

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ**
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный технический университет имени Н.Э. Баумана
(национальный исследовательский университет)»

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА
по курсу
«Data Science»

**Прогнозирование конечных свойств новых материалов
(композиционных материалов)**

Слушатель

Новиков А.А.

Москва, 2022

Оглавление

Введение.....	3
1 Анализ исходных данных и выбор методов решения	5
1.1 Описание задачи и исходных данных.....	5
1.2 Разведочный анализ и визуализация исходных данных	6
2 Разработка моделей машинного обучения	25
2.1 Предобработка данных.....	25
2.2 Разработка и обучение моделей	26
2.3 Тестирование моделей.....	27
Результаты модели k ближайших соседей для параметра «Прочность при растяжении» (KNeighborsRegressor).	27
Результаты модели повышения градиента «Модуль упругости при растяжении» (GradientBoosting Regressor).....	28
Результаты модели эпсилон-регрессия опорных векторов (SVR).	29
2.5. Разработка нейросети	30
2.6 Разработка приложения	33
3 Заключение	35
4 Список используемой литературы и ссылки на веб-ресурсы.....	36

Введение

Композитные материалы – многокомпонентные материалы, состоящие из двух (или более) компонентов с существенно различающимися физическими и/или химическими свойствами, позволяющими в сочетании достичь синергетический эффект – наличие у составного материала не только свойств компонентов, но и свойств, которыми не обладают его компоненты по отдельности.

В составе композита принято выделять матрицу/матрицы и наполнитель/наполнители. Подбором состава и свойств наполнителя и матрицы, их соотношения, ориентации, можно получить материалы с требуемым сочетанием эксплуатационных и технологических свойств.

Однако, даже при хорошо изученных свойствах отдельных компонентов, определить характеристики состоящего из них композита достаточно проблематично, так как получаемые свойства композита не представляются простой суперпозицией свойств.

Отмечается, что использование феноменологического подхода при расчете и проектировании элементов конструкций из композитов даже одного вида материала, сопряжено с высокой трудоемкостью проведения экспериментальных исследований, соответственно, полученные таким образом экспериментальные данные представляют высокую ценность, а их обработка требует дополнительного анализа.

Вопрос построения математической модели состоит в поиске достаточно надежного описания количественных взаимосвязей между свойствами компонентов композита и композитного материала при различных способах их сочетания.

Современным подходом к решению задач такого типа является применение технологий машинного обучения в целях исследования влияния одной или нескольких независимых переменных на зависимую переменную.

Актуальность решения задачи обусловлена широким использованием различных композитных материалов в отраслях, где создаются какие-либо сооружения, конструкции, техника и т.п., и, соответственно, высоким спросом на

создание различных композитов, обладающих теми или иными требуемыми свойствами.

При этом прогнозные модели могут существенно сократить количество проводимых испытаний, а также пополнить базу данных материалов возможными новыми характеристиками материалов, и цифровыми двойниками новых композитов.

1 Анализ исходных данных и выбор методов решения

1.1 Описание задачи и исходных данных

Предметом настоящей работы является построение при помощи методов машинного обучения моделей прогнозирования характеристик «модуль упругости при растяжении» и «прочность при растяжении», рекомендации «соотношение матрица-наполнитель».

Исходные данные о свойствах композиционных материалов и способах их компоновки получены структурным подразделением МГТУ им. Н.Э. Баумана – Центр НТИ «Цифровое материаловедение: новые материалы и вещества» в рамках решения производственных задач.

Соотношения и свойства используемых компонентов композитов (6 входных переменных вещественного типа), а также интересующие выходные характеристики композитов (3 выходных переменных вещественного типа), представлены в исходных данных таблицей в формате `xlsx`, содержащей 1023 строки данных.

Способы компоновки материалов композитов (3 входных переменных вещественного типа) представлены в исходных данных таблицей формата `xlsx`, содержащей 1040 строк данных.

Обе таблицы исходных данных содержат колонку с целочисленным индексом, не являющимся входным или выходным переменным, служащим для сопоставления таблиц данных.

Данная задача в рамках классификации категорий машинного обучения [2, 380] относится к машинному обучению с учителем, задача регрессии.

Анализ, предобработка данных, построение моделей выполнены посредством языка программирования Python [3] с использованием библиотек Pandas [4], Matplotlib [5] и Sklearn [6].

1.2 Разведочный анализ и визуализация исходных данных

Целями разведочного анализа является получение первоначальных представлений о характерах распределений переменных исходного набора данных, формирование оценки качества исходных данных (наличие пропусков, выбросов), выявление характера взаимосвязи между переменными с целью последующего выдвижения гипотез о наиболее подходящих для решения задачи моделях машинного обучения.

Разведочный анализ данных в рамках данной задачи проведен над датафреймом Pandas, полученным путем импорта (и объединения по типу INNER по полю индекса) таблиц исходных данных, где наименования параметров (для краткости при дальнейшей обработке) переименованы в транслитированные в латиницу аббревиатуры.

Таблица 1 – Сокращения наименований параметров

Исходная таблица	Наименование параметра	Вход / выход
Способы компоновки компонентов композита	Угол нашивки, град	Вход
	Шаг нашивки	Вход
	Плотность нашивки	Вход
Соотношения и свойства компонентов композита	Соотношение матрица-наполнитель	Выход
	Плотность, кг/м ³	Вход
	модуль упругости, ГПа	Вход
	Количество отвердителя, м. %	Вход
	Содержание эпоксидных групп, % 2	Вход
	Температура вспышки, С 2	Вход
	Поверхностная плотность, г/м ²	Вход
	Модуль упругости при растяжении, ГПа	Выход
	Прочность при растяжении, МПа	Выход
	Потребление смолы, г/м ²	Вход

Следует отметить, что часть информации из исходных данных – 17 строк таблицы способов компоновки композитов – не представляется возможным включить в работу в силу того, что они не имеют соответствующих строк в таблице соотношений и свойств используемых компонентов композитов.

Сформированный исходный датафрейм содержит 1023 записи с 9 входными параметрами и 3 выходными параметрами вещественного типа, пропуски значений отсутствуют.

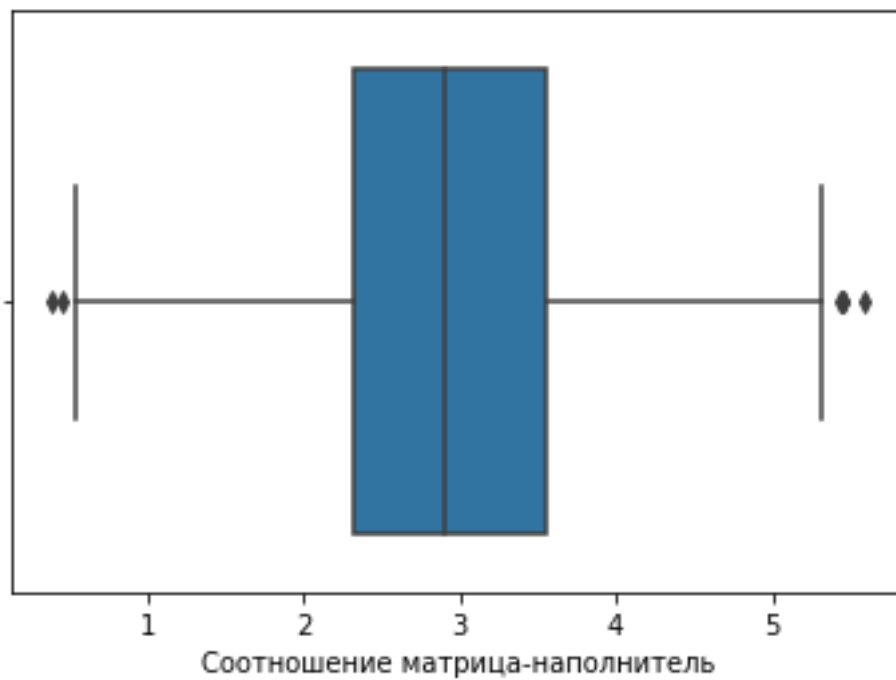
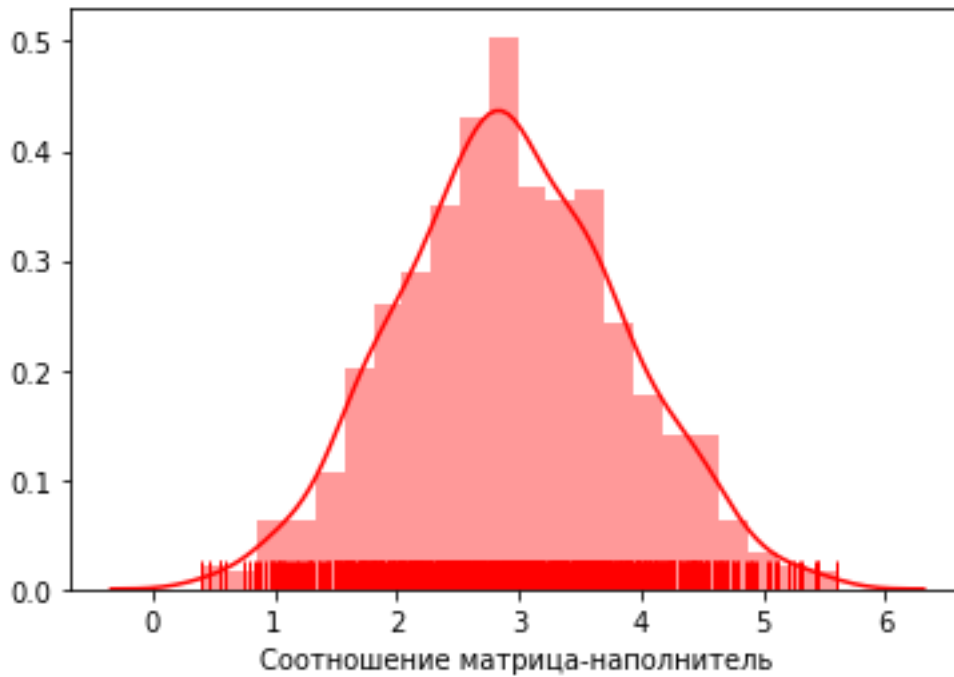
Показатели описательной статистики и визуализация гистограмм и/или диаграмм размаха («ящик с усами») позволяют получить наглядное представление о характерах распределений переменных.

Описательная статистика и диаграммы размаха представлены, соответственно, в таблице и на рисунке ниже.

