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 (композитных материалов)

Слушатель Ветчинов Николай Викторович

Москва 2022

Содержание

[**Введение** 3](#_Toc105951911)

[**Глава 1. Аналитическая часть** 9](#_Toc105951912)

[**1.1 Постановка задачи** 9](#_Toc105951913)

[**1.2 Описание используемых методов** 10](#_Toc105951914)

[**Глава 2. Практическая часть** 13](#_Toc105951915)

[**2.1. Предобработка данных** 13](#_Toc105951916)

[**2.1 Обучение моделей** 19](#_Toc105951917)

[2.3. Нейронная сеть 24](#_Toc105951918)

[Разработка приложения 31](#_Toc105951919)

[Репозиторий и результаты 38](#_Toc105951920)

[Заключение 39](#_Toc105951921)

[Библиографический список 40](#_Toc105951922)

# **Введение**

Первым создателем КМ была природа. Стволы деревьев, кости жи­вотных, зубы имеют характерную волокнистую структуру. Она состоит из сравнительно пластичного матричного вещества и более твердых и прочных веществ, имеющих форму волокон. Например, древесина - это композиция, состоящая из пучков высокопрочных целлюлозных волокон трубчатого сечения, связанных между собой матрицей из органического вещества - лигнина, придающего древесине поперечную прочность. Древесина - природный анизотропный материал с различными свойст­вами. Зубы людей и животных состоят из твердого вязкого поверхност­ного слоя - эмали и более мягкой сердцевины - дентина. И эмаль, и дентин содержат неорганические кристаллы игольчатой формы, распо­ложенные в мягкой органической матрице. Примерную структуру имеет и слоновая кость.

Такие природные образования, как минералы, также можно рас­сматривать как КМ. Например, нефрит состоит из плотно упакованных игольчатых кристаллов, связанных друг с другом на поверхностях раз­дела, что обеспечивает высокую вязкость разрушения.

История использования и развития КМ связана с производственной деятельностью человека. Уже более миллиона лет назад, с момента выделения человека из животного мира, люди сознательно использова­ли природные КМ - кости животных, кожу, древесину и др. За 5 тыс. лет до н.э. при изготовлении глиняных изделий и кирпичей в них добавляли песок, измельченные камни, солому, камыш и т.п. для снижения усадки, уменьшения растрескивания, повышения прочности. В Библии упоми­нается, что за 4000-2000 лет до н.э. вавилоняне использовали в строи­тельстве материалы на основе тростника, пропитанного битумом. Поз­же из них египтяне строили свои суда. В Азии более чем за 1000 лет до н.э. изготавливали луки, стрелы из композиционных материалов, в которых использовали древесину, слои рога, сухожилия животных, соеди­ненных с помощью клея. Такие луки характеризовались большой убой­ной силой и радиусом действия. В Китае, Индии в качестве связующего широко использовался лак на основе природной смолы - шеллака (сложная смесь полиэфирных смол). Лак смешивали с песком, спекали и получали точильные камни. В 1555-1560 гг. при постройке храма Ва­силия Блаженного в Москве зодчие Барма и Постник использовали ар­мированные железными полосами каменные плиты. Таких примеров можно привести много, однако это все были хоть и удачные, но случай­ные находки.

В некоторых источниках первым осознанно созданным КМ назы­вают железобетон, запатентованный в 1867 году французским ученым Ж. Монье и получивший широкое применение с конца XIX века. Железо­бетон можно отнести к числу первых образцов армированной керамики.

Первый патент на полимерный композиционный материал (ПКМ) был выдан в 1909 г. Он предусматривал упрочнение синтетических смол природными волокнами. Армировали первые ПКМ рубленными природными волокнами, целлюлозной бумагой, хлопчатобумажными и льняными тканями. Хотя о приоритете патента на ПКМ вопрос спорный. В описании технологии изготовления первого фанерного самолета братьями Уилбером и Орвиллом Райт, на котором они 17 декабря 1903 года совершили первый полет продолжительностью 59 сек указа­но, что для защиты фанерных крыльев от дождя и других атмосферных воздействий их оклеивали пропитанной смолой тканью - а это и есть ПКМ.

Стеклопластики запатентованы в 1935 году - это первые полимер­ные КМ, в которых в качестве упрочняющего элемента использовались неорганические волокна. Промышленный выпуск стеклопластиков на­лажен после Второй мировой войны, и с тех пор их интенсивно исполь­зуют в технике.

В 1941 году в США был подписан первый правительственный кон­тракт на создание материала из хлопкового волокна, пропитанного фенольной смолой. Целенаправленно стеклопластики в авиастроении на­чали использовать с 1943 года, когда из них стали изготавливать кресла летчиков для учебных самолетов и облицовывать кабины.

