

**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»

Тема: Прогнозирование конечных свойств новых материалов
(композиционных материалов)

Слушатель

Боркунов Юрий Александрович

Москва, 2023 год

Содержание

1. Введение.....	3
2. Аналитическая часть	
2.1. Постановка задачи	
2.2. Описание используемых методов.....	
2.3. Разведочный анализ данных.....	
3. Практическая часть	
3.1. Предобработка данных.....	
3.2. Разработка, обучение и тестирование моделей машинного обучения	
3.3. Написание нейронной сети, которая будет рекомендовать соотношение матрица-наполнитель.....	
3.4. Разработка приложения.....	
3.5. Создание удаленного репозитория и загрузка результатов работы на него.....	
4. Заключение.....	
5. Список литературы.....	

Введение

Композиционный материал - неоднородный сплошной материал, состоящий из двух или более компонентов, среди которых можно выделить армирующие элементы, обеспечивающие необходимые механические характеристики материала, и матрицу, обеспечивающую совместную работу армирующих элементов. Механическое поведение композита определяется соотношением свойств армирующих элементов и матрицы, а также прочностью связи между ними. Эффективность и работоспособность материала зависят от правильного выбора исходных компонентов и технологии их совмещения, призванной обеспечить прочную связь между компонентами при сохранении их первоначальных характеристик. В результате совмещения армирующих элементов и матрицы образуется комплекс свойств композита, не только отражающий исходные характеристики его компонентов, но и включающий свойства, которыми изолированные компоненты не обладают. В частности, наличие границ раздела между армирующими элементами и матрицей существенно повышает трещиностойкость материала, и в композитах, в отличие от металлов, повышение статической прочности приводит не к снижению, а, как правило, к повышению характеристик вязкости разрушения.

Преимущества композиционных материалов:

- высокая удельная прочность;
- высокая жёсткость (модуль упругости 130...140 ГПа);
- высокая износостойкость;
- высокая усталостная прочность;

Из КМ возможно изготовить размеростабильные конструкции, причём, разные классы композитов могут обладать одним или несколькими преимуществами.

Наиболее частые недостатки композиционных материалов:

- высокая стоимость;
- анизотропия свойств;
- повышенная наукоёмкость производства, необходимость специального дорогостоящего оборудования и сырья, а следовательно развитого промышленного производства и научной базы страны.

Классификация композиционных материалов.

Композиты - многокомпонентные материалы, состоящие из полимерной, металлической, углеродной, керамической или др. основы (матрицы), армированной наполнителями из волокон, нитевидных кристаллов, тонкодисперсных частиц и др. Путем подбора состава и свойств наполнителя и матрицы (связующего), их соотношения, ориентации наполнителя можно получить материалы с требуемым сочетанием эксплуатационных и технологических свойств. Использование в одном материале нескольких матриц (полиматричные композиционные материалы) или наполнителей различной природы (гибридные композиционные материалы) значительно расширяет возможности регулирования свойств композиционных материалов. Армирующие наполнители воспринимают основную долю нагрузки композиционных материалов.

По структуре наполнителя композиционные материалы подразделяют на волокнистые (армированы волокнами и нитевидными кристаллами), слоистые (армиро-

ваны пленками, пластинками, слоистыми наполнителями), дисперсноармированные, или дисперсно-упрочненные (с наполнителем в виде тонкодисперсных частиц). Матрица в композиционных материалах обеспечивает монолитность материала, передачу и распределение напряжения в наполнителе, определяет тепло-, влаго-, огне- и хим. стойкость.

По природе матричного материала различают полимерные, металлические, углеродные, керамические и др. композиты.

Основным недостатком композиционных материалов с одно и двумерным армированием является низкое сопротивление межслойному сдвигу и поперечному обрыву. Этого лишены материалы с объемным армированием.

методов изготовления полимерных и металлических волокнистых и слоистых композиционных материалов - выращивание кристаллов наполнителя в матрице непосредственно в процессе изготовления деталей. Такой метод применяют, напр., при создании эвтектических жаропрочных сплавов на основе Ni и Co. Легирование расплавов карбидными и интерметаллическими соединениями, образующими при охлаждении в контролируемых условиях волокнистые или пластинчатые кристаллы, приводит к упрочнению сплавов и позволяет повысить температуру их эксплуатации на 60-80°C. Композиционные материалы на основе углерода сочетают низкую плотность с высокой теплопроводностью, хим. стойкостью, постоянством размеров при резких перепадах температур, а также с возрастанием прочности и модуля упругости при нагреве до 2000 °C в инертной среде. Высокопрочные композиционные материалы на основе керамики получают при армировании волокнистыми наполнителями, а также металлическими и керамическими дисперсными частицами. Армирование непрерывными волокнами SiC позволяет получать композиционные материалы, характеризующиеся повышенной вязкостью, прочностью на изгиб и высокой стойкостью к окислению при высоких температурах.

Использование композитов в качестве конструкционных, теплозащитных, антифрикционных, радио - и электротехнических и др. материалов позволяет снизить массу конструкции, повысить ресурсы и мощности машин и агрегатов, создать принципиально новые узлы, детали и конструкции. Все виды композиционные материалы применяют в химической, текстильной, горнорудной, металлургической промышленности, машиностроении, на транспорте, для изготовления спортивного снаряжения и др.

Экономическая эффективность применения композиционных материалов.

Области применения композиционных материалов не ограничены. Они применяются в авиации для высоконагруженных деталей (обшивки, лонжеронов, нервюр, панелей, лопаток компрессора и турбины и т. д.), в космической технике для узлов силовых конструкций аппаратов, для элементов жесткости, панелей, в автомобилестроении для облегчения кузовов, рессор, рам, панелей кузовов, бамперов и т. д., в горной промышленности (буровой инструмент, детали комбайнов и т. д.), в гражданском строительстве (пролеты мостов, элементы сборных конструкций высотных сооружений и т. д.) и в других областях народного хозяйства.