Таблица 2 – Описательная статистика исходных данных

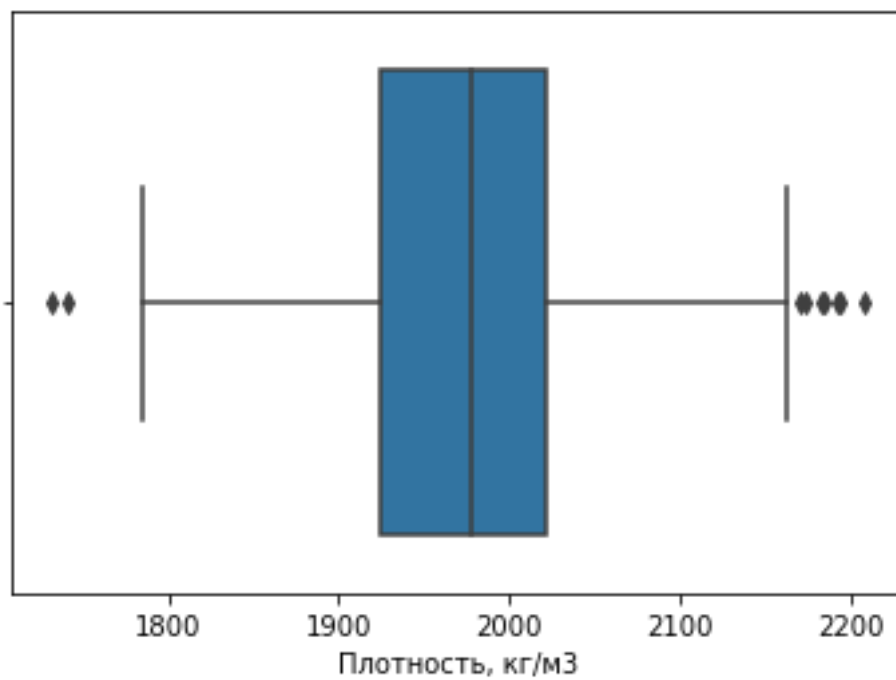
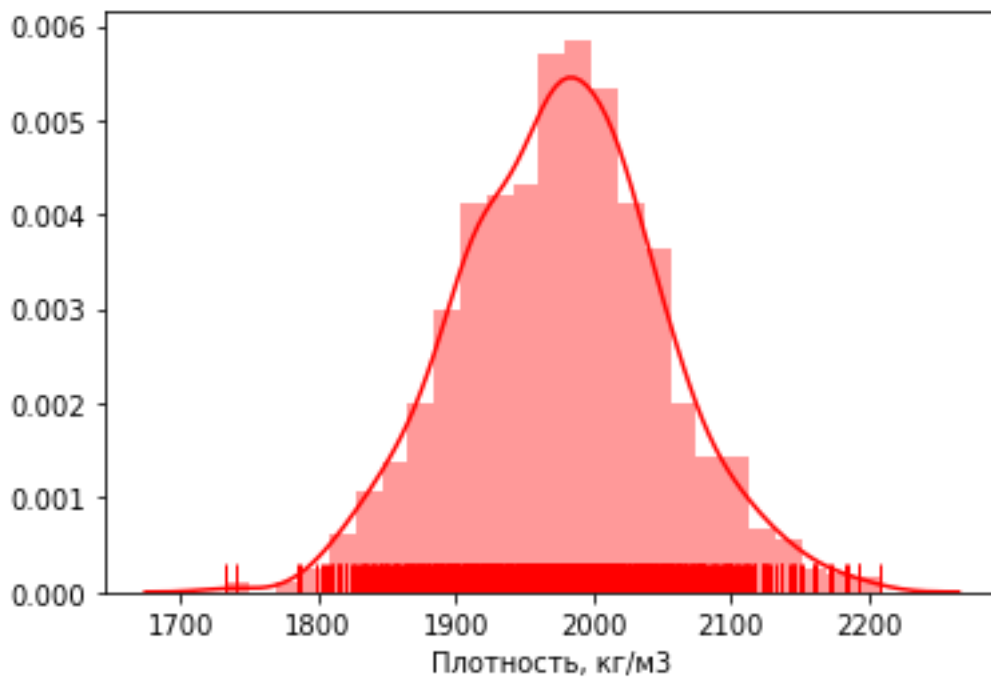
	Соотношение матрица-наполнитель	Плотность, кг/м3	модуль упругости, ГПа	Количество отвердителя, м. %	Содержание эпоксидных групп, %_2	Температура вспышки, С_2	Поверхностная плотность, г/м2	Модуль упругости при растяжении, ГПа	Прочность при растяжении, МПа	Потребление смолы, г/м2	Угол нашивки, град	Шаг нашивки	Плотность нашивки
count	936	936	936	936	936	936	936	936	936	936	936	936	936
mean	2,93	1974,04	738,25	110,92	22,21	286,04	482,99	73,31	2467,49	217,61	46,06	6,92	57,45
std	0,89	70,81	328,71	27,04	2,39	39,40	280,19	3,04	463,84	57,83	45,01	2,51	11,24
min	0,55	1784,48	2,44	38,67	15,70	179,37	0,60	65,55	1250,39	63,69	0	0,04	27,27
25%	2,32	1923,44	498,58	92,52	20,57	259,18	266,00	71,25	2146,94	179,49	0	5,13	50,21
50%	2,90	1977,26	738,74	111,11	22,18	286,02	457,73	73,26	2457,96	218,39	90	6,94	57,58
75%	3,55	2020,16	958,42	130,00	23,96	312,99	695,90	75,31	2755,17	256,40	90	8,59	64,80
max	5,31	2161,57	1649,42	181,83	28,96	386,07	1291,34	81,42	3705,67	359,05	90	13,73	86,01

Гистограммы распределений и диаграммы размаха переменных:



Минимальное значение: 0.389402605178414
 Максимальное значение: 5.59174159869754
 Среднее значение: 2.9303657734325506
 Медианное значение: 2.90687765033521

Плотность, кг/м³



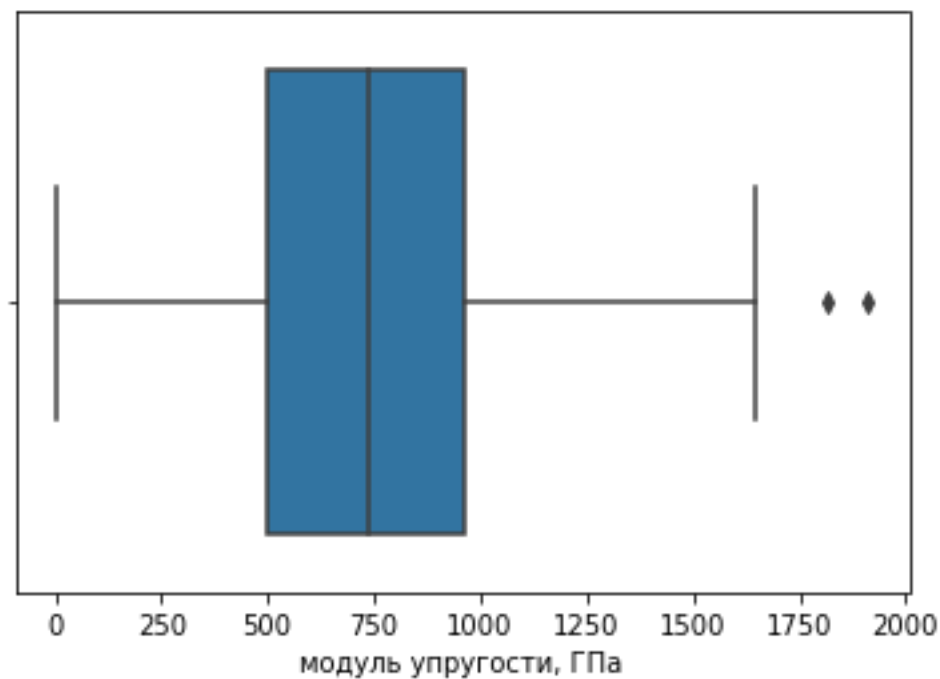
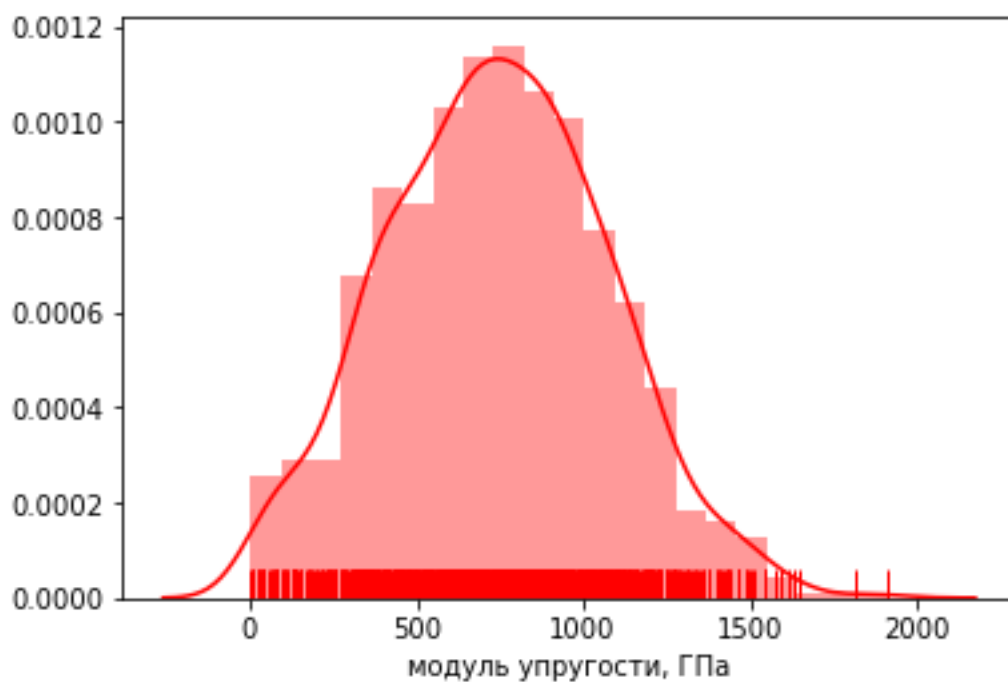
Минимальное значение: 1731.764635096