В 1940-1950 гг. появляются современные, созданные сознательно, композиционные материалы - полимерные, керамические, металличе­ские и другие. Они были вызваны к жизни потребностями передовых областей науки и техники: судо- и авиастроения, военной техники, за­рождающейся космонавтики и др. (в США, Германии, Франции, СССР и др.). Их создание и совершенствование продолжается и в настоящее время.

В 50-х годах XX века было обнаружено, что многие материалы в виде тонких монокристаллов игольчатой формы обладают фантасти­чески высокой прочностью 10000 МПа и выше. Были получены новые виды неорганических поликристаллических волокон - углеродные, бор­ные с прочностью 3000-3500 МПа и модулем упругости (3-5)∙105 МПа. Эти сверхпрочные волокна стали использовать для армирования раз­личных матриц.

История развития современных КМ насчитывает чуть больше по­ловины столетия, но успехи в этом направлении достигнуты значитель­ные. Они нашли применение в самолетостроении и аэрокосмической промышленности, в автомобилестроении, в медицине и строительстве, в судостроении и производстве спортинвентаря, в быту и во многих дру­гих направлениях деятельности человека.

Сейчас самолеты-невидимки на 75-85% по массе состоят из КМ, престижные иномарки автомобилей содержат от 120 до 150 кг деталей из КМ, не только облицовочные материалы в строительстве, но и целые временные постройки выполнены из КМ. По объему производства эти материалы в развитых странах находятся на третьем, а в некоторых -на втором месте.

**Внедрение КМ в технике**

Основой научно-технического прогресса, его фундаментом всегда было и остается машиностроение. От уровня развития машинострое­ния, от степени совершенства машин в значительной мере зависят про­изводительность общественного труда и благосостояние народа. В на­стоящее время возможности повышения качества изделий лежат не только в совершенствовании их конструкций, но и в дальнейшем улуч­шении свойств конструкционных материалов, в использовании принци­пиально новых технологических процессов их обработки.

Перспективной тенденцией современного машиностроения являет­ся замена традиционных конструкционных материалов (поликристалли­ческих) на КМ в деталях различного назначения, как несущих, так и не­несущих. Это вызвано целым рядом причин научного, технического, экономического и социального характера.

Совершенство конструкций в значительной степени определяется параметрами прочности и жесткости конструкционных материалов, из которых они изготавливаются. В связи с этим рассмотрим, каковы ре­зервы повышения прочностных характеристик традиционных конструкционных материалов, например, стали и алюминиевых сплавов, для чего сравним их характеристики в начале 40-х годов и по прошествии 50-ти лет.

Предел прочности алюминиевых сплавов в 40-х годах был порядка 400 МПа, а стали - 1200 МПа. К началу 90-х годов они изменились в сторону увеличения и составили соответственно 650 МПа и 2200 МПа, причем дальнейшее увеличение этих характеристик дается с большим трудом. Специалисты приходят к выводу, что в связи с исчерпанием потенциальных возможностей механическая технология вступает в по­лосу своего заката. Совершенствование механических методов воздей­ствия на предмет труда не дает теперь большого экономического эф­фекта, не приносит больших качественных изменений в продуктах тру­да. Следовательно, резервы прочностных характеристик стали и алю­миниевых сплавов исчерпываются. Анализируя другую, не менее важ­ную характеристику материалов - модуль упругости первого рода, предварительно отметим, что в настоящее время особое значение при­дается материалоемкости и весовым параметрам конструкций. В связи с этим при оценке механических свойств материалов необходимо об­ращать внимание и на его удельный вес - γ. Оказалось, что обобщаю­щими и показательными характеристиками являются удельные модуль упругости и предел прочности, т.е. отношение модуля или предела прочности к удельному весу. Для алюминиевых сплавов γ = 2,5 г/см3, для стали - 7,8 г/см3. Оказывается, что значение удельного модуля уп­ругости для всех металлов изменяется в довольно узком диапазоне -2300-2600 км, причем изменить эту характеристику обычными техноло­гическими приемами невозможно.

Таким образом, можно сделать вывод: механические характеристи­ки традиционных конструкционных материалов, широко используемых в промышленности, находятся на предельном уровне.

Второй ограничительный фактор, которому до настоящего времени не уделялось должное внимание, - это исчерпание энергетических ре­сурсов, постепенное истощение наиболее богатых и наиболее доступ­ных месторождений полезных ископаемых, большинство из которых принадлежит к невозобновляемым, их добыча становится все сложнее и дороже.

