ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

по курсу

«Data Science»

Тема: «Прогнозирование конечных свойств новых материалов (композиционных материалов)»

Слушатель: Москалева Юлия Михайловна

План работы

1. Аналитическая часть

1.1 Постановка задачи

1.2 Описание используемых методов

1.3 Разведочный анализ данных

2. Практическая часть

2.1 Предобработка данных

2.2 Разработка и обучение моделей

2.3 Рекомендательные нейросети для соотношения `` «матрица – наполнитель»

3. Заключение

Цели и задачи работы. Методы реализации

\

Целью настоящей работы является разработка модели для прогноза модуля упругости при растяжении, прочности при растяжении и рекомендации оптимального соотношения «матрица-наполнитель».

Для каждой колонки необходимо получить среднее, медианное значение, провести анализ и исключение выбросов, проверить наличие пропусков; пред обработать данные: удалить шумы и выбросы, сделать нормализацию и стандартизацию. Обучить несколько моделей для прогноза модуля упругости при растяжении и прочности при растяжении. Написать нейронную сеть, которая будет рекомендовать соотношение матрица-наполнитель. Разработать приложение с графическим интерфейсом, которое будет выдавать прогноз соотношения «матрица-наполнитель». Оценить точность модели на тренировочном и тестовом датасете. Создать репозиторий в GitHub и разместить код исследования.

Данная задача в рамках классификации категорий машинного обучения относится к машинному обучению с учителем и традиционно это задача регрессии. Цель любого алгоритма обучения с учителем — определить функцию потерь и минимизировать её, поэтому для наилучшего решения в процессе исследования были применены следующие методы:

метод опорных векторов; случайный лес; линейная регрессия; k-ближайших соседей;
Известные и широкоприменяемые методы, такие как градиентный бустинг,
дерево решений, стохастический градиентный спуск, лассо-регрессия, многослойный перцептрон, стохастический градиентный спуск в данной работе не применялись, поскольку показали свою низкую эффективность на обширном предварительном этапе исследования.