Максимальное значение: 2207.77348061119

Среднее значение: 1975.7348881101548

Медианное значение: 1977.62165679058

модуль упругости, ГПа



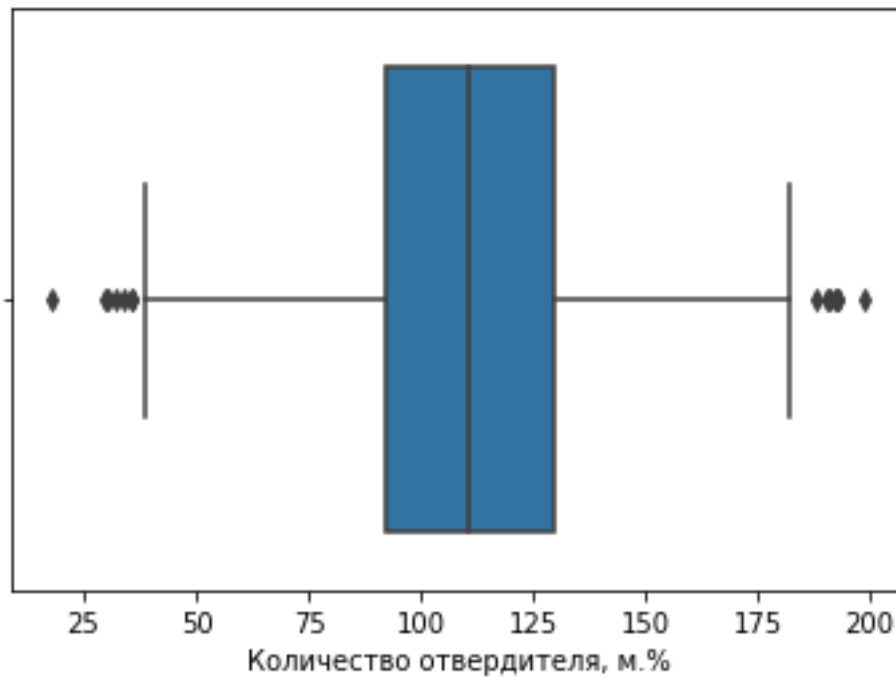
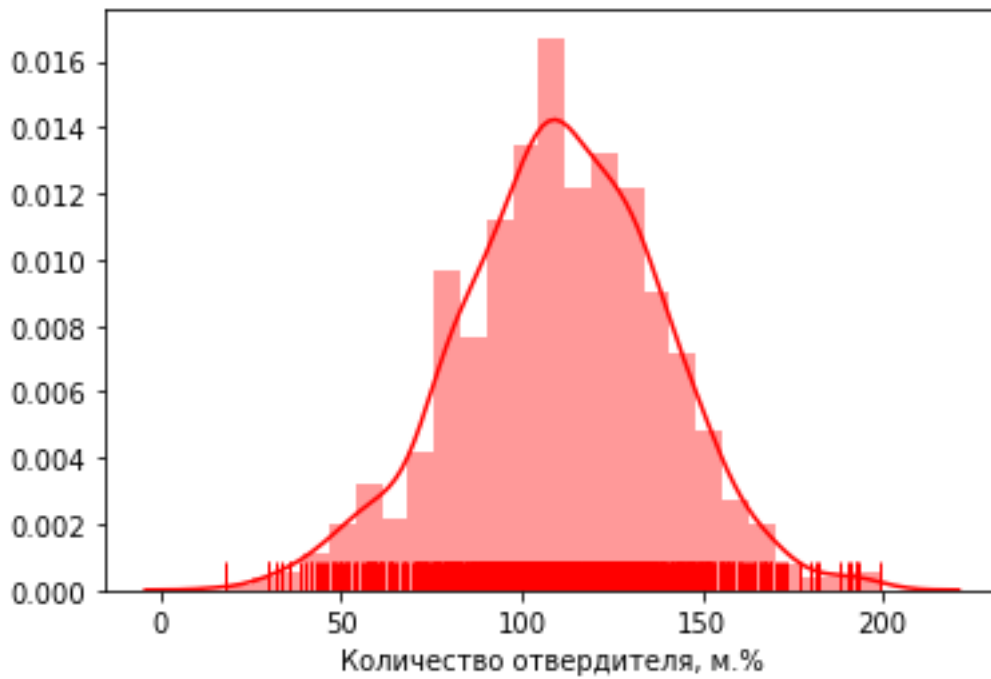
Минимальное значение: 2.4369087535075

Максимальное значение: 1911.53647700054

Среднее значение: 739.9232327560719

Медианное значение: 739.664327697792

Количество отвердителя, м.%



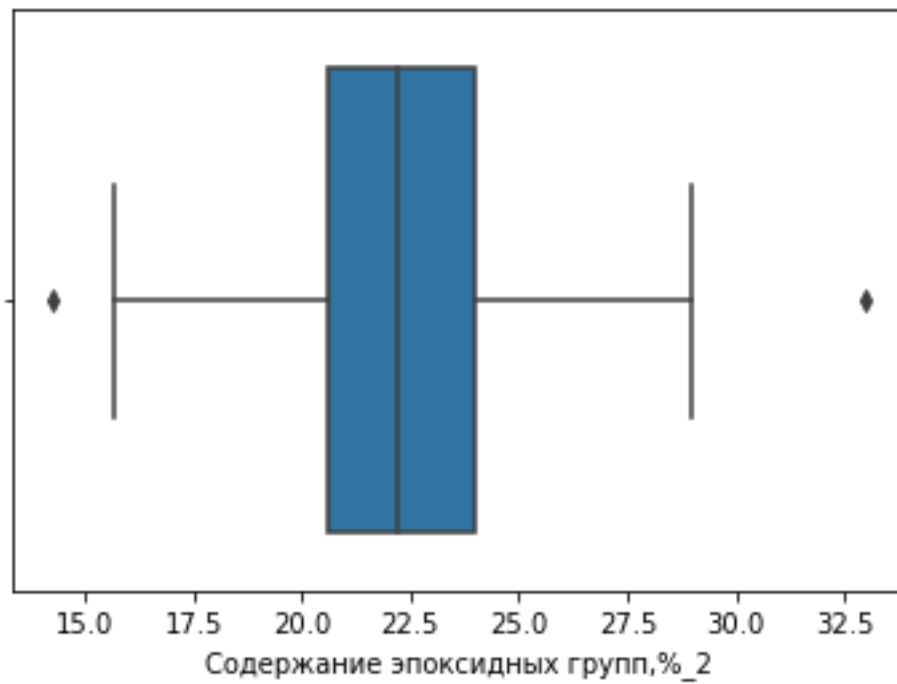
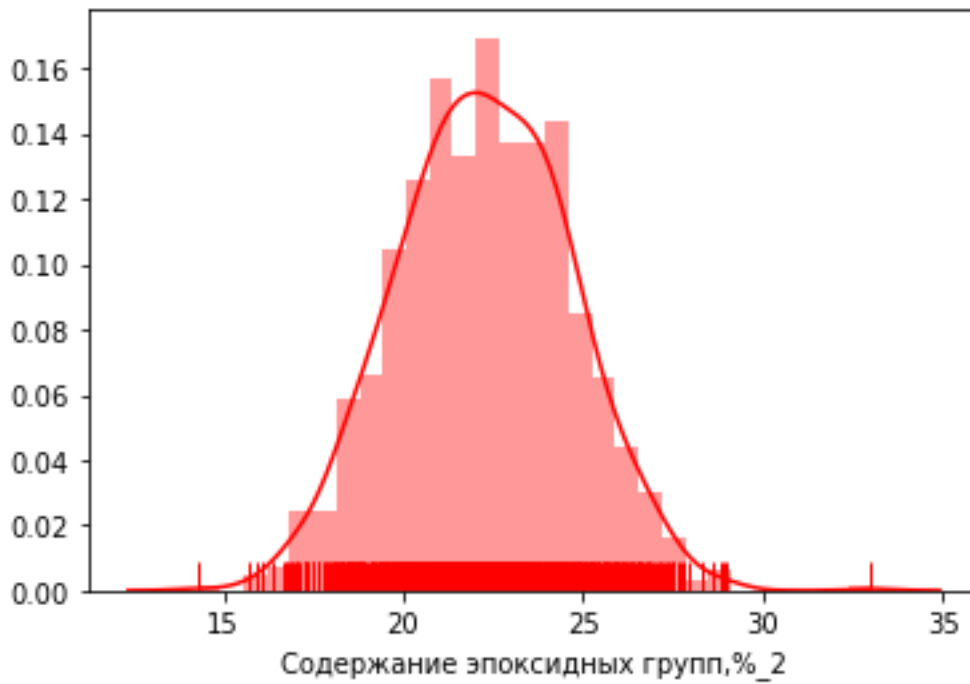
Минимальное значение: 17.7402745562519

Максимальное значение: 198.953207190451

Среднее значение: 110.57076864736258

Медианное значение: 110.564839894065

Содержание эпоксидных групп, %_2



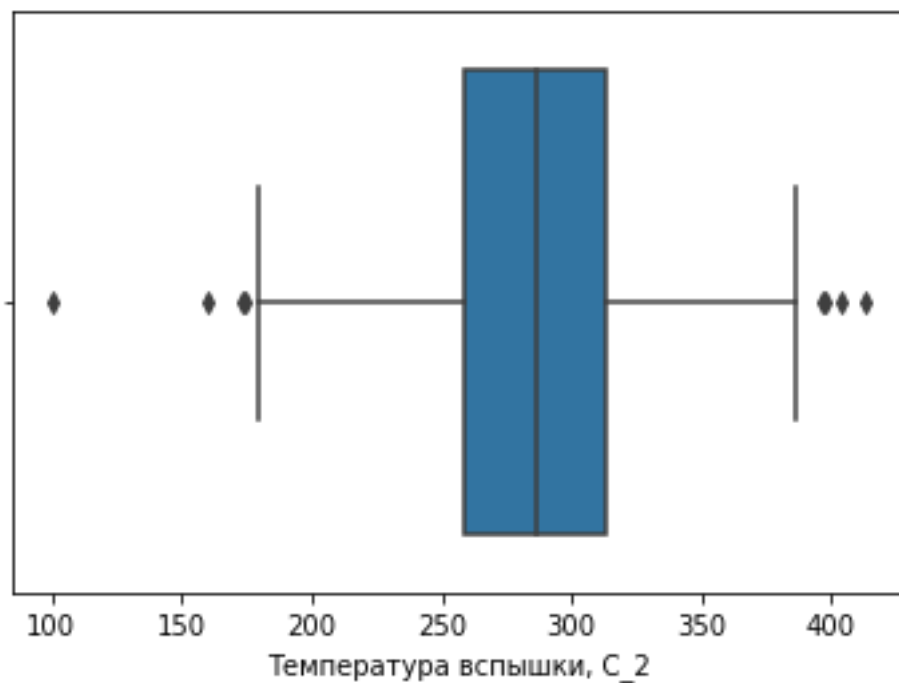
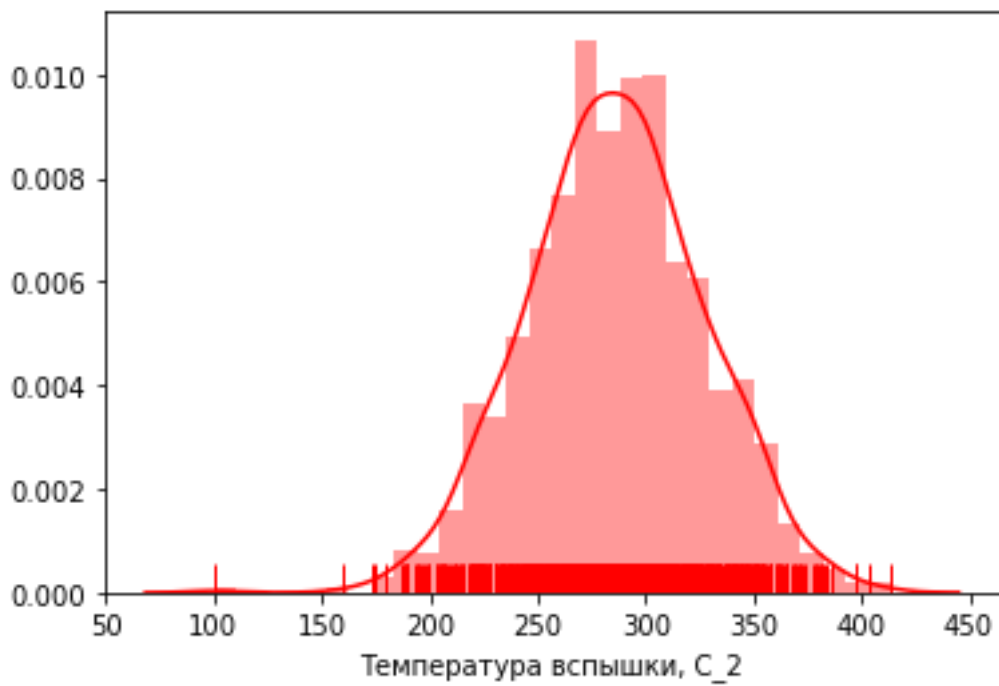
Минимальное значение: 14.2549854977161