Этот фактор характерен для большинства развитых стран. Одна из причин этого - постоянно увеличивающийся объем потребления метал­лов, что неизбежно приведет к тому, что месторождения руд, богатых по содержанию металлов, наиболее близко расположенных к сущест­вующим коммуникациям, удобных для добычи, будут в ближайшем бу­дущем выработаны. В период 1970-1990 гг. мировое производство же леза увеличилось в 2,7 раза, меди - в 2,3, алюминия - в 4,7, никеля - в 4, цинка - в 2 и титана - в 17 раз. По мнению специалистов, объемы производства и потребления металлов будут расти и дальше. Если предположить, что среднегодовое потребление, например, стали во всем мире приблизится к сегодняшнему уровню развитых стран (500 кг на душу населения), и условно допустить, что дальнейшего роста объ­емов потребления в этих странах не произойдет, то тогда на земном шаре должно производиться 2,5...3 млрд. тонн стали ежегодно.

Из анализа тенденций роста производства металлов естественно вытекает вопрос: является ли увеличение их производства единствен­ной формой удовлетворения растущих с каждым годом потребностей человечества в металлах или есть рациональная альтернатива - более эффективное их использование, появление и стремительное увеличе­ние использования наполненных КМ. В развитых странах (США, госу­дарствах Западной Европы, Японии) за последние 25 лет объем произ­водства КМ возрос более чем в 100 раз, что позволило наращивать производство товаров, снизив потребление стали. Производство стали в США снизилось со 134 млн. тонн в 1977 г. до 99 млн. тонн в 1990 г., в странах Евросоюза - с 117 млн. тонн до 103 млн. тонн за тот же период. В этой связи интересно сопоставить потребление стали в бывшем Со­ветском Союзе: потребление возросло за сопоставляемый период с 145 млн. тонн до 161 млн. тонн.

В мировом потреблении стали происходит перераспределение до­ли разных стран по этому показателю: доля США и Канады снизилась с 20,9% до 15,5%, а доля развивающихся стран возросла с 12,9% до 15,5%.

Композиционные материалы можно конструировать и получать с заранее заданными физико-механическими характеристиками, их плот­ность в 3...6 раз ниже стали, они безотходны при переработке в изде­лия, инертны к окружающей и агрессивным средам, т.е. не подвержены коррозии, обладают направленной тепло- и электропроводностью, зву­конепроницаемостью и т.д. Обычно эти характеристики превосходят характеристики поликристаллических материалов в несколько раз и да­же на несколько порядков. Считается, что 1 тонна изделий из КМ заме­няет, в среднем, 10 тонн стали.

Создание и внедрение КМ сопровождается рядом сопутствующих положительных эффектов. Так, во многих случаях существенно упро­щается технология изготовления деталей машин и конструкций из этих материалов, что позволяет экономить энергетические, человеческие и материальные ресурсы. Важны также эффекты снижения эксплуатационных расходов, выражающиеся в экономии топлива, увеличении дол­говечности, удлинении сроков межремонтной эксплуатации и т.д.

Композиционные материалы, за редким исключением, еще не­сколько десятков лет назад относились к разряду экзотических. Они создавались с целью обеспечения все возрастающих требований к ле­тательным аппаратам в области весового совершенства, обеспечения надежности, которые, в свою очередь, определялись экстремальными условиями эксплуатации элементов конструкции.

Следует отметить различия в подходах при освоении КМ, сущест­вовавшие в бывшем Советском Союзе и в зарубежных странах. Отече­ственные разработки по КМ по многим направлениям не уступали, а в некоторых случаях превосходили зарубежные, но в нашей стране они засекречивались, и только после снижения их актуальности разреша­лось их использование в гражданской авиации, автомобилестроении и т.д. Все это ограничивало объемы применения и приводило к высокой стоимости КМ в сравнении с металлами за единицу веса. За рубежом они, помимо использования в военно-промышленном комплексе в уп­рощенном варианте применялись при производстве большого количе­ства изделий народного потребления. Массовый и крупносерийный ха­рактер производства обеспечивал совершенствование технологий их переработки, высокое качество и степень автоматизации используемого оборудования.

В настоящее время композиты все шире используются в различных отраслях, где преобладают обычные, традиционные условия эксплуа­тации, далекие от экстремальных. Это приводит к существенному уде­шевлению их стоимости и созданию большой группы КМ из разнообраз­ных составляющих элементов, различающихся как по геометрическим, так и по физическим параметрам. Однако внедрение КМ связано с ря­дом специфических факторов, которые надо учитывать при разработке элементов конструкций на их основе.