Применение композиционных материалов обеспечивает новый качественный скачек в увеличении мощности двигателей, энергетических и транспортных устано-

вок, уменьшении массы машин и приборов. Композиционные материалы с неметаллической матрицей, а именно полимерные карбоволокониты используют в судо- и автомобилестроении (кузова гоночных машин, шасси, гребные винты); из них изготавливают подшипники, панели отопления, спортивный инвентарь, части ЭВМ. Высокомодульные карбоволокониты применяют для изготовления деталей авиационной техники, аппаратуры для химической промышленности, в рентгеновском оборудовании и другом. Карбоволокониты с углеродной матрицей заменяют различные типы графитов. Они применяются для тепловой защиты, дисков авиационных тормозов, химически стойкой аппаратуры. Изделия из карбоволоконитов применяют в авиационной и космической технике (профили, панели, роторы и лопатки компрессоров, лопасти винтов, трансмиссионные валы вертолетов и т. д.). Органоволокониты применяют в качестве изоляционного и конструкционного материала в электрорадиопромышленности, авиационной технике и т. д.

2. Аналитическая часть

2.1. Постановка задачи

В процессе выполнения данной выпускной квалификационной работы необходимо продемонстрировать навыки, полученные в данном курсе на примере решения задачи прогнозирования конечных свойств новых композиционных материалов.

А, именно:

- провести анализ и предобработку данных;
- описать методы которые предполагается использовать для решения поставленной задачи;
- разработать, обучить и протестировать модель;
- написать нейронную сеть, которая будет рекомендовать соотношение матрица-наполнитель;
- разработать приложение.

2.2.Описание используемых методов

Перед нами стоит задача регрессии – прогноз на основе выборки объектов с различными признаками. Поэтому в рамках решения данной задачи были применены следующие методы:

- Линейная регрессия;
- Метод k-ближайших соседей;
- Случайный лес.

Линейная регрессия в Машинном обучении – это подход к моделированию отношений между Целевой переменной (Target Variable) и одной или несколькими "предсказывающими" переменными (Predictor Variable). Проще говоря, это «линия наилучшего соответствия», которая помогает спрогнозировать положение других точек в будущем.

Метод k-ближайших соседей - метрический алгоритм для автоматической классификации объектов или регрессии. На интуитивном уровне суть метода проста: посмотри на соседей вокруг, какие из них преобладают, таковым ты и являешься.

Случайный лес, как и следует из его названия, состоит из большого количества отдельных деревьев решений, которые работают как ансамбль методов. Каждое дерево в случайном лесу возвращает прогноз класса, и класс с наибольшим количеством голосов становится прогнозом леса. Фундаментальная концепция в основе случайного леса проста, но сильна — это мудрость толпы. Большое число относительно некоррелированных деревьев, работающих совместно, будет превосходить любую из их отдельных составляющих.

2.3.Разведочный анализ данных

В исходных данных мы имеем датасет с двумя Excel-таблицами. Далее с ними мы и будем работать.

Описательная статистика датасета.

```
#Описательная статистика датасета
# count - количество значений
#mean - среднее значение
#std - стандартное отклонение
#min - минимум
#25% - верхнее значение первого квартиля
#50% - медиана
#75% - верхнее значение третьего квартиля
#max - максимум
df.describe().T
```

	count	mean	std	min	25%	50%	75%	max
Соотношение матрица-наполнитель	1023.0	2.930366	0.913222	0.389403	2.317887	2.906878	3.552660	5.591742
Плотность, кг/м3	1023.0	1975.734888	73.729231	1731.764635	1924.155467	1977.621657	2021.374375	2207.773481
модуль упругости, ГПа	1023.0	739.923233	330.231581	2.436909	500.047452	739.664328	961.812526	1911.536477
Количество отвердителя, м.%	1023.0	110.570769	28.295911	17.740275	92.443497	110.564840	129.730366	198.953207
Содержание эпоксидных групп, %_2	1023.0	22.244390	2.406301	14.254985	20.608034	22.230744	23.961934	33.000000
Температура вспышки, C_2	1023.0	285.882151	40.943260	100.000000	259.066528	285.896812	313.002106	413.273418
Поверхностная плотность, г/м2	1023.0	482.731833	281.314690	0.603740	266.816645	451.864365	693.225017	1399.542362
Модуль упругости при растяжении, ГПа	1023.0	73.328571	3.118983	64.054061	71.245018	73.268805	75.356612	82.682051
Прочность при растяжении, МПа	1023.0	2466.922843	485.628006	1036.856605	2135.850448	2459.524526	2767.193119	3848.436732
Потребление смолы, г/м2	1023.0	218.423144	59.735931	33.803026	179.627520	219.198882	257.481724	414.590628
Угол нашивки	1023.0	0.491691	0.500175	0.000000	0.000000	0.000000	1.000000	1.000000
Шаг нашивки	1023.0	6.899222	2.563467	0.000000	5.080033	6.916144	8.586293	14.440522
Плотность нашивки	1023.0	57.153929	12.350969	0.000000	49.799212	57.341920	64.944961	103.988901

Рисунок1- таблица с основными параметрами датасета

Произведено построение диаграмм с целью выявления характера взаимосвязи между переменными и последующего подбора моделей машинного обучения, подходящих для решения задачи.