Максимальное значение: 33.0

Среднее значение: 22.24438954776772

Медианное значение: 22.2307437560244

Температура вспышки, C_2



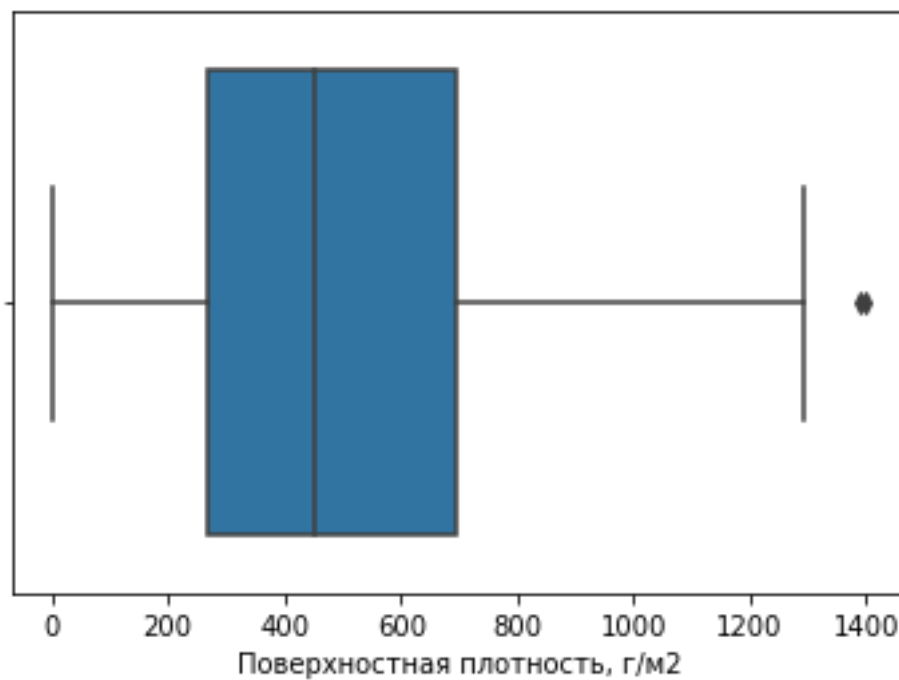
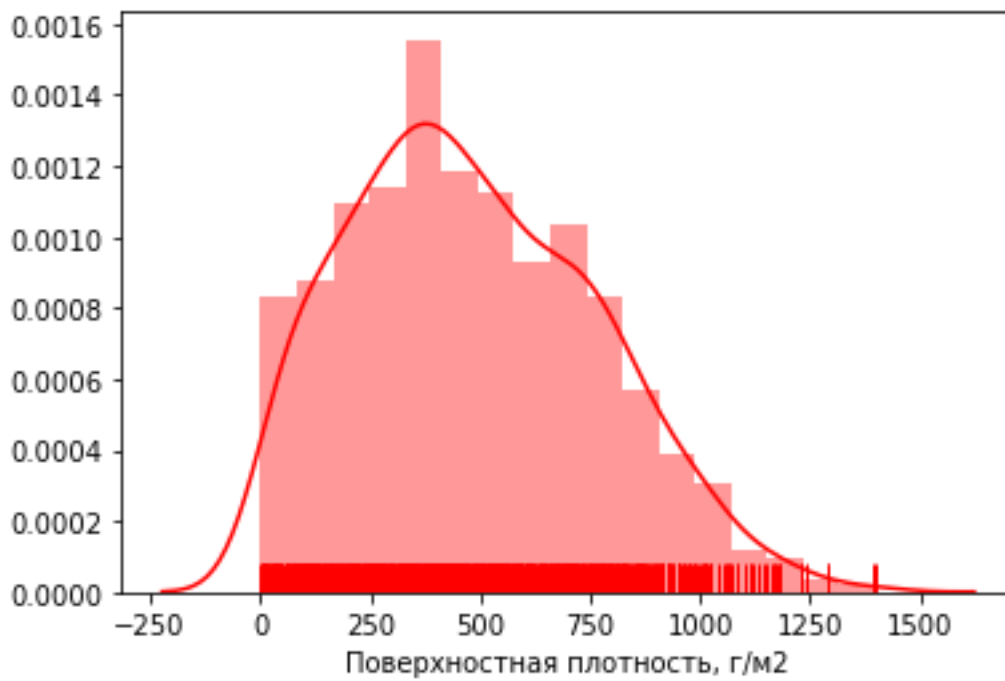
Минимальное значение: 100.0