# **Глава 1. Аналитическая часть**

## **1.1 Постановка задачи**

Необходимо спрогнозировать ряд конечных свойств получаемых композиционных материалов.

В их перечень входит:

* Модуль упругости при растяжении, Гпа
* Прочность при растяжении, Мпа

Анализ необходимо проводить на основе следующих свойств:

* 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

Задача основан на реальных производственных задачах Центра НТИ «Цифровое материаловедение: новые материалы и вещества» (структурное подразделение МГТУ им. Н.Э. Баумана).

Заинтересованными лицами могут выступать Российские металлурги:

* ГМК НорНикель
* Северсталь
* Русал
* En+
* НЛМК
* Росатом
* И т.д.

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

## **1.2 Описание используемых методов**

Для прогнозирования модуля упругости при растяжении и прочности при растяжении применим несколько методов машинного обучения.

* Случайный лес решений
* Логистическая регрессия

**Метод «Случайный лес решений»**

Алгоритм машинного обучения, заключающийся в использовании комитета (ансамбля) решающих деревьев. Алгоритм сочетает в себе две основные идеи: метод бэггинга Бреймана, и метод случайных подпространств, предложенный Тин Кам Хо. Алгоритм применяется для задач классификации, регрессии и кластеризации. Основная идея заключается в использовании большого ансамбля решающих деревьев, каждое из которых само по себе даёт очень невысокое качество классификации, но за счёт их большого количества результат получается хорошим.

Сама структура дерева представляет собой «листья» и «ветки». На рёбрах («ветках») дерева решения записаны признаки, от которых зависит целевая функция, в «листьях» записаны значения целевой функции, а в остальных узлах — признаки, по которым различаются случаи. Чтобы классифицировать новый случай, надо спуститься по дереву до листа и выдать соответствующее значение.

Подобные деревья решений широко используются в интеллектуальном анализе данных. Цель состоит в том, чтобы создать модель, которая предсказывает значение целевой переменной на основе нескольких переменных на входе.

Каждый лист представляет собой значение целевой переменной, изменённой в ходе движения от корня по рёбрам дерева до листа. Каждый внутренний узел сопоставляется с одной из входных переменных. Дерево может быть также «изучено» разделением исходных наборов переменных на подмножества, основанные на проверке значений признаков. Это действие повторяется на каждом из полученных подмножеств. Рекурсия завершается тогда, когда подмножество в узле имеет те же значения целевой переменной, таким образом, оно не добавляет ценности для предсказаний.

**Линейная регрессия**

Используемая в статистике регрессионная модель зависимости одной (объясняемой, зависимой) переменной от другой или нескольких других переменных (факторов, регрессоров, независимых переменных) с линейной функцией зависимости. Модель линейной регрессии является часто используемой и наиболее изученной в эконометрике. А именно изучены свойства оценок параметров, получаемых различными методами при предположениях о вероятностных характеристиках факторов, и случайных ошибок модели. Предельные (асимптотические) свойства оценок нелинейных моделей также выводятся исходя из аппроксимации последних линейными моделями. С эконометрической точки зрения более важное значение имеет линейность по параметрам, чем линейность по факторам модели.

Регрессионная модель



где b — параметры модели,  — случайная ошибка модели; называется линейной регрессией, если функция регрессии f(x,b) имеет вид



где  — параметры (коэффициенты) регрессии,  — регрессоры (факторы модели), k — количество факторов модели. Коэффициенты линейной регрессии показывают скорость изменения зависимой переменной по данному фактору, при фиксированных остальных факторах (в линейной модели эта скорость постоянна):

# **Глава 2. Практическая часть**

## **2.1. Предобработка данных**

Объединяем 2 файла датасета:

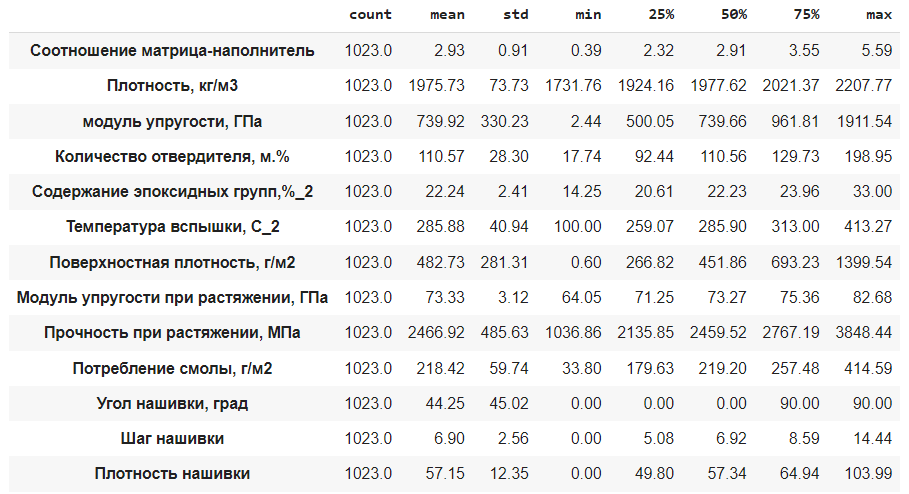
df = data\_bp.merge(data\_nup, left\_index=True, right\_index=True, how='inner')