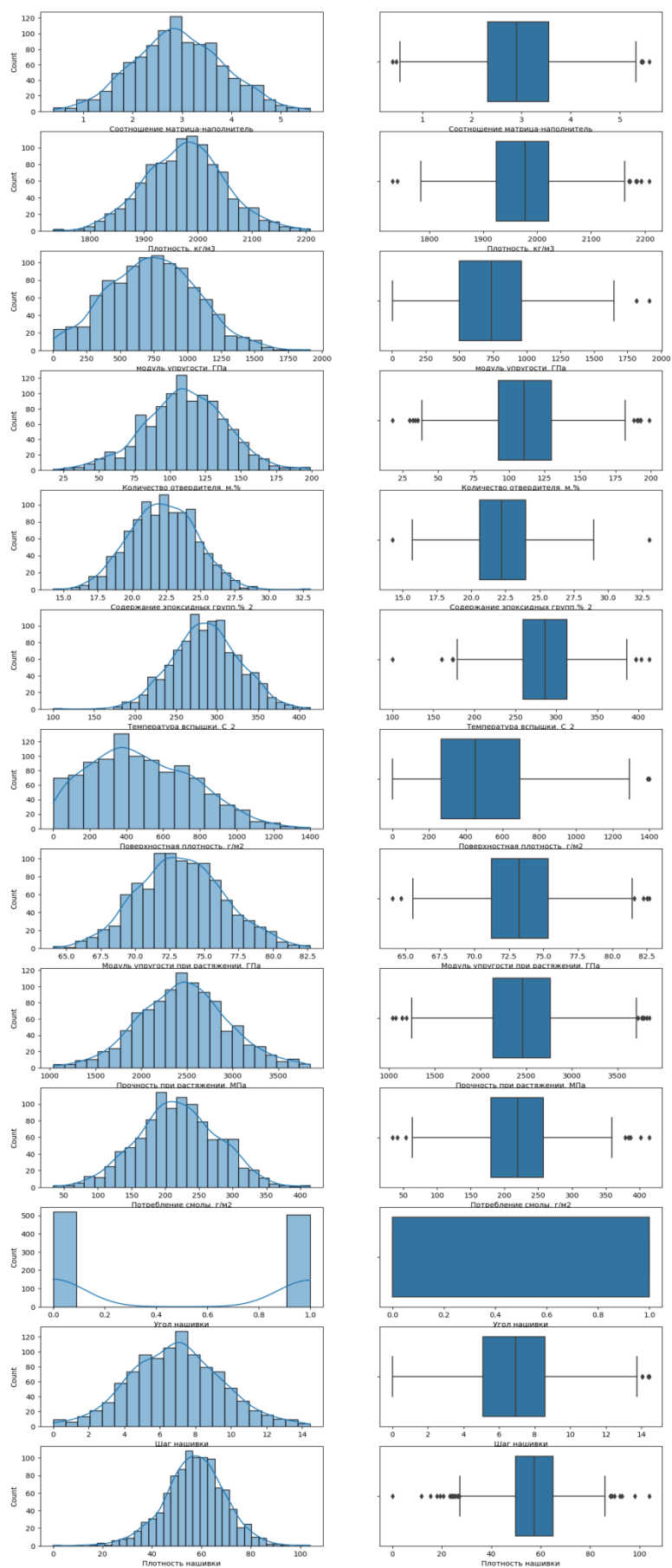


Рисунок 2 - Гистограммы распределения параметров переменных и "ящик с усами"