Максимальное значение: 413.273418243566

Среднее значение: 285.88215135162227

Медианное значение: 285.896812331237

Поверхностная плотность, г/м²



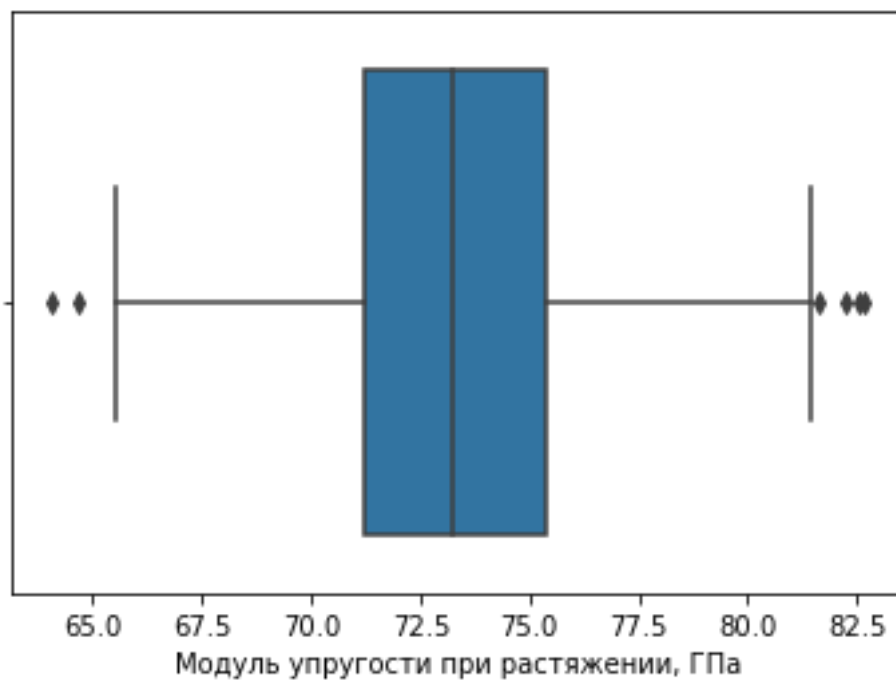
Минимальное значение: 0.603739925153945

Максимальное значение: 1399.54236233989

Среднее значение: 482.7318330384181

Медианное значение: 451.86436518306

Модуль упругости при растяжении, ГПа



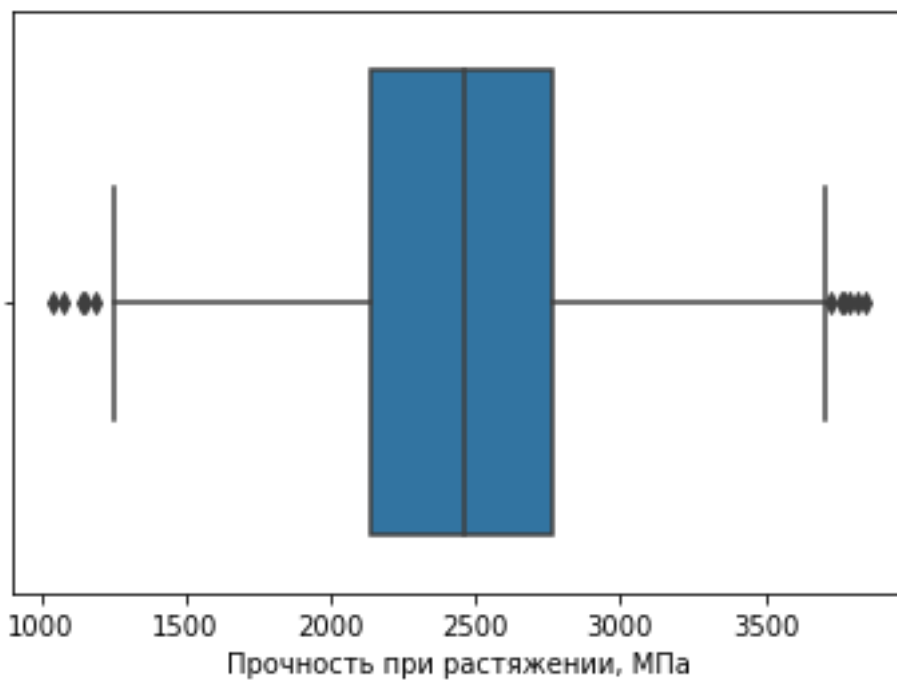
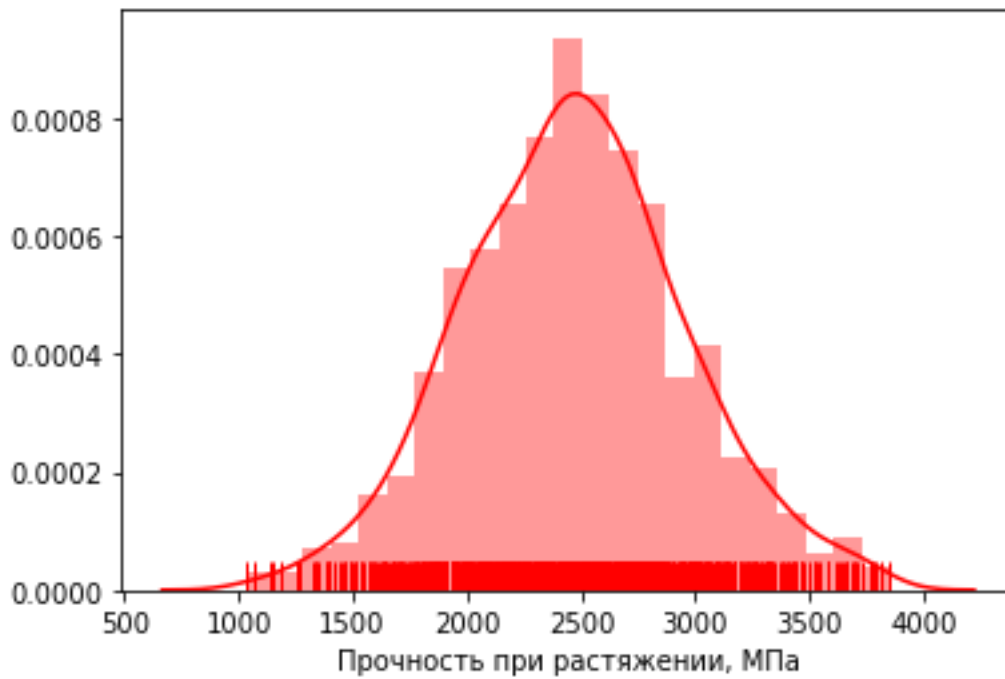
Минимальное значение: 64.0540605597917

Максимальное значение: 82.682051035271

Среднее значение: 73.32857125009073

Медианное значение: 73.2688045943481

Прочность при растяжении, МПа



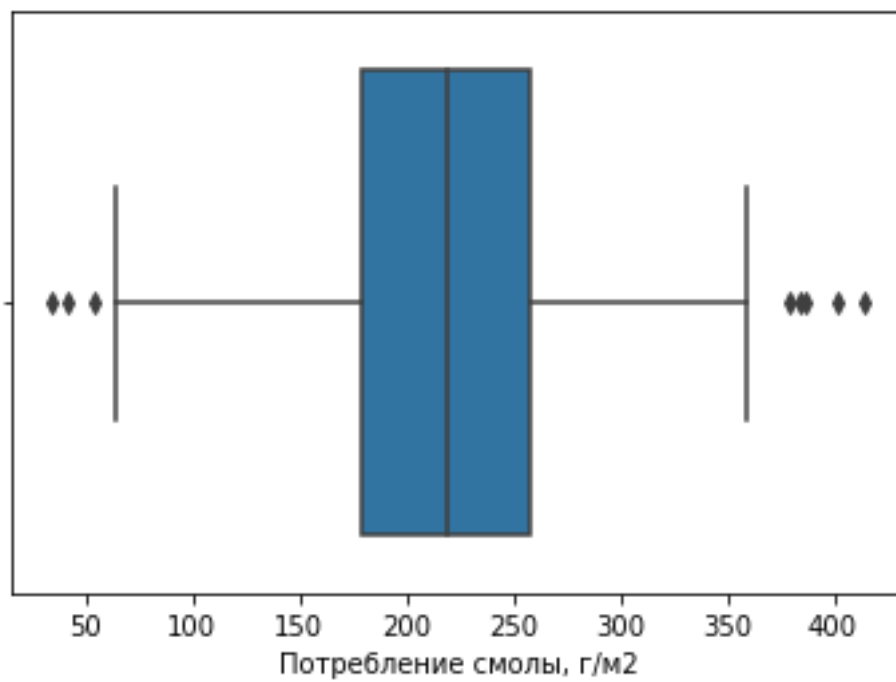
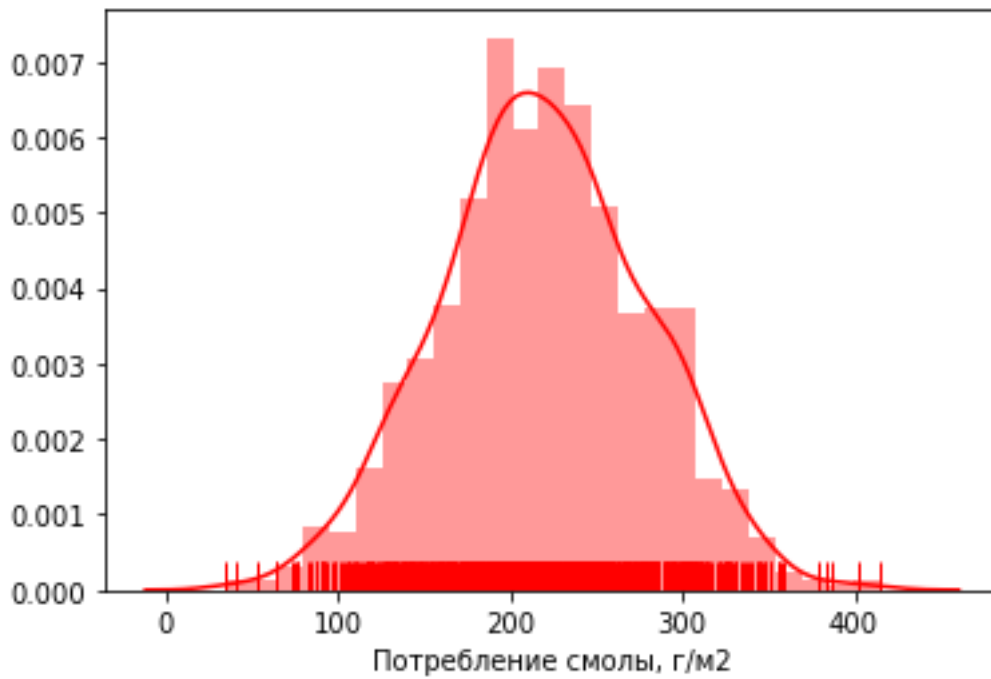
Минимальное значение: 1036.85660535

Максимальное значение: 3848.43673187618

Среднее значение: 2466.9228426979025

Медианное значение: 2459.52452600309

Потребление смолы, г/м²



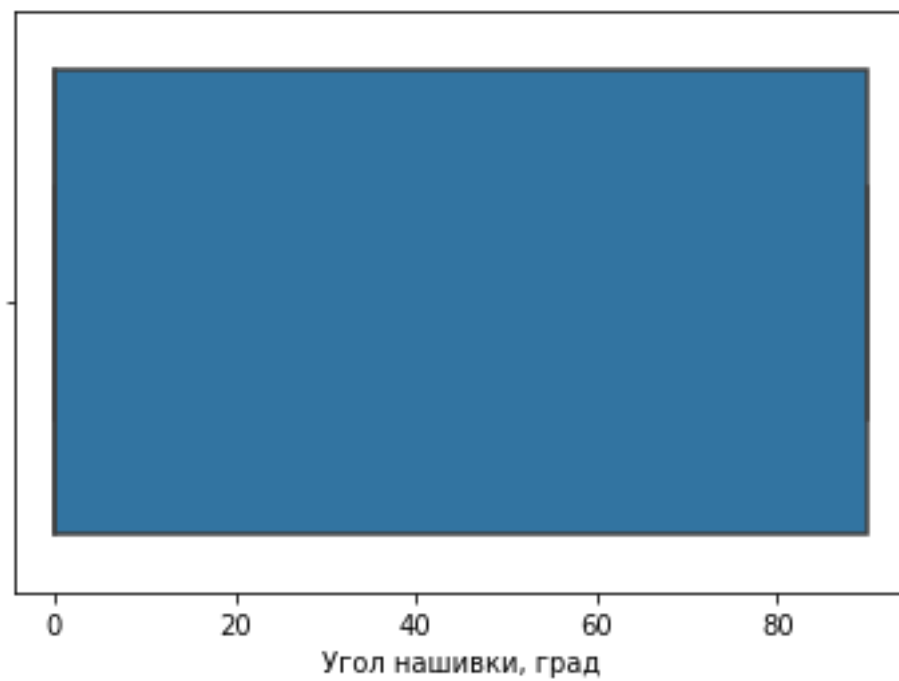
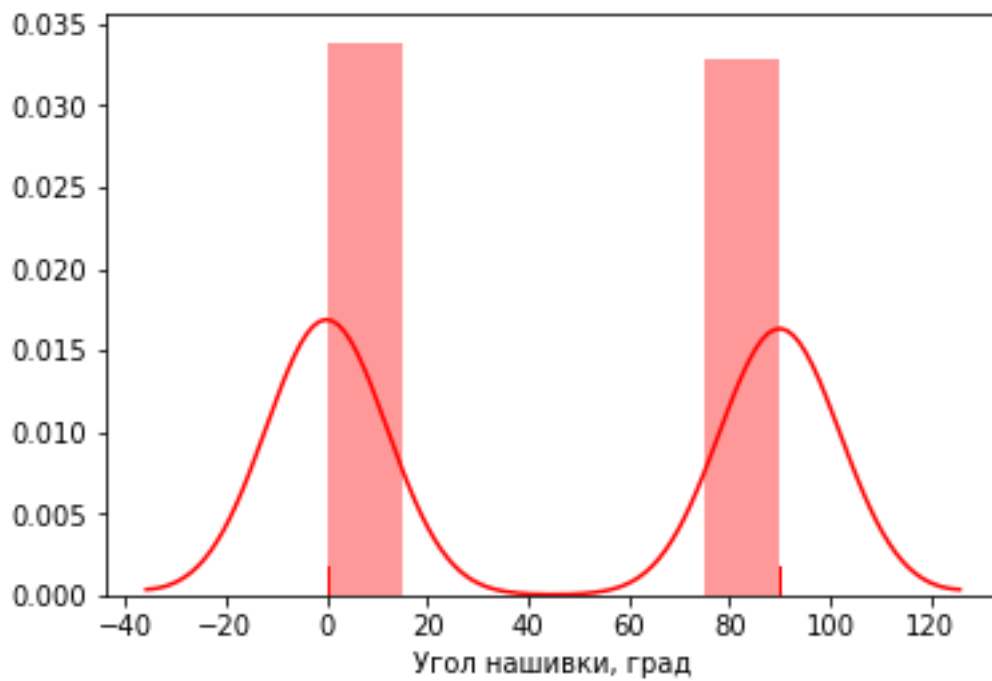
Минимальное значение: 33.8030255329625