Выводим первые 5 строк:



Выводим описательную статистику по датасету:

df.describe().round(2).T



Проверяем:

* Корреляцию
* Распределение

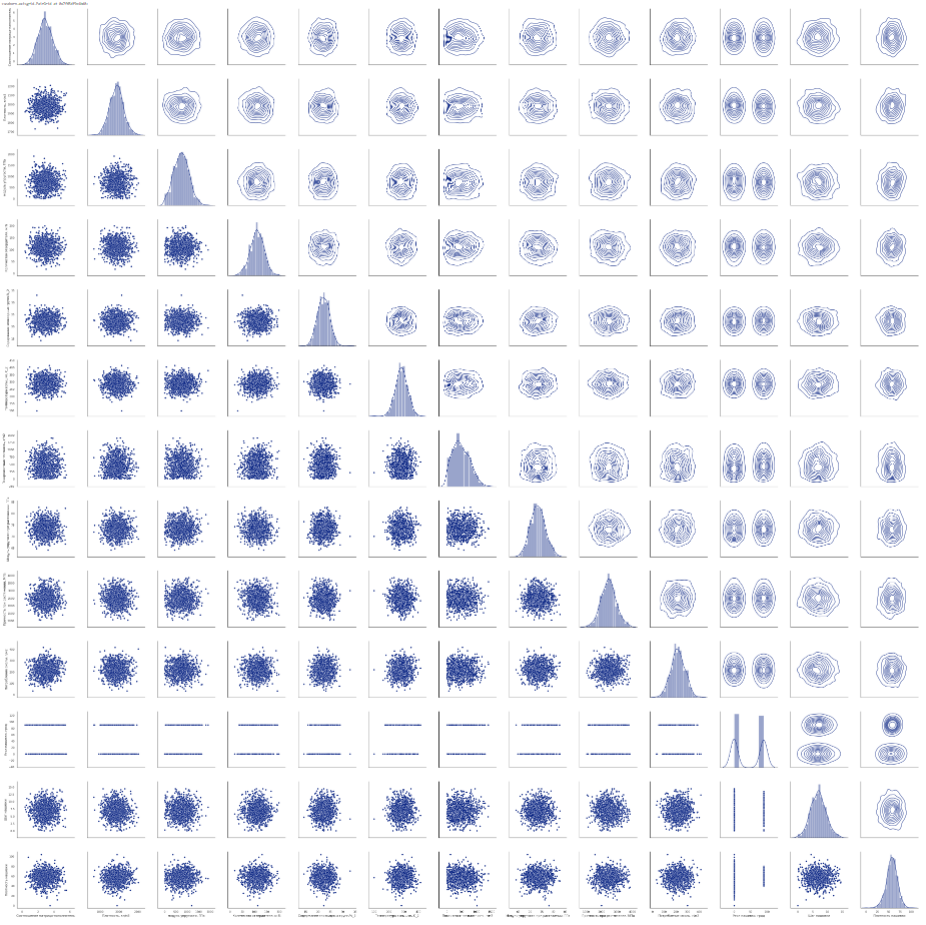
sns.set(style = 'white', palette ='dark')

grids = sns.PairGrid(df, diag\_sharey = False, height =4)

grids.map\_diag(sns.distplot, kde = True)

grids.map\_upper(sns.kdeplot)

grids.map\_lower(sns.scatterplot)



Распределение приближено к нормальному. Отсутствует корреляция

Дополнительно строим тепловую матрицу:

plt.figure(figsize=(16, 6))

mask = np.triu(np.ones\_like(df.corr(), dtype=np.bool))

heatmap = sns.heatmap(df.corr(), mask=mask, vmin=-1, vmax=1, annot=True, cmap='Blues')

heatmap.set\_title('Карта корреляции параметров', fontdict={'fontsize':18}, pad=16)



Проверим на выбросы. Построим ящик с усами. Предварительно надо нормализовать датасет.