Разведочный анализ данных

Разведочный анализ данных. Гистограмма

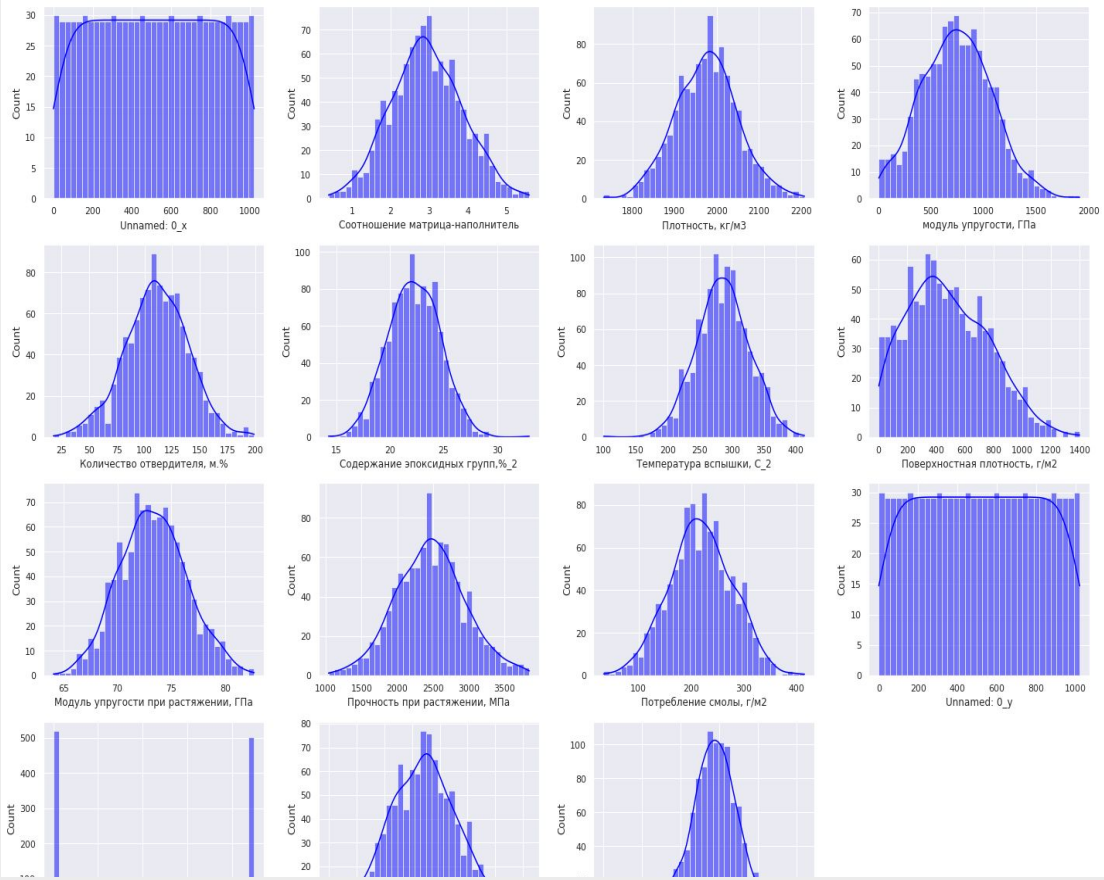
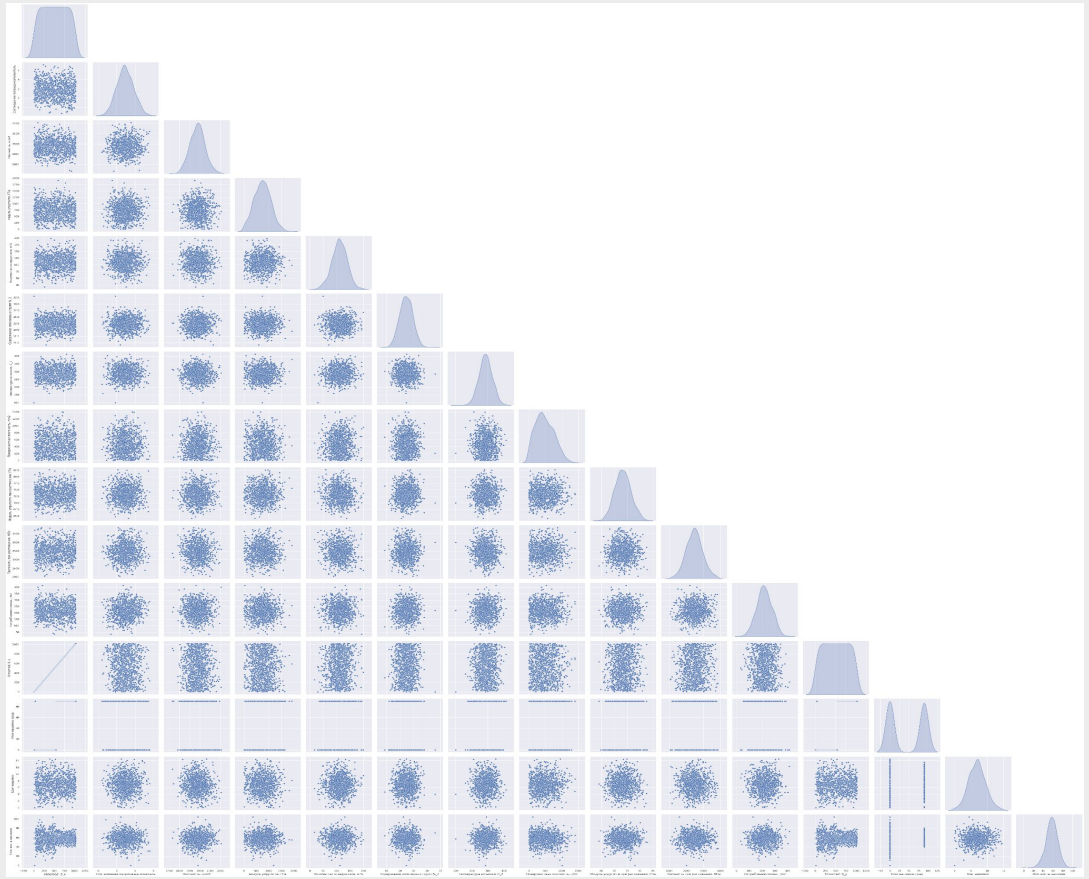