Максимальное значение: 414.590628361534

Среднее значение: 218.42314367654265

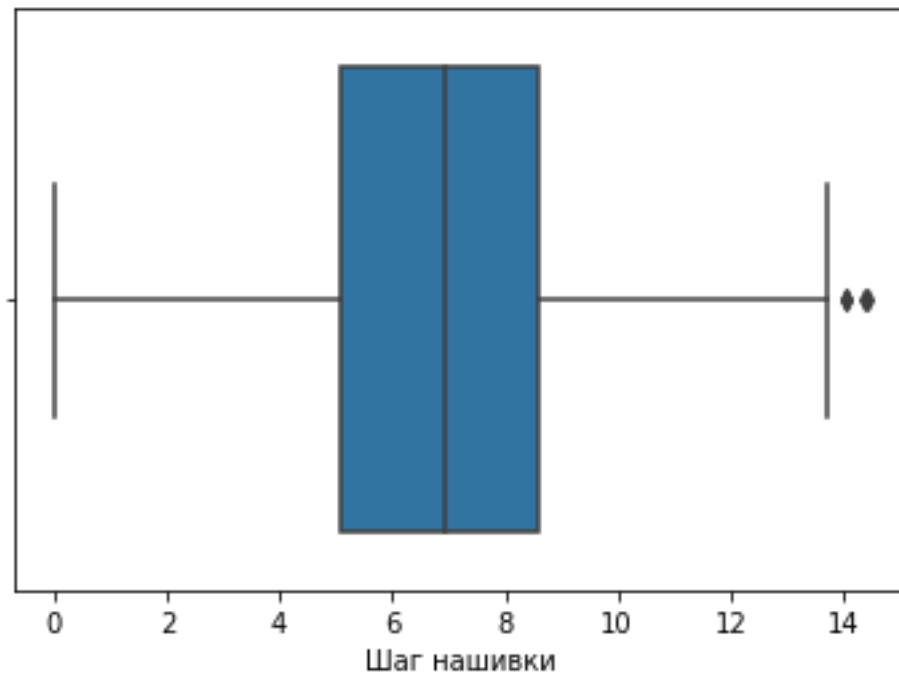
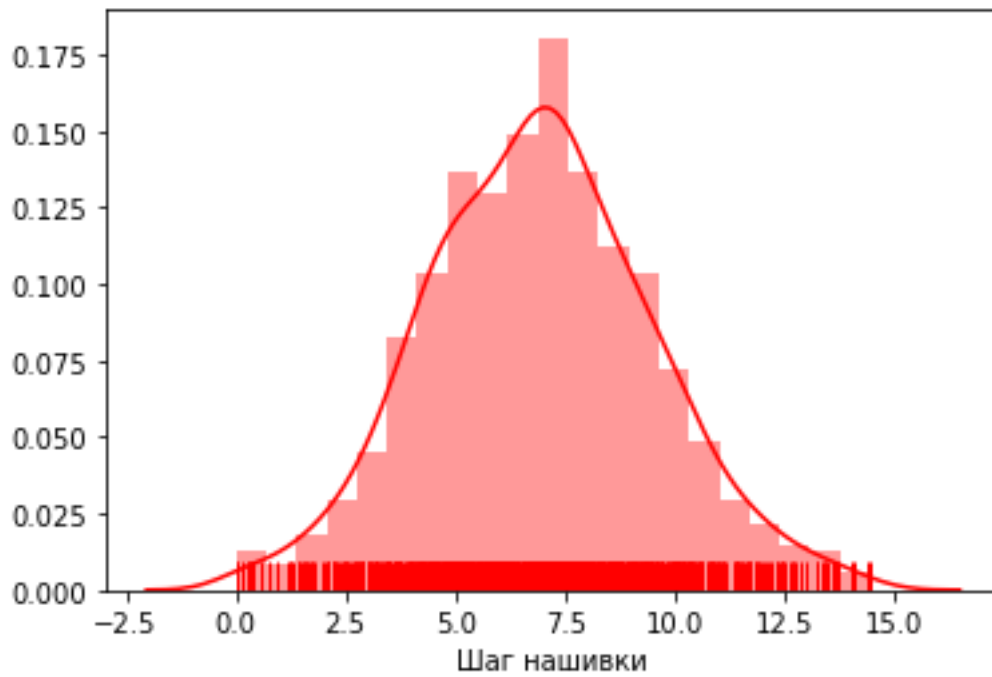
Медианное значение: 219.198882195134

Угол нашивки, град



Минимальное значение: 0
 Максимальное значение: 90
 Среднее значение: 44.252199413489734
 Медианное значение: 0.0

Шаг нашивки



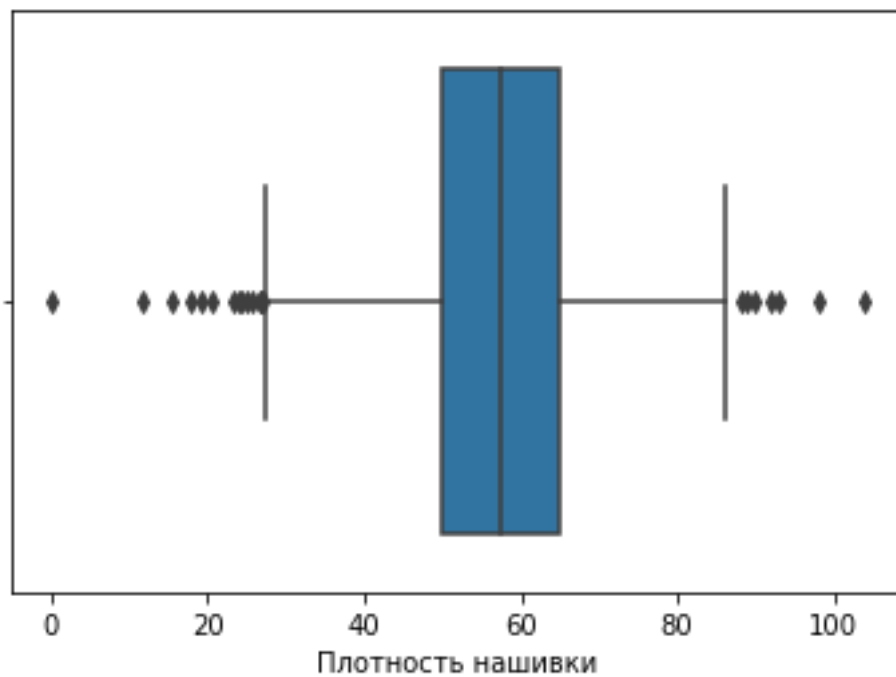
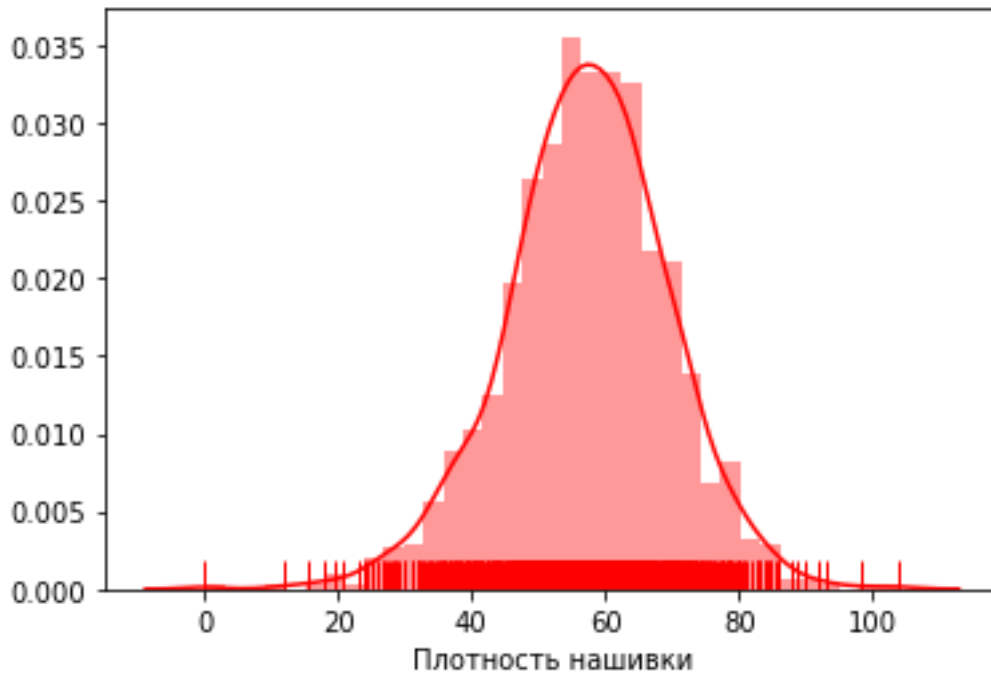
Минимальное значение: 0.0

Максимальное значение: 14.4405218753969

Среднее значение: 6.899222077675024

Медианное значение: 6.9161438559491

Плотность нашивки



Минимальное значение: 0.0
 Максимальное значение: 103.988901301494
 Среднее значение: 57.15392943285763
 Медианное значение: 57.3419198469929

Из всех параметров, помимо «Угол нашивки, град», представленного всего двумя возможными значениями, сильнее прочих также выделяется «Поверхностная плотность, г/м²», медиана которого разнится с выборочным средним сильнее, чем у других параметров, а форма распределения менее других походит на нормальное.

Также «ящики с усами» по каждому из параметров, кроме «Угол нашивки, град», показывают наличие некоторого количества значений, находящихся за пределами полутора межквартильных расстояний от первого и третьего квартилей.

Вместе с тем, трактовка сходу всех таких записей выбросами идет в разрез с принципом [7, 46], состоящим в том, что разведочный анализ данных позволяет поставить под сомнение те или иные значения, но при этом немаловажен учет специфики предметной области.