Убираем столбцы и нормализуем датасет

minmax\_scaler = MinMaxScaler()

df\_normal = minmax\_scaler.fit\_transform(np.array(df\_normal[['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']]))

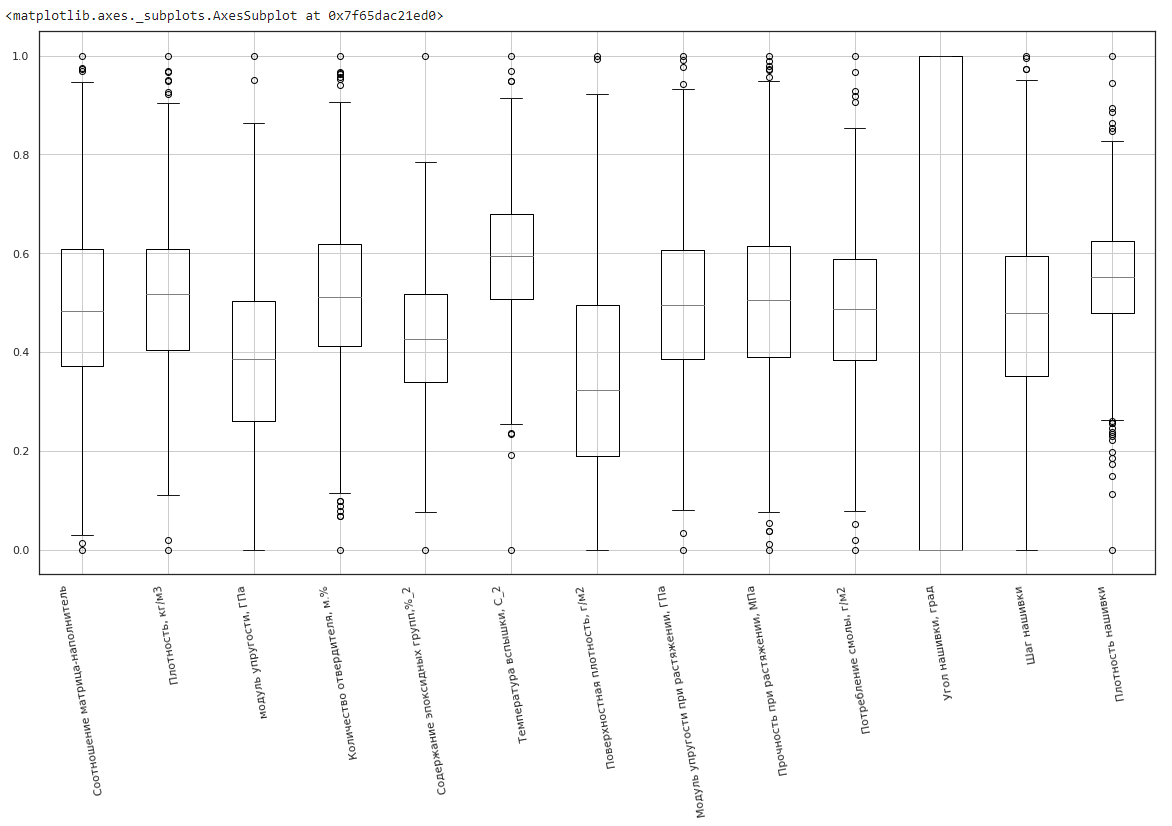
Возвращаем столбцы

dataset\_norm\_df = pd.DataFrame(data = df\_normal, columns = ['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'])

dataset\_norm\_df.head()



dataset\_norm\_df.boxplot(rot = 100, figsize=(20, 10))