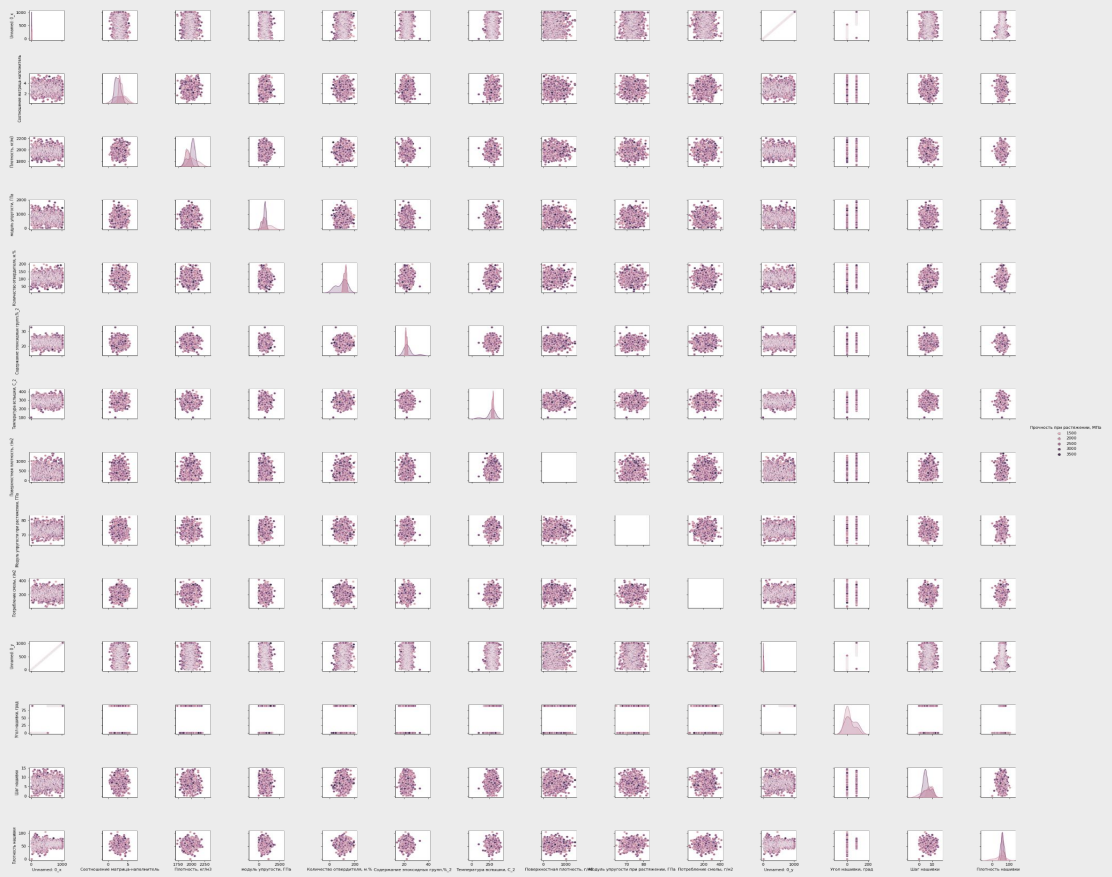
График попарного рассеивания точек



Разведочный анализ данных

График попарного рассеивания точек в привязке к целевой переменной “Модуль упругости при растяжении, ГПа”

График попарного рассеивания точек в привязке к целевой переменной “Прочность при растяжении, МПа”