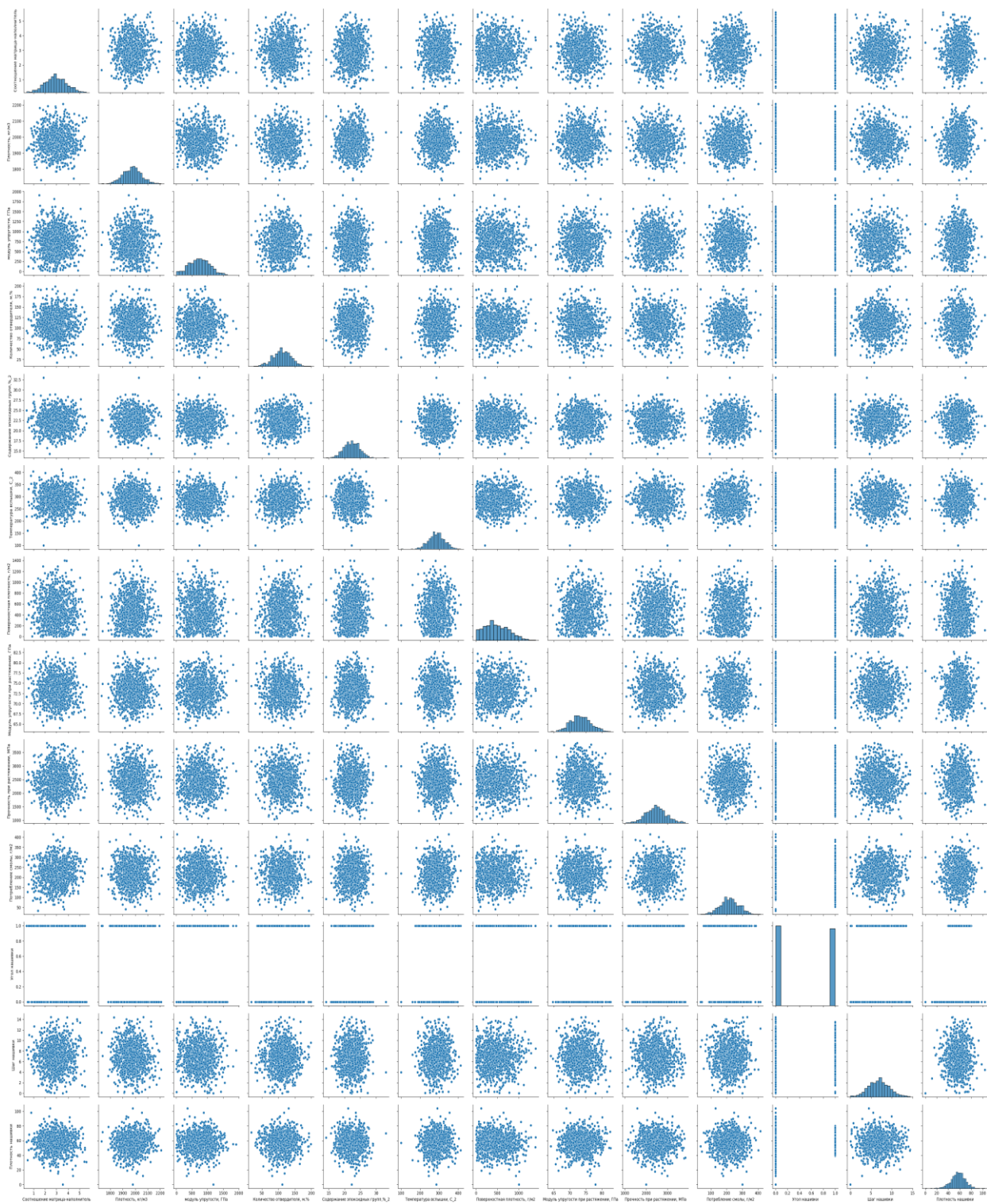


Рисунок 3 - Попарные графики рассеяния точек

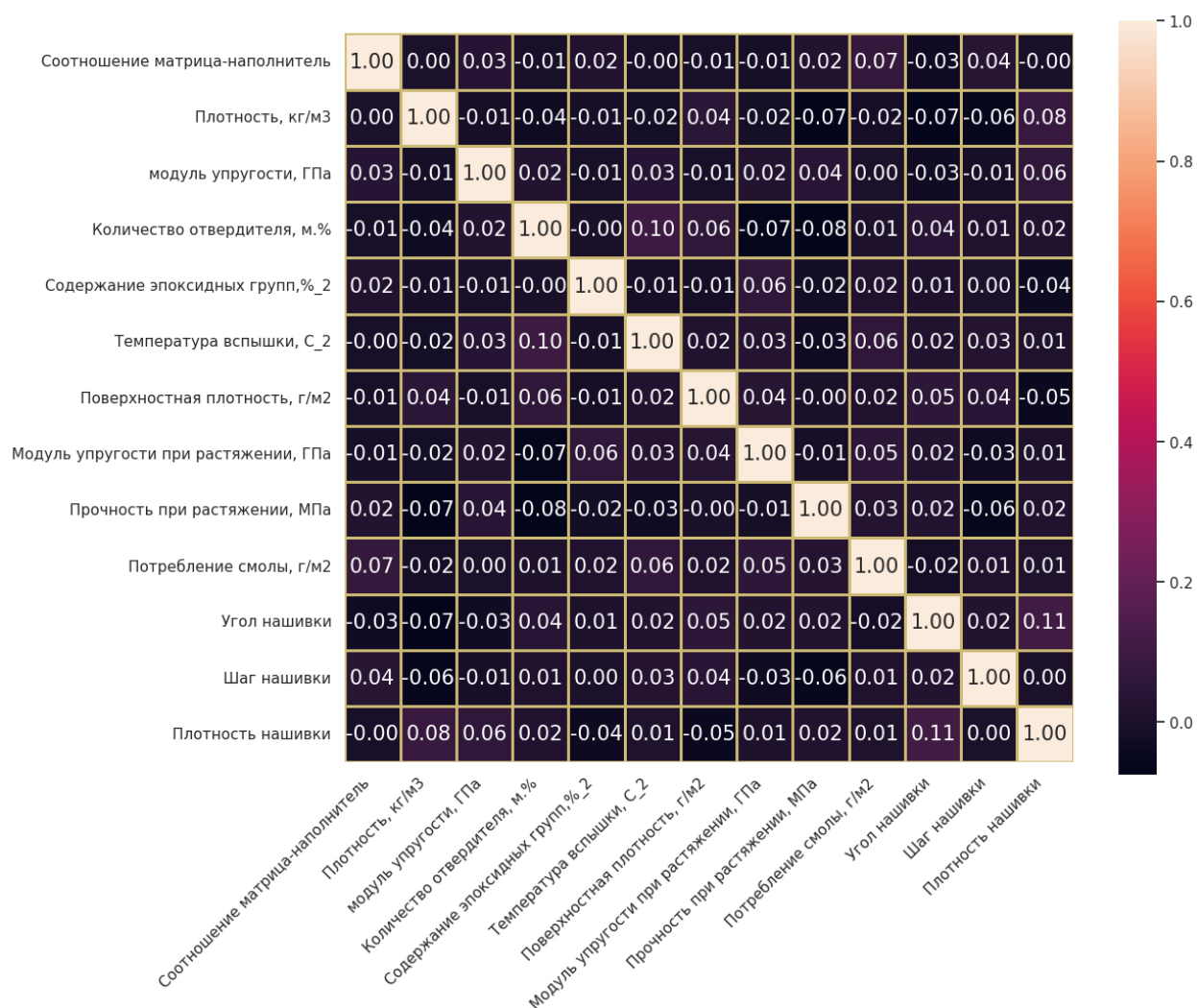


Рисунок 4 - Корреляционная матрица,представляющую собой тепловую карту, в которой измеряется линейная зависимость между парой признаков

Произведено обнаружение выбросов методом трех сигм и межквартильного расстояния. Правило трех сигм (3-sigma rule) - правило, утверждающее, что вероятность того, что случайная величина отклонится от своего математического ожидания более чем на три среднеквадратических отклонения. Межквартильный диапазон набора данных, часто сокращенно IQR, представляет собой разницу между первым квартилем (25-й процентиль) и третьим квартилем (75-й процентиль) набора данных.

Далее осуществлена очистка данных от выбросов методом межквартильного расстояния.

3.Практическая часть

3.1.Предобработка данных

Произведено объединение двух Excel-таблиц с данными в один датасет.

В параметре "угол нашивки,Град" только 2 значения, поэтому для удобства мы определили переменные как 0 и 1 и переименовали в “угол нашивки”.

	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	Шаг нашивки	Плотность нашивки
0	1.857143	2030.000000	738.736842	30.000000	22.267857	100.000000	210.000000	70.000000	3000.000000	220.000000	0	4.000000	57.000000
1	1.857143	2030.000000	738.736842	50.000000	23.750000	284.615385	210.000000	70.000000	3000.000000	220.000000	0	4.000000	60.000000
2	1.857143	2030.000000	738.736842	49.900000	33.000000	284.615385	210.000000	70.000000	3000.000000	220.000000	0	4.000000	70.000000
3	1.857143	2030.000000	738.736842	129.000000	21.250000	300.000000	210.000000	70.000000	3000.000000	220.000000	0	5.000000	47.000000
4	2.771331	2030.000000	753.000000	111.860000	22.267857	284.615385	210.000000	70.000000	3000.000000	220.000000	0	5.000000	57.000000
...
1018	2.271346	1952.087902	912.855545	86.992183	20.123249	324.774576	209.196700	73.090961	2387.292495	125.007669	1	9.076380	47.019770
1019	3.444022	2050.089171	444.732634	145.981978	19.599769	254.215401	350.660830	72.920827	2360.392784	117.730099	1	10.565614	53.750790
1020	3.280604	1972.372865	416.836524	110.533477	23.967502	248.423047	740.142791	74.734344	2662.906040	236.606764	1	4.161154	67.629684
1021	3.705351	2066.799773	741.475517	141.397963	19.246945	275.779640	641.468152	74.042708	2071.715856	197.126067	1	6.313201	58.261074
1022	3.808020	1890.413468	417.316232	129.183416	27.474763	300.952708	758.747882	74.309704	2856.328932	194.754342	1	6.078902	77.434468