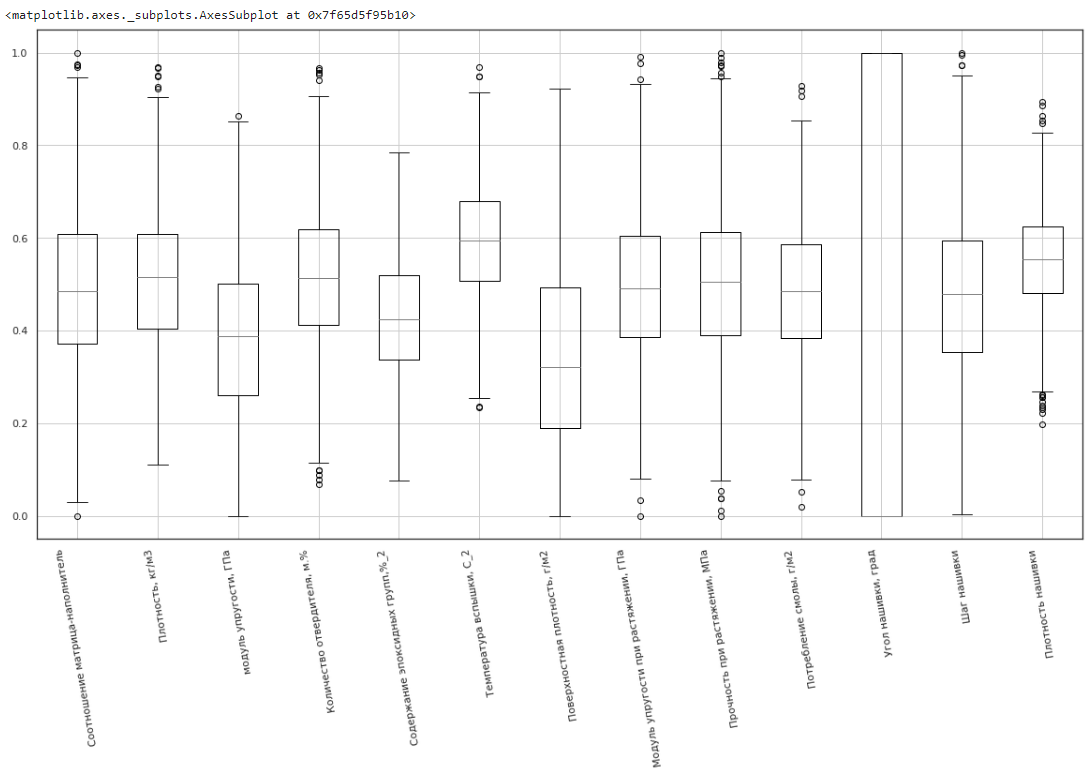
Убираем выбросы по правилу 3 сигм

sigma\_normal = dataset\_norm\_df[(np.abs(stats.zscore(dataset\_norm\_df)) < 3).all(axis=1)]

sigma\_normal



Строим повторно ящик с усами

Проведем прямую через точки для «Модуль упругости при растяжении, ГПа»

sns.pairplot(df\_MinMax, x\_vars=["Соотношение матрица-наполнитель", "Прочность при растяжении, МПа", 'Плотность, кг/м3',

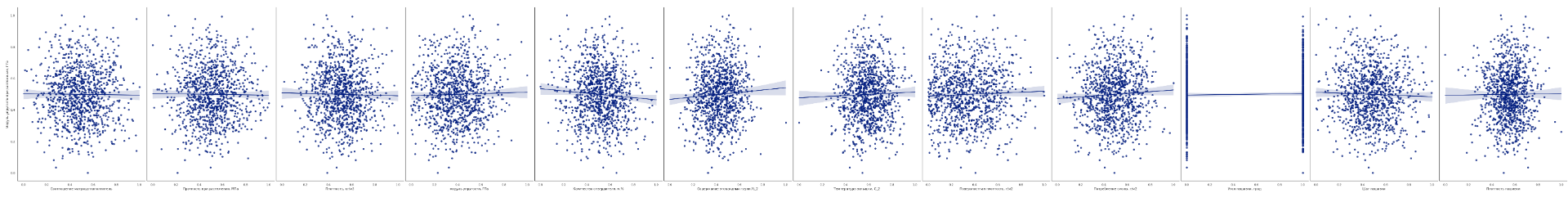
                                'модуль упругости, ГПа', 'Количество отвердителя, м.%', 'Содержание эпоксидных групп,%\_2',

                                'Температура вспышки, С\_2', 'Поверхностная плотность, г/м2', 'Потребление смолы, г/м2',

                                'Угол нашивки, град', 'Шаг нашивки', 'Плотность нашивки'], y\_vars='Модуль упругости при растяжении, ГПа', size=7, aspect=0.8,kind = 'reg')

plt.savefig("pairplot.jpg")

plt.show()



Проведем прямую через точки для «Прочность при растяжении, МПа»

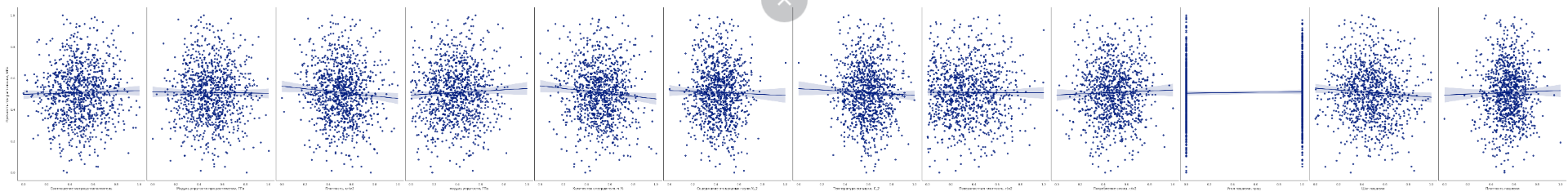
sns.pairplot(dataset\_norm\_df, x\_vars=["Соотношение матрица-наполнитель", "Модуль упругости при растяжении, ГПа", 'Плотность, кг/м3',