Матрица корреляции данных и выбросы

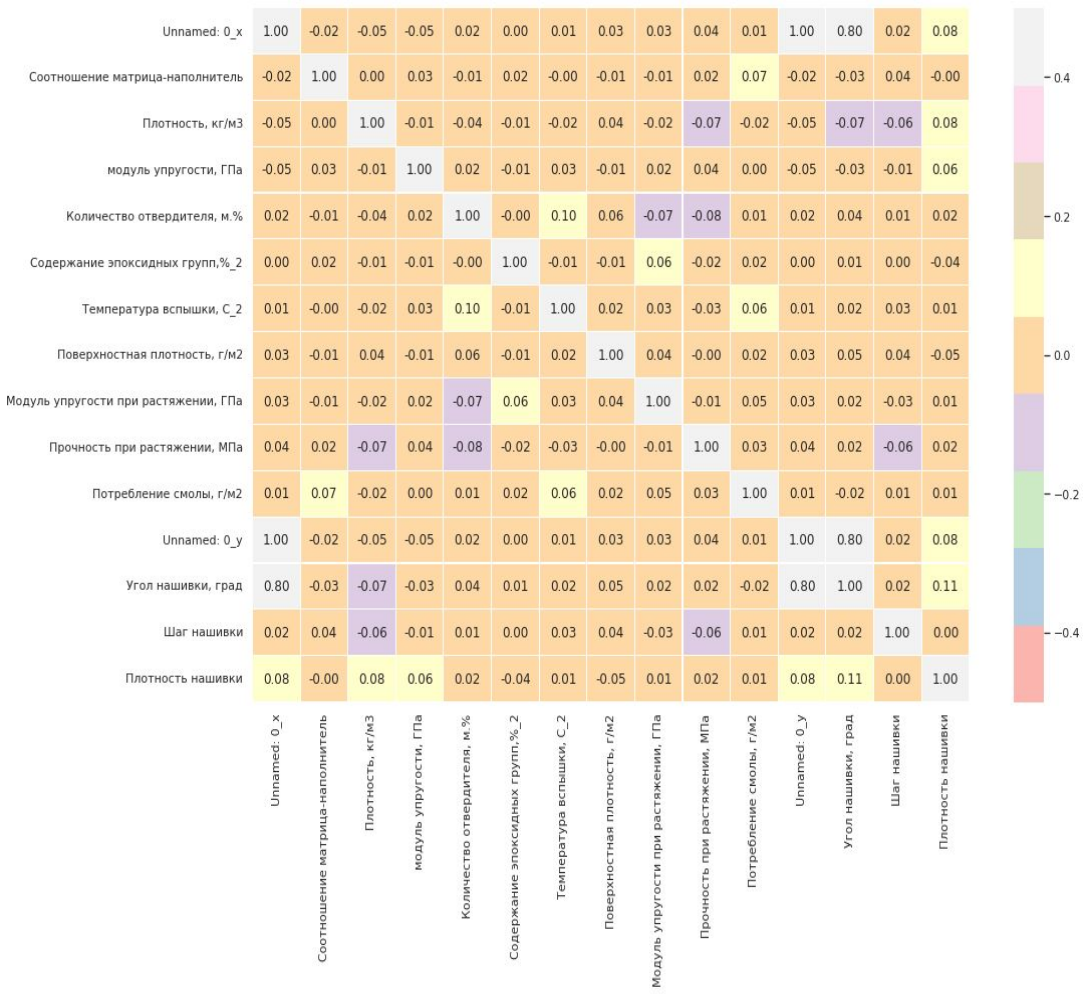