Рисунок 5 – Объединённый датасет

Произведена нормализация данных методом MinMaxScaler

	count	mean	std	min	25%	50%	75%	max
Соотношение матрица-наполнитель	922.0	0.499412	0.187858	0.0	0.371909	0.495189	0.629774	1.0
Плотность, кг/м3	922.0	0.502904	0.188395	0.0	0.368184	0.511396	0.624719	1.0
модуль упругости, ГПа	922.0	0.451341	0.201534	0.0	0.305188	0.451377	0.587193	1.0
Количество отвердителя, м.%	922.0	0.506200	0.186876	0.0	0.378514	0.506382	0.638735	1.0
Содержание эпоксидных групп,%_2	922.0	0.490578	0.180548	0.0	0.366571	0.488852	0.623046	1.0
Температура вспышки, С_2	922.0	0.516739	0.190721	0.0	0.386228	0.516931	0.646553	1.0
Поверхностная плотность, г/м2	922.0	0.373295	0.217269	0.0	0.204335	0.354161	0.538397	1.0
Модуль упругости при растяжении, ГПа	922.0	0.487343	0.196366	0.0	0.353512	0.483718	0.617568	1.0
Прочность при растяжении, МПа	922.0	0.503776	0.188668	0.0	0.373447	0.501481	0.624299	1.0
Потребление смолы, г/м2	922.0	0.507876	0.199418	0.0	0.374647	0.510143	0.642511	1.0
Угол нашивки	922.0	0.510846	0.500154	0.0	0.000000	1.000000	1.000000	1.0
Шаг нашивки	922.0	0.503426	0.183587	0.0	0.372844	0.506414	0.626112	1.0
Плотность нашивки	922.0	0.503938	0.193933	0.0	0.376869	0.504310	0.630842	1.0

Рисунок 6 – Нормализованный датасет

3.2.Разработка и обучение модели

Перед нами стоит задача регрессии – прогноз на основе выборки объектов с различными признаками. Построим модели для прогноза модуля упругости при растяжении и прочности при растяжении.

```
X_u = df_norm.drop(['Модуль упругости при растяжении, ГПа'], axis=1)
X_p = df_norm.drop(['Прочность при растяжении, МПа'], axis=1)
y_u = df_norm[['Модуль упругости при растяжении, ГПа']]
y_p = df_norm[['Прочность при растяжении, МПа']]

X_train_u, X_test_u, y_train_u, y_test_u = train_test_split(X_u, y_u, test_size=0.3, random_state=42)
X_train_p, X_test_p, y_train_p, y_test_p = train_test_split(X_p, y_p, test_size=0.3, random_state=42)
```

Рисунок 6 - Разбиваем данные на обучающую и тестовую выборки

Метод линейной регрессии.

Определим несколько параметров:

- R2(или Коэффициент детерминации)- это статистическая мера, которая показывает степень вариации зависимой переменной из-за независимой переменной;
- Средняя квадратическая ошибка;
- Средняя абсолютная ошибка.

```
model = LinearRegression()
model.fit(X_train_u,y_train_u)
predictions = model.predict(X_test_u)
r_sq = model.score(X_u, y_u)
print('coefficient of determination:', r_sq) # коэффициент детерминации. R2 (или Коэффициент детерминации)
print(
    'mean_squared_error : ', mean_squared_error(y_test_u, predictions)) #Средняя квадратическая ошибка
print(
    'mean_absolute_error : ', mean_absolute_error(y_test_u, predictions)) #Средняя абсолютная ошибка
```

```
coefficient of determination: 0.0054214924703653855
mean_squared_error : 0.042832542319332874
mean_absolute_error : 0.16952044227222318
```

```
model = LinearRegression()
model.fit(X_train_p,y_train_p)
predictions = model.predict(X_test_p)
r_sq = model.score(X_p, y_p)
print('coefficient of determination:', r_sq) # коэффициент детерминации
print(
    'mean_squared_error : ', mean_squared_error(y_test_p, predictions)) #средняя квадратическая ошибка
print(
    'mean_absolute_error : ', mean_absolute_error(y_test_p, predictions)) #Средняя абсолютная ошибка
```

```
coefficient of determination: 0.01040217342102201
mean_squared_error : 0.037238059902490565
mean_absolute_error : 0.1551905647370025
```

Рисунок 7 – Оценки производительности модели для модуля упругости при растяжении и прочности при растяжении

Получим прогнозы и выведем соответствующие графики.