Принимая во внимание, во-первых, источник формирования данных – решение производственных задач (данные измерений), во-вторых то, что нетипичные значения параметров, хотя и находятся за пределами «усов», не демонстрируют экстремально больших отклонений, и в-третьих то, что такие значения присутствуют в том числе и у целевых параметров (а задача исследований в данной области – получение композитов с уникальными свойствами), такие значения вне дополнительных уточнений не следует трактовать как выбросы, по крайней мере, до тех пор, пока их наличие в обучающей и тестовых выборках не будет негативно сказываться на точности предсказаний модели.

Менее радикальным способом оценки качества исходных данных является применение «правила трех сигм».

Число значений по каждому из параметров, а также общее число записей, у которых хотя бы один из параметров находится на расстоянии более «трех сигм» от выборочного среднего, приведено в таблице ниже.

Примечательно, что данная группа записей, как и при отборе «ящиком с усами», также включает строку с нулевыми значениями параметров «Шаг нашивки» и

«Плотность нашивки», физический смысл которой малопонятен, но зато не содержит записей с предельными значениями целевых параметров.

Описательная статистика и диаграммы размаха датасета после удаления выбросов с применением «правила трех сигм» представлены, соответственно, в таблице и на рисунке ниже.

Таблица 5 – Описательная статистика датасета после очистки выбросов

	Соотношение матрица-наполнитель	Плотность, кг/м3	модуль упругости, ГПа	Количество отвердителя, м.%	Содержание эпоксидных групп, %_2	Температура вспышки, С_2	Поверхностная плотность, г/м2	Модуль упругости при растяжении, ГПа	Прочность при растяжении, МПа	Потребление смолы, г/м2	Угол нашивки, град	Шаг нашивки	Плотность нашивки
count	936	936	936	936	936	936	936	936	936	936	936	936	936
mean	0,50	0,50	0,45	0,50	0,49	0,52	0,37	0,49	0,50	0,52	0,51	0,50	0,51
std	0,19	0,19	0,20	0,19	0,18	0,19	0,22	0,19	0,19	0,20	0,50	0,18	0,19
min	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
25%	0,37	0,37	0,30	0,38	0,37	0,39	0,21	0,36	0,37	0,39	0,00	0,37	0,39
50%	0,49	0,51	0,45	0,51	0,49	0,52	0,35	0,49	0,49	0,52	1,00	0,50	0,52
75%	0,63	0,62	0,58	0,64	0,62	0,65	0,54	0,62	0,61	0,65	1,00	0,62	0,64
max	1	1	1	1	1	1	1	1	1	1	1	1	1

После очистки выбросов по «правилу трех сигм» медианы и выборочные средние параметров «подтянулись» ближе друг к другу, за исключением параметра «Поверхностная плотность, г/м2», чья форма распределения, отличная от нормального, также сохранилась.

Построение матрицы корреляции и/или визуализация матрицы рассеивания позволяют получить представление о том, как попарно связаны между собой те или иные параметры.

Визуализация матрицы корреляции и матрицы рассеяния исходных данных данной задачи, представленная на рисунках ниже, показывает около нулевую попарную корреляцию между параметрами и, соответственно, указывают на нелинейный характер связей между ними.

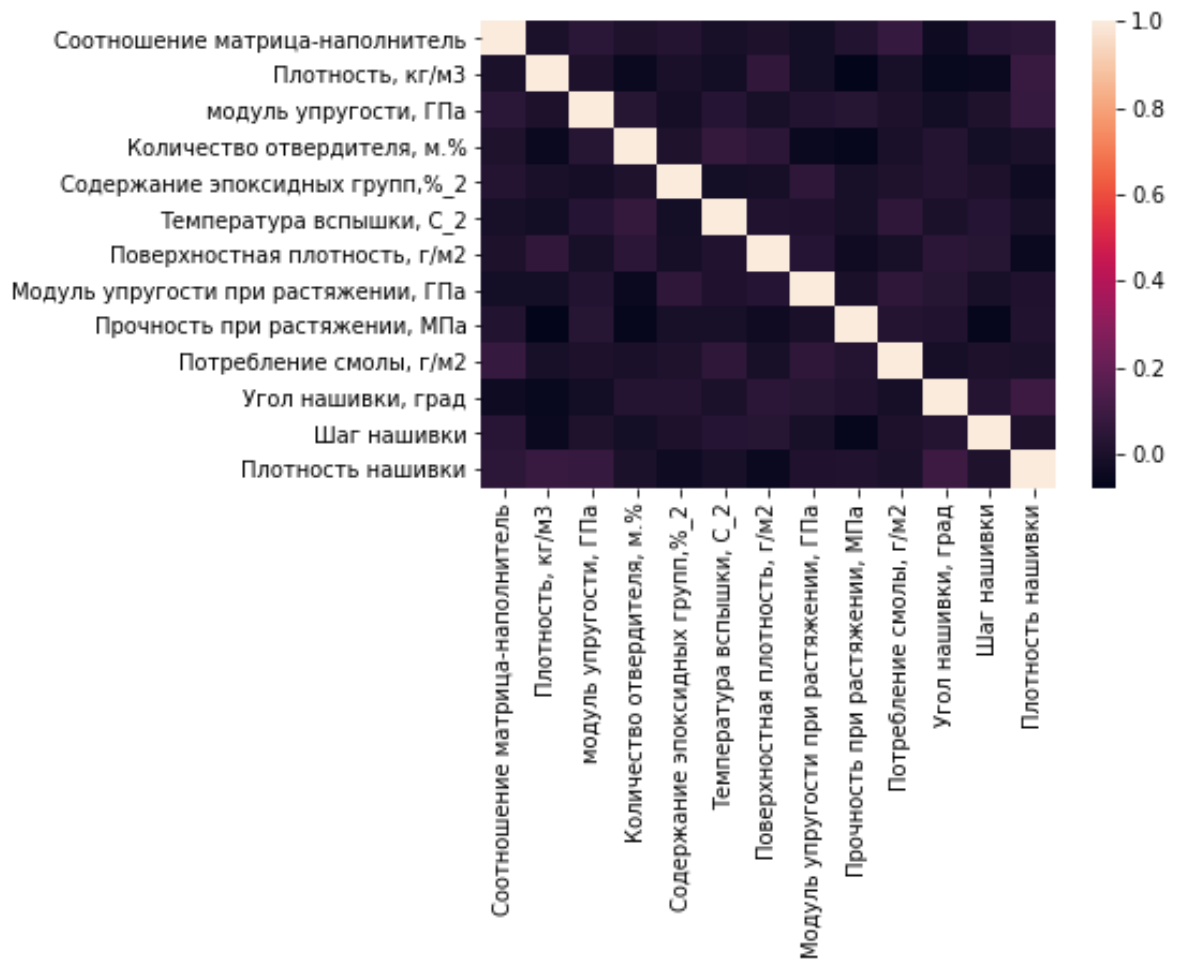


Рисунок 3 – Матрица корреляции датасета

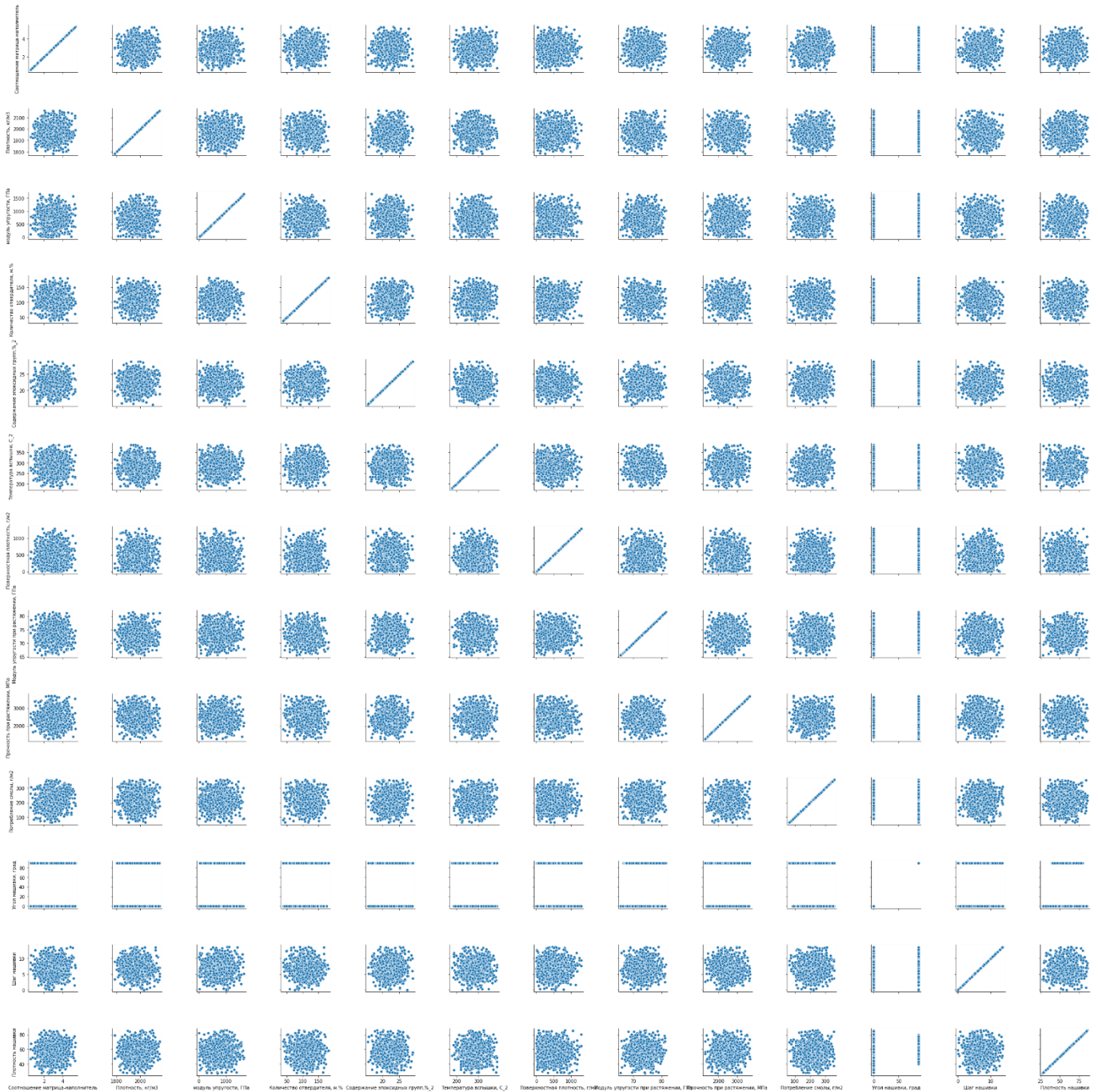


Рисунок 4 – Матрица попарной зависимости датасета

В условиях нелинейных зависимостей, для возможности «взглянуть» на данные в совокупности признаков и попытаться проследить какие-то взаимосвязи, следует обратиться к методам обучения на базе многообразий – класс оценщиков без учителя, нацеленных на описание наборов данных как низкоразмерных многообразий, вложенных в пространство большей размерности [2, 501].