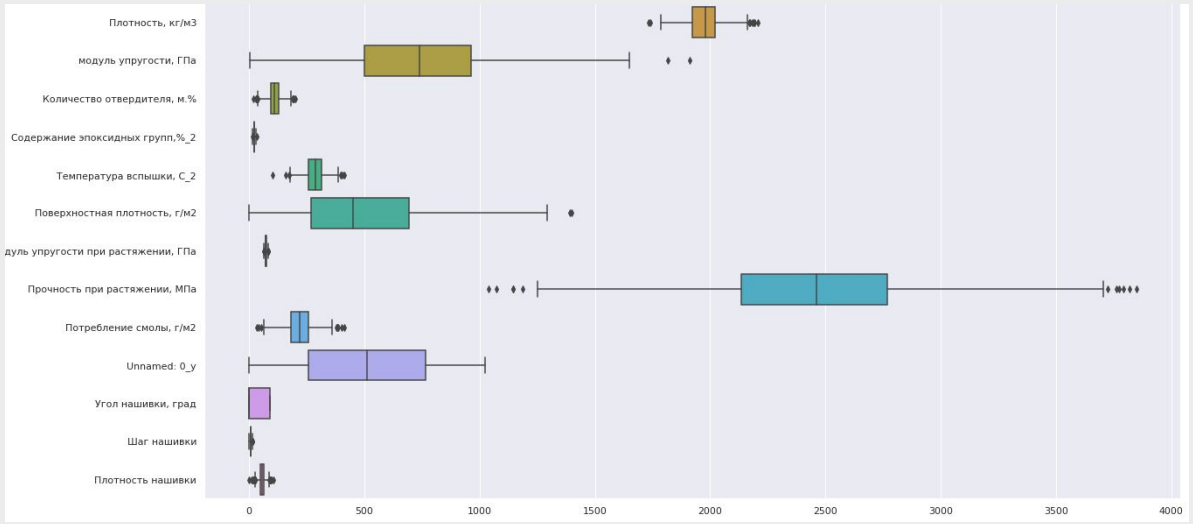
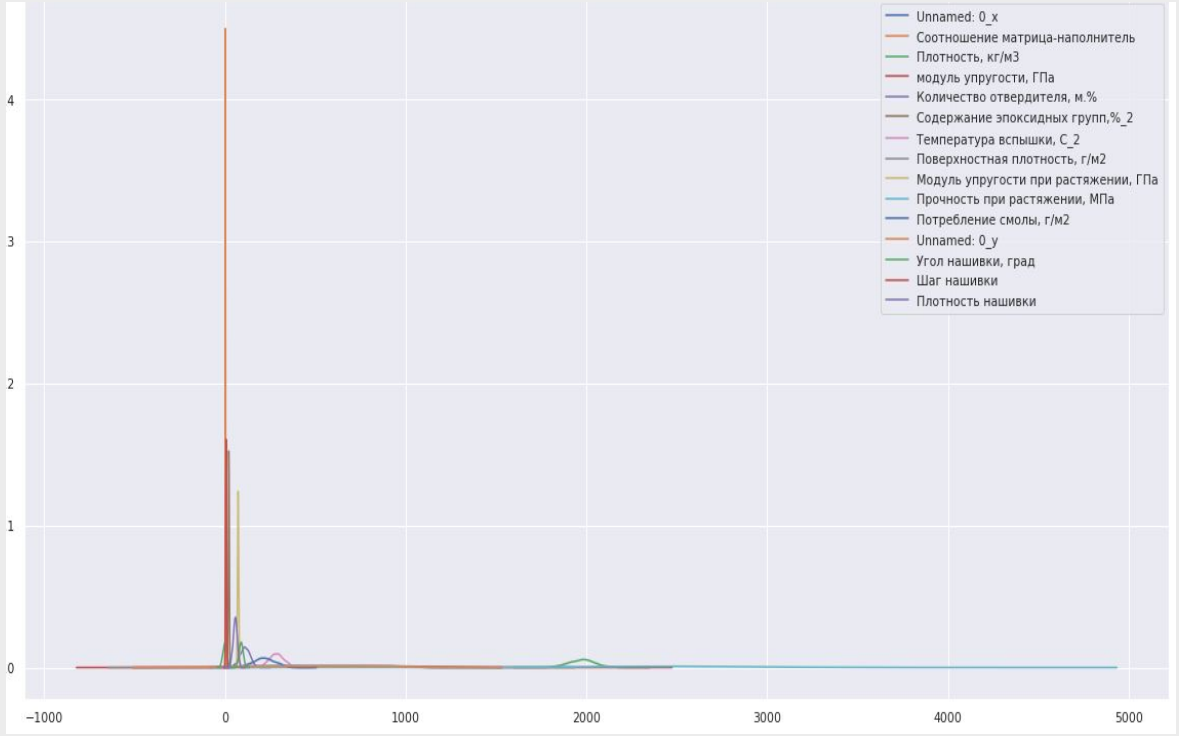