                                'модуль упругости, ГПа', 'Количество отвердителя, м.%', 'Содержание эпоксидных групп,%\_2',

                                'Температура вспышки, С\_2', 'Поверхностная плотность, г/м2', 'Потребление смолы, г/м2',

                                'Угол нашивки, град', 'Шаг нашивки', 'Плотность нашивки'], y\_vars='Прочность при растяжении, МПа', size=10, aspect=0.7,kind = 'reg')

plt.show()



## **2.1 Обучение моделей**

Создаем переменные для наших тестовых и обучающих выборок.

Не нормализованный датасет:

* х\y\_1 - Модуль упругости при растяжении, ГПа
* х\y\_2 - Прочность при растяжении, МПа
* х\y\_all - объединенный

Нормализованный датасет:

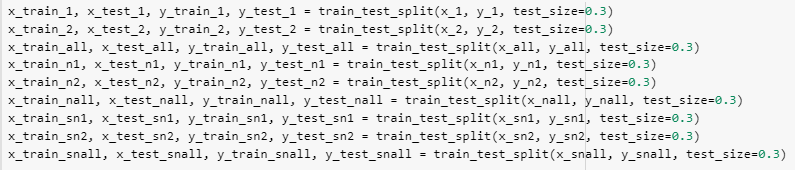
* х\y\_n1 - Модуль упругости при растяжении, ГПа
* х\y\_n2 - Прочность при растяжении, МПа
* х\y\_nall – объединенный

Нормализованный датасет без выбросов:

* х\y \_sn1 - Модуль упругости при растяжении, ГПа
* х\y \_sn2 - Прочность при растяжении, МПа
* х\y \_snall – объединенный



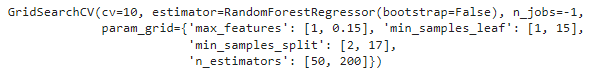
Разбиваем датасет на тестовую и обучающую выборку



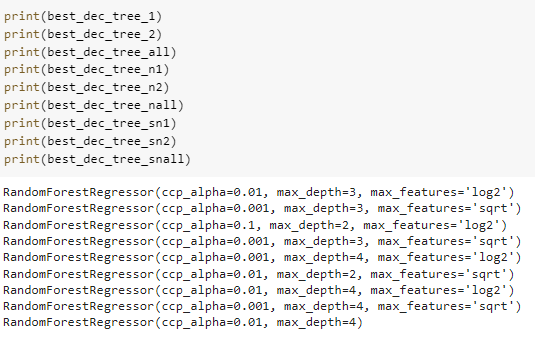
Случайный лес:

Для среднеквадратичной и абсолютной ошибки будем использовать параметры:

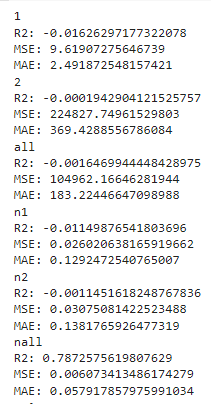


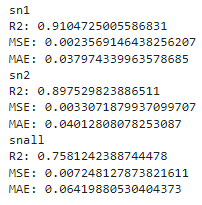


Для score будем использовать лучшее дерево решений

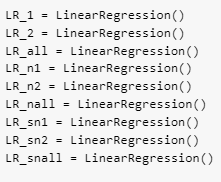


Выводим результаты предсказания:

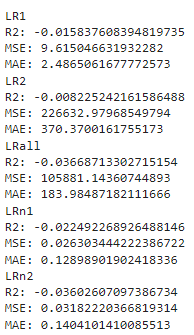


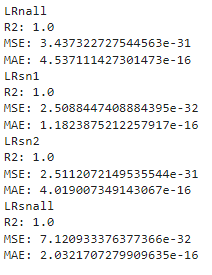


Линейная регрессия



Выводим результат предсказания:





Большая часть моделей показывают результат хуже случайного. Из всех попыток можно выделить результаты «best\_dec\_tree» на нормализованных данных и на данных без выбросов.

**2.2.** **Нейронная сеть для рекомендации соотношения матрица-наполнитель**

ав

## Разработка приложения

## Репозиторий и результаты

# Заключение

# Библиографический список