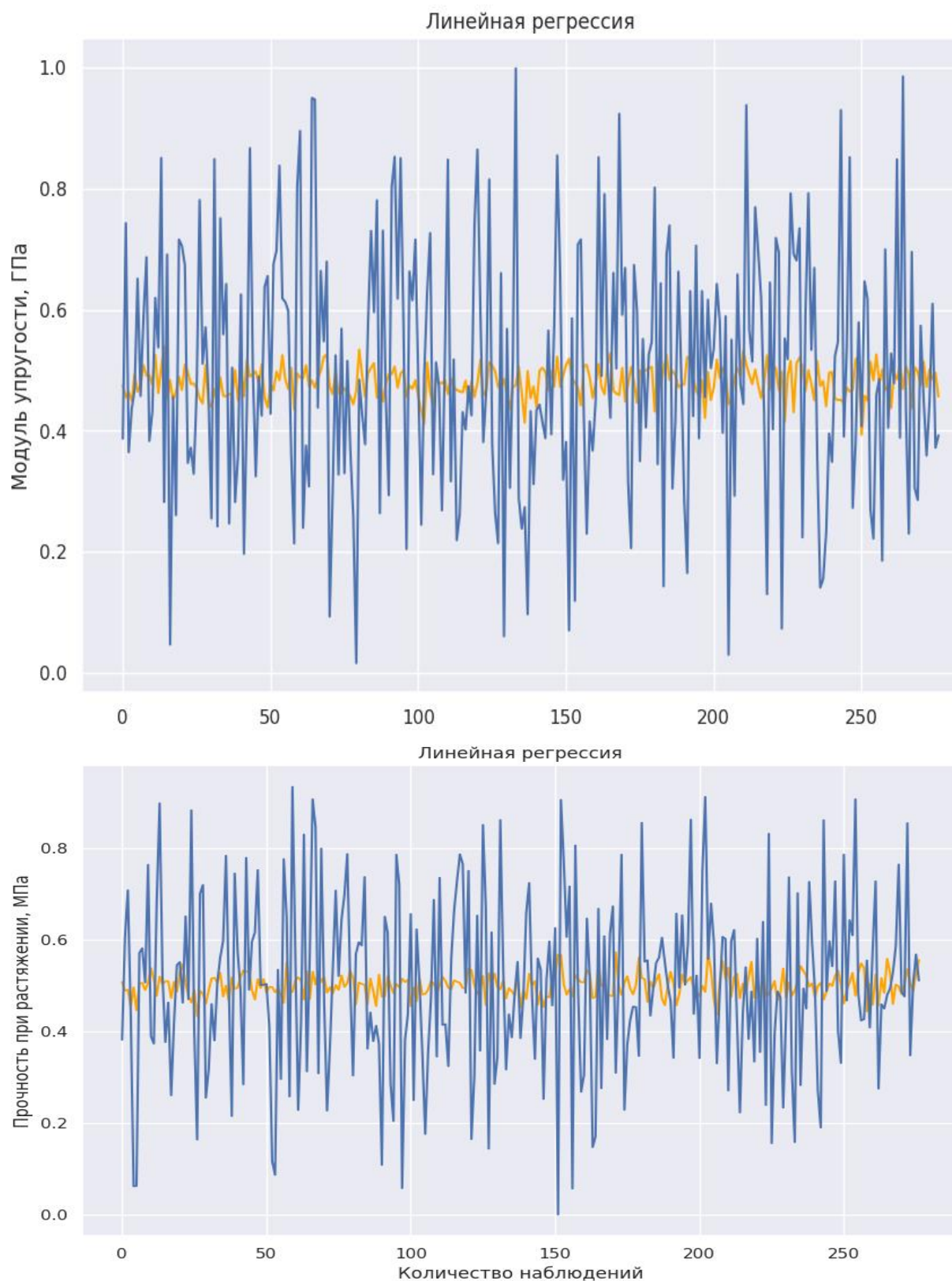


Рисунок 8 - графики прогноза и тестовых данных

Метод k-ближайших соседей.

```

knn = KNeighborsRegressor(n_neighbors=5)
knn.fit(X_train_u, y_train_u)
y_pred_u_knn = knn.predict(X_test_u)
mae_knr = mean_absolute_error(y_pred_u_knn, y_test_u)
mse_knn_elast = mean_squared_error(y_test_u, y_pred_u_knn)

print('K Neighbors Regressor Results Train:')
print("Test score: {:.2f}".format(knn.score(X_train_u, y_train_u)))# Оценка тренировочной выборки
print('K Neighbors Regressor Results:') # результаты регрессора соседей
print('KNN_MAE: ', round(mean_absolute_error(y_test_u, y_pred_u_knn))) # Средняя абсолютная ошибка
print('KNN_MAPE: {:.2f}'.format(mean_absolute_percentage_error(y_test_u, y_pred_u_knn))) # Средняя абсолютная ошибка в процентах
print('KNN_MSE: {:.2f}'.format(mse_knn_elast)) # средняя квадратическая ошибка
print("KNN_RMSE: {:.2f}".format(np.sqrt(mse_knn_elast))) # Среднеквадратическая ошибка
print("Test score: {:.2f}".format(knn.score(X_test_u, y_test_u)))# Оценка тестовой выборки

```

K Neighbors Regressor Results Train:
Test score: 0.18
K Neighbors Regressor Results:
KNN_MAE: 0
KNN_MAPE: 0.68
KNN_MSE: 0.05
KNN_RMSE: 0.23
Test score: -0.22

Рисунок 8 – оценочные характеристики модели

Здесь будет применена функция GridSearch — поиск лучших параметров в фиксированной сетке возможных значений.

CV – перекрёстная проверка (кросс-валидация, Cross-validation), метод, который показывает, что модель не переобучилась.

Результаты:

R2-score Модуль упругости при растяжении: -0.014

R2-score Прочность при растяжении: -0.002

R2<0-разработанная модель даёт прогноз даже хуже, чем простое усреднение.

Случайный лес.

Здесь будет применена функция RandomizedSearchCV. Она реализует метод «подгонки» и «оценки». В отличие от GridSearchCV, проверяются не все значения параметров, а из указанных распределений выбирается фиксированное количество значений параметров.

```

knn = KNeighborsRegressor(n_neighbors=5)
knn.fit(X_train_u, y_train_u)
y_pred_u_knn = knn.predict(X_test_u)
mae_knr = mean_absolute_error(y_pred_u_knn, y_test_u)
mse_knn_elast = mean_squared_error(y_test_u, y_pred_u_knn)

print('K Neighbors Regressor Results Train:')
print("Test score: {:.2f}".format(knn.score(X_train_u, y_train_u)))# Оценка тренировочной выборки
print('K Neighbors Regressor Results:') # результаты регрессора соседей
print('KNN_MAE: ', round(mean_absolute_error(y_test_u, y_pred_u_knn))) # Средняя абсолютная ошибка
print('KNN_MAPE: {:.2f}'.format(mean_absolute_percentage_error(y_test_u, y_pred_u_knn))) # Средняя абсолютная ошибка в процентах
print('KNN_MSE: {:.2f}'.format(mse_knn_elast)) # средняя квадратическая ошибка
print("KNN_RMSE: {:.2f}".format(np.sqrt(mse_knn_elast))) # Среднеквадратическая ошибка
print("Test score: {:.2f}".format(knn.score(X_test_u, y_test_u)))# Оценка тестовой выборки