Диаграмма “Ящики с усами” до нормализации данных

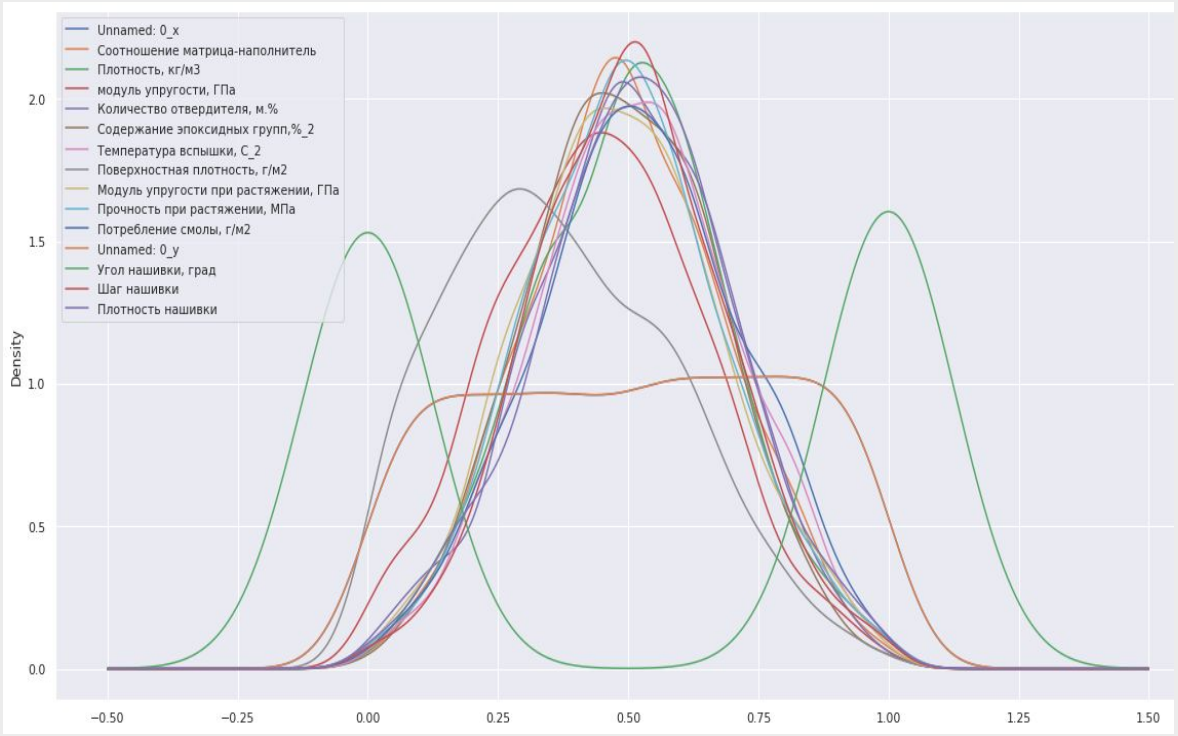


Оценка распределения

До нормализации

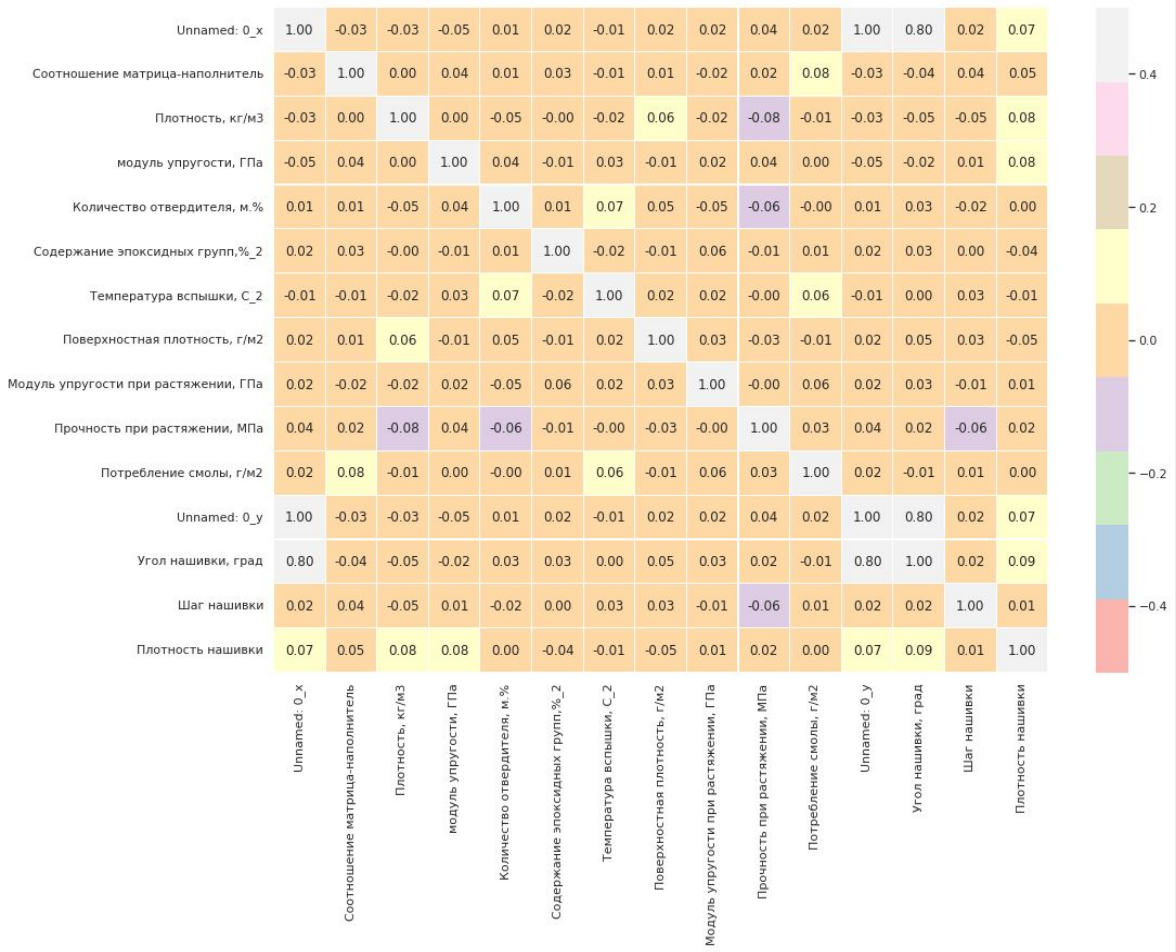


