Statistics=-0.807, p-value=0.000</p> <p>Количество отвердителя, м.% & Содержание эпоксидных групп,%_2 : Statistics=-0.471, p-value=0.007</p> <p>Количество отвердителя, м.% & Температура вспышки, C_2 : Statistics=0.554, p-value=0.001</p> <p>Количество отвердителя, м.% & Модуль упругости при растяжении, ГПа : Statistics=0.452, p-value=0.011</p> <p>Количество отвердителя, м.% & Шаг нашивки : Statistics=0.428, p-value=0.016</p> <p>Поверхностная плотность, г/м2 & Модуль упругости при растяжении, ГПа : Statistics=0.745, p-value=0.000</p> <p>Поверхностная плотность, г/м2 & Прочность при растяжении, МПа : Statistics=-0.426, p-value=0.017</p> <p>Поверхностная плотность, г/м2 & Потребление смолы, г/м2 : Statistics=0.527, p-value=0.002</p> <p>Поверхностная плотность, г/м2 & Плотность нашивки : Statistics=-0.376, p-value=0.037</p> <p>Поверхностная плотность, г/м2 & Относительное удлинение при растяжении : Statistics=-0.516, p-value=0.003</p> <p>Модуль упругости при растяжении, ГПа & Прочность при растяжении, МПа : Statistics=-0.617, p-value=0.000</p> <p>Модуль упругости при растяжении, ГПа & Относительное удлинение при растяжении : Statistics=-0.734, p-value=0.000</p> <p>Модуль упругости при растяжении, ГПа & Относительное удлинение : Statistics=-0.424, p-value=0.017</p> <p>Прочность при растяжении, МПа & Относительное удлинение при растяжении : Statistics=0.987, p-value=0.000</p> <p>Прочность при растяжении, МПа & Относительное удлинение : Statistics=0.366, p-value=0.043</p> <p>Относительное удлинение при растяжении & Относительное удлинение : Statistics=0.403, p-value=0.025</p> <p>В 1_1_1-ой части датасета по критерию Спирмена значимыми являются корреляции между</p>

Статистика по срезам исходного датасета
<p>следующими параметрами:</p> <p>Соотношение матрица-наполнитель & Содержание эпоксидных групп,%_2 : Statistics=-0.479, p-value=0.006</p> <p>Соотношение матрица-наполнитель & Прочность при растяжении, МПа : Statistics=-0.394, p-value=0.028</p> <p>Соотношение матрица-наполнитель & Шаг нашивки : Statistics=0.422, p-value=0.018</p> <p>Соотношение матрица-наполнитель & Относительное удлинение при растяжении : Statistics=-0.405, p-value=0.024</p> <p>модуль упругости, ГПа & Модуль упругости при растяжении, ГПа : Statistics=0.455, p-value=0.010</p> <p>модуль упругости, ГПа & Относительное удлинение : Statistics=-0.796, p-value=0.000</p> <p>Количество отвердителя, м.% & Содержание эпоксидных групп,%_2 : Statistics=-0.406, p-value=0.023</p> <p>Количество отвердителя, м.% & Температура вспышки, C_2 : Statistics=0.443, p-value=0.013</p> <p>Количество отвердителя, м.% & Модуль упругости при растяжении, ГПа : Statistics=0.410, p-value=0.022</p> <p>Количество отвердителя, м.% & Шаг нашивки : Statistics=0.488, p-value=0.005</p> <p>Содержание эпоксидных групп,%_2 & Шаг нашивки : Statistics=-0.654, p-value=0.000</p> <p>Температура вспышки, C_2 & Шаг нашивки : Statistics=0.446, p-value=0.012</p> <p>Поверхностная плотность, г/м2 & Модуль упругости при растяжении, ГПа : Statistics=0.732, p-value=0.000</p> <p>Поверхностная плотность, г/м2 & Прочность при растяжении, МПа : Statistics=-0.428, p-value=0.016</p> <p>Поверхностная плотность, г/м2 & Потребление смолы, г/м2 : Statistics=0.560, p-value=0.001</p> <p>Поверхностная плотность, г/м2 & Угол нашивки, град : Statistics=0.376, p-value=0.037</p> <p>Поверхностная плотность, г/м2 & Относительное удлинение при растяжении : Statistics=-0.468, p-value=0.008</p> <p>Модуль упругости при растяжении, ГПа & Прочность при растяжении, МПа : Statistics=-0.682, p-value=0.000</p> <p>Модуль упругости при растяжении, ГПа & Относительное удлинение при растяжении : Statistics=-0.770, p-value=0.000</p> <p>Модуль упругости при растяжении, ГПа & Относительное удлинение : Statistics=-0.636, p-value=0.000</p> <p>Прочность при растяжении, МПа & Относительное удлинение при растяжении : Statistics=0.961, p-value=0.000</p> <p>Прочность при растяжении, МПа & Относительное удлинение : Statistics=0.618, p-value=0.000</p> <p>Относительное удлинение при растяжении & Относительное удлинение : Statistics=0.643, p-value=0.000</p>
Срез 1_1_2 [32 : 63], уровень III
<p>Наименование колонок с данными БЕЗ нормального распределения : ['Шаг нашивки', 'Плотность нашивки']</p> <p>В 1_1_2-ой части датасета по критерию Пирсона значимыми являются корреляции между следующими параметрами:</p> <p>модуль упругости, ГПа & Количество отвердителя, м.% : Statistics=0.508, p-value=0.004</p> <p>модуль упругости, ГПа & Потребление смолы, г/м2 : Statistics=-0.363, p-value=0.045</p> <p>модуль упругости, ГПа & Шаг нашивки : Statistics=0.363, p-value=0.045</p> <p>модуль упругости, ГПа & Относительное удлинение : Statistics=-0.882, p-value=0.000</p> <p>Количество отвердителя, м.% & Температура вспышки, C_2 : Statistics=0.377, p-value=0.037</p> <p>Количество отвердителя, м.% & Относительное удлинение : Statistics=-0.509, p-value=0.003</p> <p>Прочность при растяжении, МПа & Относительное удлинение при растяжении : Statistics=0.976, p-value=0.000</p> <p>Потребление смолы, г/м2 & Относительное удлинение : Statistics=0.432, p-value=0.015</p> <p>В 1_1_2-ой части датасета по критерию Спирмена значимыми являются корреляции между следующими параметрами:</p> <p>Соотношение матрица-наполнитель & Содержание эпоксидных групп,%_2 : Statistics=0.379, p-</p>

Статистика по срезам исходного датасета
<p>value=0.036</p> <p>модуль упругости, ГПа & Количество отвердителя, м.% : Statistics=0.554, p-value=0.001</p> <p>модуль упругости, ГПа & Потребление смолы, г/м2 : Statistics=-0.421, p-value=0.018</p> <p>модуль упругости, ГПа & Относительное удлинение : Statistics=-0.936, p-value=0.000</p> <p>Количество отвердителя, м.% & Температура вспышки, C_2 : Statistics=0.379, p-value=0.035</p> <p>Количество отвердителя, м.% & Потребление смолы, г/м2 : Statistics=-0.403, p-value=0.024</p> <p>Количество отвердителя, м.% & Относительное удлинение : Statistics=-0.583, p-value=0.001</p> <p>Содержание эпоксидных групп,%_2 & Угол нашивки, град : Statistics=0.437, p-value=0.014</p> <p>Прочность при растяжении, МПа & Относительное удлинение при растяжении : Statistics=0.962, p-value=0.000</p> <p>Потребление смолы, г/м2 & Относительное удлинение : Statistics=0.500, p-value=0.004</p> <p>Угол нашивки, град & Шаг нашивки : Statistics=0.562, p-value=0.001</p>
Срез 1_2_1 [64 : 95], уровень III