```

K Neighbors Regressor Results Train:
Test score: 0.18
K Neighbors Regressor Results:
KNN_MAE: 0
KNN_MAPE: 0.68
KNN_MSE: 0.05
KNN_RMSE: 0.23
Test score: -0.22

Рисунок 9 – оценочные метрики для модуля упругости при растяжении

```

rfr = RandomForestRegressor(n_estimators=15, max_depth=7, random_state=33)
rfr.fit(X_train_p, y_train_p.values)
y_pred_p_forest = rfr.predict(X_test_p)
mae_rfr = mean_absolute_error(y_pred_p_forest, y_test_p)
mse_rfr_elast = mean_squared_error(y_test_p, y_pred_p_forest)

print('Random Forest Regressor Results Train:')
print("Test score: {:.2f}".format(rfr.score(X_train_p, y_train_p)))
print('Random Forest Regressor Results:')
print('RF_MAE: ', round(mean_absolute_error(y_test_p, y_pred_p_forest)))
print('RF_MAPE: {:.2f}'.format(mean_absolute_percentage_error(y_test_p, y_pred_p_forest)))
print('RF_MSE: {:.2f}'.format(mse_rfr_elast))
print("RF_RMSE: {:.2f}".format(np.sqrt(mse_rfr_elast)))
print("Test score: {:.2f}".format(rfr.score(X_test_p, y_test_p)))

Random Forest Regressor Results Train:
Test score: 0.48
Random Forest Regressor Results:
RF_MAE: 0
RF_MAPE: 8769553567468.01
RF_MSE: 0.04
RF_RMSE: 0.20

```

Рисунок 10 – оценочные метрики для прочности при растяжении

R2-score Модуль упругости при растяжении: -0.013

R2-score Прочность при растяжении: -0.002

3.3. Написание нейронной сети, которая будет рекомендовать соотношение матрица-наполнитель.

Сформируем выборки:

```
# Определяем входы и выходы
X_MN = df.drop(['Соотношение матрица-наполнитель'], axis=1)
y_MN = df[['Соотношение матрица-наполнитель']]
# Разбиваем выборки на обучающую и тестовую
X_train_MN, X_test_MN, y_train_MN, y_test_MN = train_test_split(X_MN, y_MN, test_size=0.3, random_state=1)
```

Рисунок 11 - Выборки

```
x_array = np.array(df[['Соотношение матрица-наполнитель']])
normalized_arr = preprocessing.normalize([x_array])
print(normalized_arr)
```

Рисунок 12 – Нормализация

Применим функцию активации SELU, она сочетает в себе оба преимущества классического RELU со свойствами самонормализации.

```
modelMN = Sequential()
modelMN.add(Dense(128))
modelMN.add(BatchNormalization())
modelMN.add(ReLU())
modelMN.add(Dense(128, activation='selu'))
modelMN.add(BatchNormalization())
modelMN.add(Dense(64, activation='selu'))
modelMN.add(BatchNormalization())
modelMN.add(Dense(32, activation='selu'))
modelMN.add(BatchNormalization())
modelMN.add(ReLU())
modelMN.add(Dense(16, activation='selu'))
modelMN.add(BatchNormalization())
modelMN.add(Dense(1))
modelMN.add(Activation('selu'))
```

Рисунок 12 – Слои нейронной сети

Построение модели и определение её параметров.

SGD-стохастический градиентный спуск с мини-пакетами — вариант, при котором коэффициенты меняются после обсчета N элементов выборки, то есть для каждой тренировочной итерации алгоритм выбирает случайное подмножество набора данных. Частота обновления параметров выше, меньше требуется оперативной памяти, эффективность вычислений высокая.

```
modelMN.compile(optimizer=tf.optimizers.SGD(learning_rate=0.02, momentum=0.5),
    loss='mean_absolute_error',metrics=['mse', "mape"])
```

```
tf.keras.callbacks.EarlyStopping(
    monitor='val_loss',
    min_delta=0,
    patience=10,
    verbose=1,
    mode='auto',
    baseline=None,
    restore_best_weights=False,
    ... start_from_epoch=0
)# Параметры ранней остановки для предотвращения "переоснащения" модели.
```

```
<keras.callbacks.EarlyStopping at 0x7feaf83fc070>
```

```
# Минимизируемая функция потерь
loss=keras.losses.SparseCategoricalCrossentropy()
```

```
historyMN=modelMN.fit(
    X_train_MN,
    y_train_MN,
    batch_size = 64,
    epochs=100,
    verbose=1,
    validation_split = 0.2,
)
```

Рисунок 13 – Формирование нейросети

Оценим модель.

```
modelMN.evaluate(X_test_MN, y_test_MN)
```

```
9/9 [=====] - 0s 2ms/step - loss: 0.7727 - mse: 0.9130 - mape: 33.5275
[0.7727224230766296, 0.9130035638809204, 33.527462005615234]
```

```
scores = modelMN.evaluate(X_train_MN,y_train_MN)
print("\n%s: %.2f%%" % (modelMN.metrics_names[1], scores[1]*100))
```

```
21/21 [=====] - 0s 1ms/step - loss: 0.7020 - mse: 0.7867 - mape: 31.4950
```

```
mse: 78.67%
```

Рисунок 14 – потери, средняя квадратическая ошибка, средняя абсолютная ошибка в процентах.