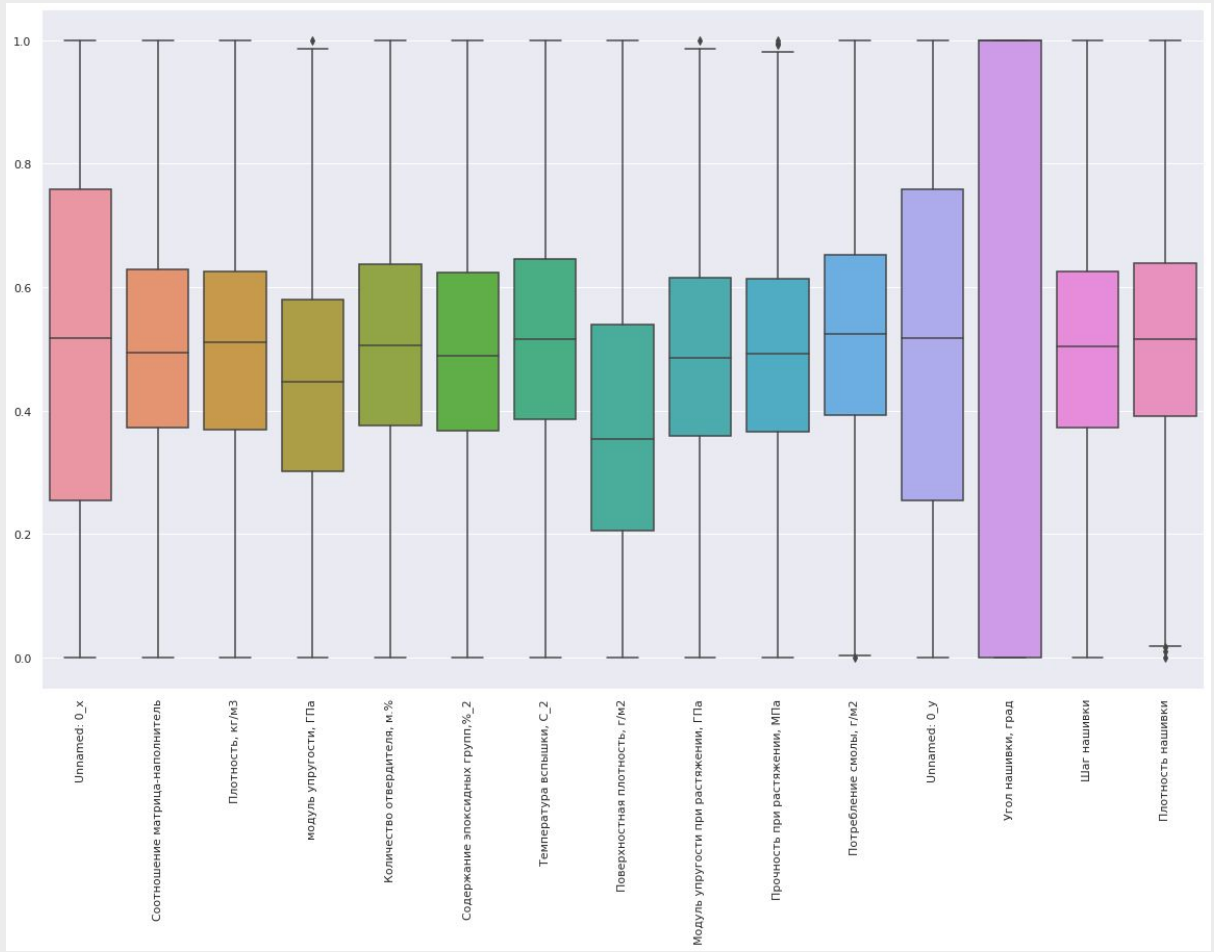
После нормализации



Матрица корреляции данных и диаграмма “Ящики с усами”

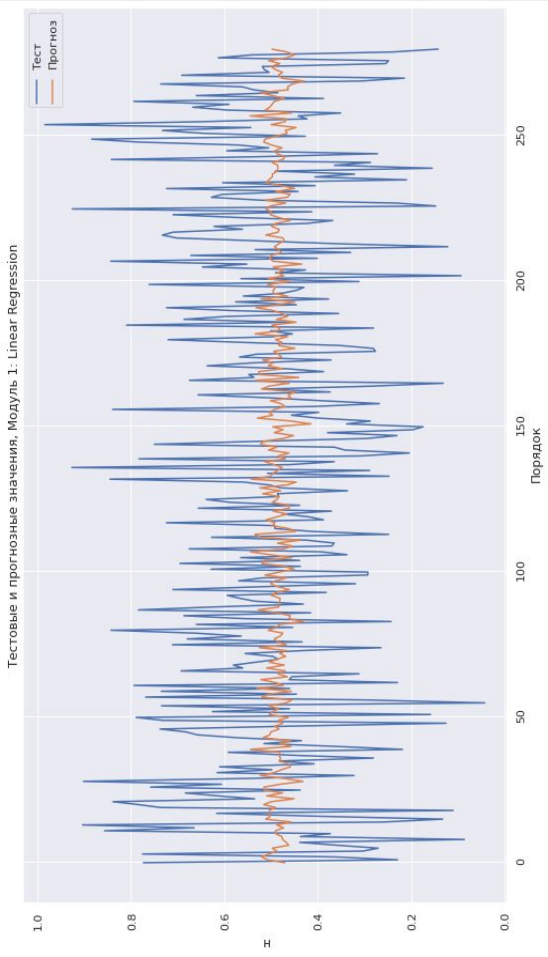
после нормализации MinMaxScaler()