<p>Наименование колонок с данными БЕЗ нормального распределения : ['Относительное удлинение']</p> <p>В 1_2_1-ой части датасета по критерию Пирсона значимыми являются корреляции между следующими параметрами:</p> <p>Соотношение матрица-наполнитель & Температура вспышки, C_2 : Statistics=-0.465, p-value=0.008</p> <p>Плотность, кг/м3 & модуль упругости, ГПа : Statistics=-0.441, p-value=0.013</p> <p>модуль упругости, ГПа & Относительное удлинение при растяжении : Statistics=0.363, p-value=0.044</p> <p>модуль упругости, ГПа & Относительное удлинение : Statistics=-0.612, p-value=0.000</p> <p>Поверхностная плотность, г/м2 & Относительное удлинение : Statistics=0.389, p-value=0.031</p> <p>Модуль упругости при растяжении, ГПа & Потребление смолы, г/м2 : Statistics=0.397, p-value=0.027</p> <p>Прочность при растяжении, МПа & Относительное удлинение при растяжении : Statistics=0.989, p-value=0.000</p> <p>В 1_2_1-ой части датасета по критерию Спирмена значимыми являются корреляции между следующими параметрами:</p> <p>Соотношение матрица-наполнитель & Температура вспышки, C_2 : Statistics=-0.473, p-value=0.007</p> <p>Плотность, кг/м3 & модуль упругости, ГПа : Statistics=-0.393, p-value=0.029</p> <p>модуль упругости, ГПа & Относительное удлинение : Statistics=-0.919, p-value=0.000</p> <p>Количество отвердителя, м.% & Плотность нашивки : Statistics=0.360, p-value=0.047</p> <p>Модуль упругости при растяжении, ГПа & Потребление смолы, г/м2 : Statistics=0.381, p-value=0.034</p> <p>Прочность при растяжении, МПа & Относительное удлинение при растяжении : Statistics=0.984, p-value=0.000</p> <p>Потребление смолы, г/м2 & Относительное удлинение : Statistics=0.367, p-value=0.042</p>
Срез 1_2_2 [96 : 127], уровень III
<p>Наименование колонок с данными БЕЗ нормального распределения : ['Потребление смолы, г/м2', 'Относительное удлинение при растяжении', 'Относительное удлинение']</p> <p>В 1_2_2-ой части датасета по критерию Пирсона значимыми являются корреляции между следующими параметрами:</p> <p>Соотношение матрица-наполнитель & Шаг нашивки : Statistics=0.364, p-value=0.044</p> <p>модуль упругости, ГПа & Количество отвердителя, м.% : Statistics=0.509, p-value=0.003</p> <p>модуль упругости, ГПа & Модуль упругости при растяжении, ГПа : Statistics=-0.382, p-value=0.034</p> <p>модуль упругости, ГПа & Шаг нашивки : Statistics=0.441, p-value=0.013</p> <p>модуль упругости, ГПа & Относительное удлинение : Statistics=-0.425, p-value=0.017</p> <p>Количество отвердителя, м.% & Относительное удлинение : Statistics=-0.392, p-value=0.029</p> <p>Прочность при растяжении, МПа & Относительное удлинение при растяжении : Statistics=0.977, p-value=0.000</p>

Статистика по срезам исходного датасета
Шаг нашивки & Относительное удлинение : Statistics=-0.430, p-value=0.016
В 1_2_2-ой части датасета по критерию Спирмена значимыми являются корреляции между следующими параметрами:
модуль упругости, ГПа & Количество отвердителя, м.% : Statistics=0.448, p-value=0.012
модуль упругости, ГПа & Модуль упругости при растяжении, ГПа : Statistics=-0.376, p-value=0.037
модуль упругости, ГПа & Шаг нашивки : Statistics=0.408, p-value=0.023
модуль упругости, ГПа & Относительное удлинение : Statistics=-0.942, p-value=0.000
Количество отвердителя, м.% & Относительное удлинение : Statistics=-0.496, p-value=0.005
Содержание эпоксидных групп,%_2 & Модуль упругости при растяжении, ГПа : Statistics=-0.394, p-value=0.028
Прочность при растяжении, МПа & Относительное удлинение при растяжении : Statistics=0.935, p-value=0.000
Шаг нашивки & Относительное удлинение : Statistics=-0.451, p-value=0.011



Благодарю за
внимание!