Тестовые и прогнозные значения

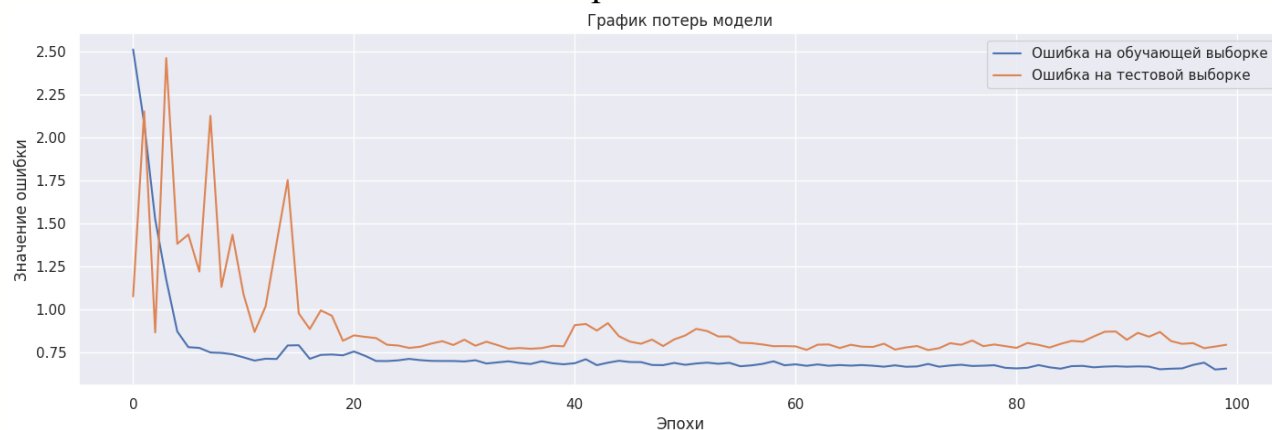


Рисунок 15

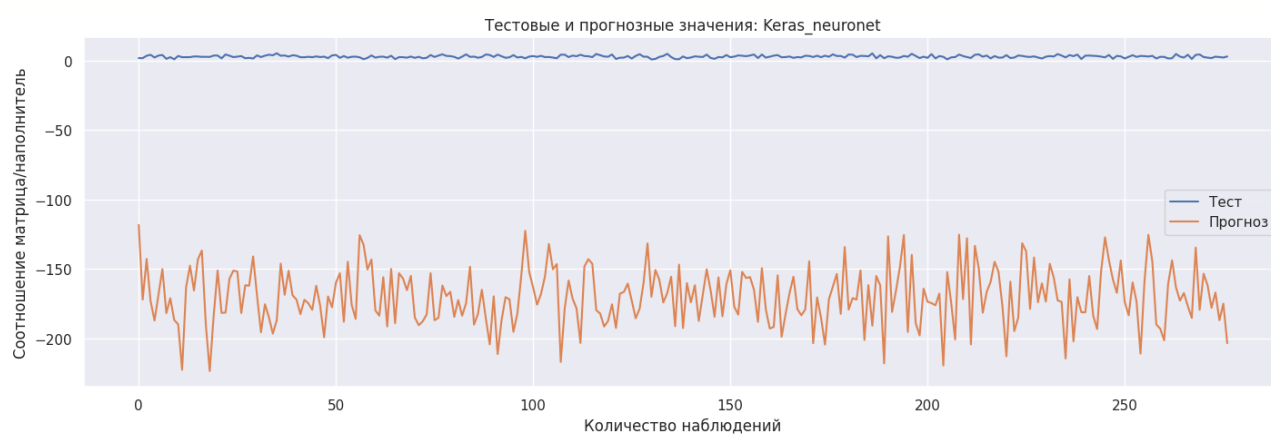


Рисунок 16

Средняя абсолютная ошибка.

9/9 [=====] - 0s 2ms/step —

loss: 0.7727 - mse: 0.9130 - mape: 33.5275

Model MAE: [0.7727224230766296, 0.9130035638809204, 33.527462005615234]

MAE среднего значения: Соотношение матрица-наполнитель 0.741552

Заключение

В процессе выполнения данной выпускной квалификационной работы были продемонстрированы навыки, полученные в данном курсе на примере решения задачи прогнозирования конечных свойств новых композиционных материалов.

Проведён анализ и предобработка данных, описаны методы которые предполагается использовать для решения поставленной задачи, разработаны, обучены и протестированы модели, написана нейронная сеть, которая будет рекомендовать соотношение матрица-наполнитель, разработано приложение.

Были сделаны следующие выводы. Распределение полученных данных в объединённом датасете близко к нормальному, но коэффициенты корреляции между парами признаков стремятся к нулю. Используемые при разработке моделей подходы не позволили получить сколько-нибудь достоверных прогнозов. Применённые модели регрессии не показали высокой эффективности в прогнозировании свойств композитов.

Созданная нейронная сеть, которая будет рекомендовать соотношение матрица-наполнитель также далека от идеала.

Список литературы

1. ГрасД. Data Science. Наука о данных с нуля: Пер. с англ. - 2-е изд., перераб. и доп. - СПб.: БХВ-Петербург, 2021. - 416 с.: ил.
2. Силен Дэви, Мейсман Арно, Али Мохамед. Основы Data Science и Big Data. Python и наука о данных. – СПб.: Питер, 2017. – 336 с.: ил.
3. Композиционные материалы : учебное пособие для вузов / Д. А. Иванов, А. И. Ситников, С. Д. Шляпин ; под редакцией А. А. Ильина. — Москва : Издательство Юрайт, 2019 — 253 с. — (Высшее образование). — Текст : непосредственный.
4. Документация по языку программирования python: – Режим доступа: <https://docs.python.org/3.8/index.html>.
5. Документация по библиотеке matplotlib: – Режим доступа: <https://matplotlib.org/stable/users/index.html>.
6. Документация по библиотеке seaborn: – Режим доступа: <https://seaborn.pydata.org/tutorial.html>.
7. Документация по библиотеке sklearn: – Режим доступа: https://scikit-learn.org/stable/user_guide.html.
8. Документация по библиотеке keras: – Режим доступа: <https://keras.io/api/>.
9. Документация по библиотеке numpy: – Режим доступа: <https://numpy.org/doc/1.22/user/index.html#user>.

19. Документация по библиотеке pandas: – Режим доступа:
https://pandas.pydata.org/docs/user_guide/index.html#user-guide.