2 Разработка моделей машинного обучения

2.1 Предобработка данных

Для предобработки данных использовались следующие процедуры:

1. анализ на пропуски и удаление пропусков

информационно, метод .info() датасета

2. удаление выбросов

замена данных, за пределами второго и третьего квантиля на пустые, затем удаление строк, содержащие пустые значения

сумма выбросов по каждому столбцу:

Соотношение матрица-наполнитель	6
Плотность, кг/м3	9
модуль упругости, ГПа	2
Количество отвердителя, м.%	14
Содержание эпоксидных групп,%_2	2
Температура вспышки, С_2	8
Поверхностная плотность, г/м2	2
Модуль упругости при растяжении, ГПа	6
Прочность при растяжении, МПа	11
Потребление смолы, г/м2	8
Угол нашивки, град	0
Шаг нашивки	4
Плотность нашивки	21

3. нормализация данных

процедура MinMaxScaler библиотеки sklearn

Итоговая выборка представляет собой датасет с 936 уникальными строками статистика , представлена в таблице 5 (см выше)

2.2 Разработка и обучение моделей

В настоящей работе разработка и обучение моделей машинного обучения осуществляется для двух выходных параметров: «Прочность при растяжении» и «Модуль упругости при растяжении». При этом для каждого признака построение моделей осуществляется отдельно.

Для признака «Прочность при растяжении» были разработаны и обучены следующие модели:

- модель k ближайших соседей (KNeighborsRegressor);
- модель на основе градиентного бустинга (GradientBoostingRegressor()).

Для признака «Модуль упругости при растяжении» были разработаны и обучены следующие модели:

- модель на основе линейной регрессии (LinearRegression);
- модель на основе опорных векторов (SVR).

Порядок разработки модели для каждого параметра и для каждого выбранного метода можно разделить на следующие этапы:

Этап 1. Разделение нормализованных данных на обучающую и тестовую выборки (в соотношении 70 на 30%)

Этап 2. Задание сетки гиперпараметров, по которым будет происходить оптимизация модели. В качестве параметра оценки выбран коэффициент детерминации (R^2).

```
#Зададим сетку параметров, по которым будем оптимизировать модель
t_search11 = {'weights': ['uniform', 'distance'],
              'n_neighbors': list(np.linspace(5, 100, 10, dtype = int)),
              'algorithm': ['auto', 'ball_tree', 'kd_tree', 'brute'],
              'leaf_size': list(np.linspace(10, 100, 10, dtype = int))}
#В качестве первой модели будем использовать метод ближайших соседей
clf11 = KNeighborsRegressor()
```

Рисунок 1. Пример определения сетки параметров для модели k ближайших соседей.

Этап 3. Оптимизация подбора гиперпараметров модели с помощью выбора по сетке и перекрестной проверки.

Этап 4. Подстановка оптимальных гиперпараметров в модель и обучение модели на тренировочных данных.

2.3 Тестирование моделей

После обучения моделей была проведена оценка точности этих моделей на обучающей и тестовых выборках. В качестве параметра оценки модели использовалась средняя абсолютная ошибка (MAE). Помимо этого, для большей наглядности результатов работы модели на тестовых данных, были построены диаграммы рассеяния тестовых данных (реальные данные) и значений, полученных в качестве прогноза,

Результаты модели k ближайших соседей для параметра «Прочность при растяжении» (KNeighborsRegressor).

Средняя абсолютная ошибка (тест): 0.0000

Средняя абсолютная ошибка (тренир данные): 0.1483



Рисунок 2. Результаты работы модели по оценке значений параметра на основе тестовых данных (Прочность при растяжении)

Результаты модели повышения градиента «Модуль упругости при растяжении» (GradientBoosting Regressor).

Средняя абсолютная ошибка (тест) : 0.1352

Средняя абсолютная ошибка (трен): 0.1513



Рисунок 3. Результаты работы модели GradientBoostingRegressor (Модуль упругости при растяжении)

Результаты модели повышения градиента (LinearRegression).

Средняя абсолютная ошибка (тест): 0.1547

Средняя абсолютная ошибка (тренир): 0.1510



Рисунок 4. Результаты работы модели LinearRegression

Результаты модели эпсилон-регрессия опорных векторов (SVR).

Средняя абсолютная ошибка(тест): 0.1549

Средняя абсолютная ошибка (тренир): 0.1524



Рисунок 5. Результаты работы модели SVR

2.5. Разработка нейросети

Для рекомендации соотношения «матрица-наполнитель» была разработана простая модель глубокого обучения с помощью библиотеки Keras.

Архитектура нейронной сети может быть описана следующим образом.

Модель состоит из трех скрытых уровней. Первый содержит 64 нейрона, что чуть более чем в три раза превышает объем входных данных (10 входных переменных). Последующие скрытые уровни – они содержат 64, и 1 нейрона. Снижение числа нейронов на каждом уровне сжимает информацию, которую сеть обработала на предыдущих уровнях.

Скрытые уровни сети трансформируются функциями активации. Эти функции – важные элементы сетевой инфраструктуры, поскольку они привносят в систему нелинейность.

Для эксперимента были выбраны 3 функции активации: tanh (арктангенс), relu (выпрямленная линейная единица) и sigmoid (сигмоида $1/(1+\exp(-x))$)

```
def build_model():  
    model=models.Sequential()  
    model.add(layers.Dense(64, activation='tanh', input_shape=(Xtrn1.shape[1],)))  
    model.add(layers.Dense(64, activation='tanh'))  
    model.add(layers.Dense(1))  
    model.compile(optimizer='rmsprop', loss='mse', metrics=['mae'])  
    return model
```

Рисунок 6. Архитектура нейронной сети

Далее была определена функция стоимости сети, которая используется для генерации оценки отклонения между прогнозами сети и реальными результатами наблюдений в ходе обучения. Для решения проблем с регрессией используют функцию средней квадратичной ошибки (mean squared error, MSE). Данная функция вычисляет среднее квадратичное отклонение между предсказаниями и целями.

В качестве оптимизатора использовался RMSprop похож на алгоритм градиентного спуска с импульсом. Оптимизатор RMSprop ограничивает колебания в вертикальном направлении

Обучение сети останавливается по достижению максимального числа эпох или при срабатывании другого определенного заранее критерия остановки.

После обучения модели для была определена средняя абсолютная ошибка на тестовом наборе данных.

Средняя абсолютная ошибка: 0.174.

Также результаты прогноза модели на тестовых данных были представлены по аналогии с результатами для моделей машинного обучения, описанных в предыдущем разделе.



Рисунок 7. Прогнозные данные для модели с функцией \tanh



Рисунок 8. Прогнозные данные для модели с функцией relu



Рисунок 9. Прогнозные данные для модели с функцией sigmoid

2.6 Разработка приложения

Приложение разработано с интерфейсом командной строки и позволяет решать задачи прогнозирования целевой переменной на основе входных данных.

Приложение может быть запущено из командной строки следующим образом:

1. Программа и датасет находятся в единой папке /ВКР. В данном случае (на конкретном ПК) папка /## ВКР расположена по полному адресу: C:\Users\user\Desktop\Data Science\## ВКР\.

2. В нашем случае используется операционная система Windows 7¹ и PowerShell² 7.2.2. Используя командную строку, запускаем команду `cd`³ и прописываем путь до папки: `cd C:\Users\user\Desktop\Data Science\## ВКР\`. Затем даём команду на запуск консольного приложения: `command_line.py`.

3. Появляются запросы на ввод значений переменных:

Плотность, кг/м³ модуль упругости, ГПа

Количество отвердителя, м.%

Содержание эпоксидных групп,%₂

Температура вспышки, С₂

Поверхностная плотность, г/м²

Модуль упругости при растяжении, ГПа

Прочность при растяжении, МПа

Потребление смолы, г/м²

Угол нашивки, град

Шаг нашивки

Плотность нашивки

¹ Проприетарная операционная система для персональных компьютеров, разработанная компанией Microsoft в рамках семейства Windows NT. При разработке ВКР использовалась версия Windows 11 Pro, версии 22H2, сборка 22581.200.

² Расширяемое средство автоматизации от Microsoft с открытым исходным кодом, состоящее из оболочки с интерфейсом командной строки и сопутствующего языка сценариев. Новая версия PowerShell 7.2.2 устанавливается в новый каталог и работает параллельно с Windows PowerShell 5.1.

³ Отображает имя текущего каталога или изменяет текущий. Если используется только с буквой диска (например, `cd C:`), на `cd C:` диске отображаются имена текущего каталога на указанном диске.

4. После ввода всех переменных выдается прогнозное значение параметра «Соотношение матрица-наполнитель», сформированного моделью нейросети model (п 2.5, модель с функцией \tanh)

3 Заключение

Теоретически разработанный метод определения надёжности изделий из композиционных материалов, основанный на использовании статистически достоверных характеристик материалов, полученных физическим и вычислительным экспериментом, позволяет оценивать уровень надёжности изделий как в отдельных точках, так и по всему объёму в целом.

4 Список используемой литературы и ссылки на веб-ресурсы

- [1] В.В. Васильев, В.Д. Протасов, В.В. Болотин и др.: Композитные материалы: справочник. Москва: Машиностроение, 1990, 510 с.
- [2] Плас Дж. Вандер, Python для сложных задач: наука о данных и машинное обучение. Санкт-Петербург: Питер, 2018, 576 с.
- [3] Язык программирования Python: - Режим доступа: <https://www.python.org/> (дата обращения 01.04.2022)
- [4] Библиотека Pandas – Режим доступа: <https://pandas.pydata.org/> (дата обращения 01.04.2022)
- [5] Библиотека Matplotlib – Режим доступа: <https://matplotlib.org/> (дата обращения 01.04.2022)
- [6] Библиотека Sklearn – Режим доступа: <https://scikit-learn.org/stable/> (дата обращения 01.04.2022)
- [7] К. Андерсон, Аналитическая культура. От сбора данных до бизнес-результатов: монография. Москва: O'Reilly, 2017, 392 с.
- [8] How to choose a machine learning model in Python? – Режим доступа: <https://www.codestar.com/choose-machine-learning-models-python/> (дата обращения 03.04.2022)