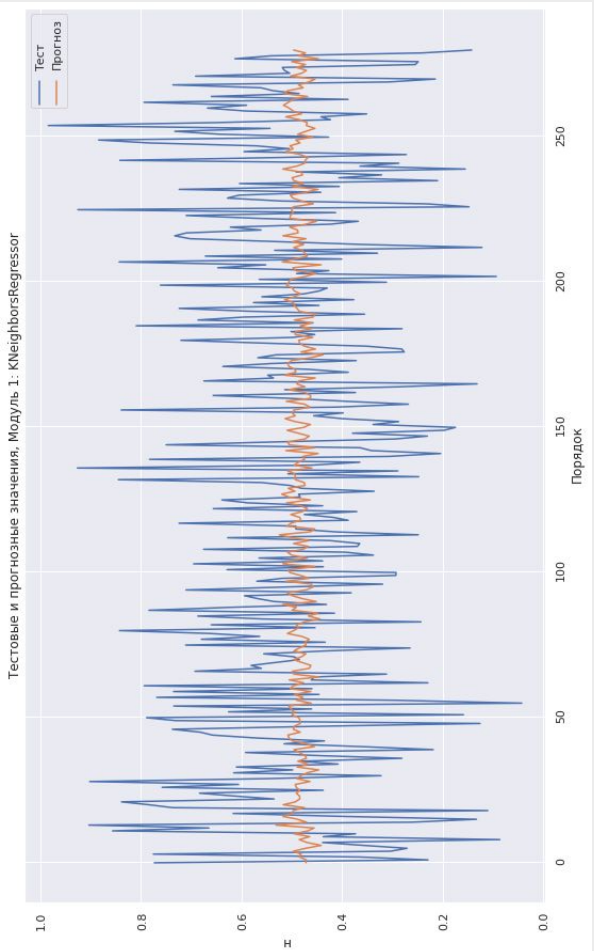


Разработка и обучение моделей. Методы

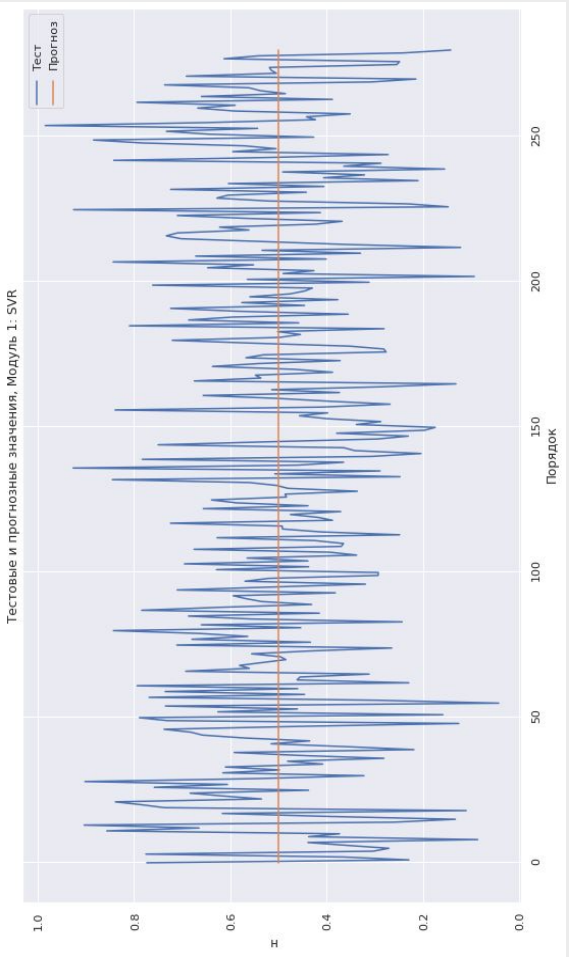
Linear regression



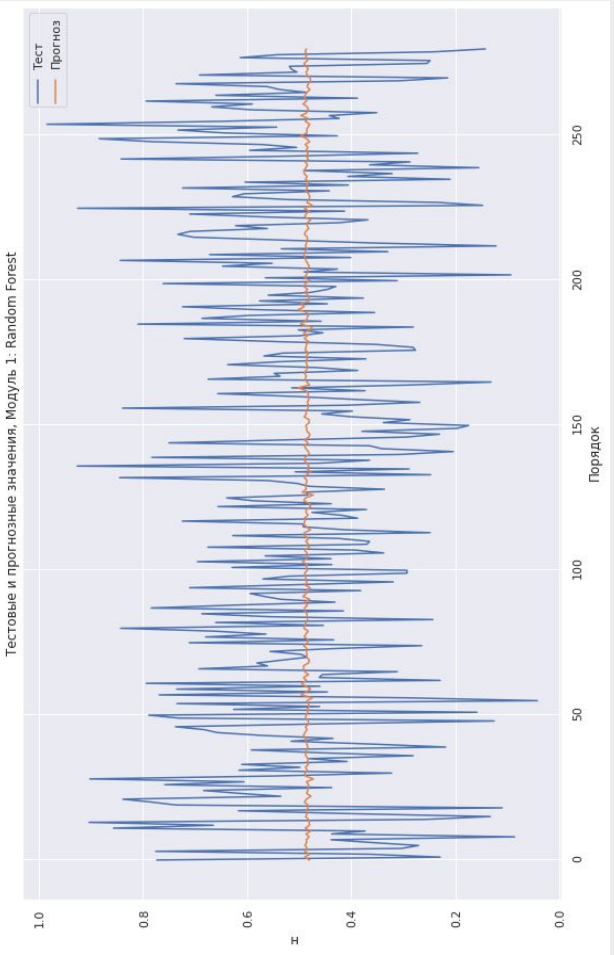
K-neighbors regression



SVR



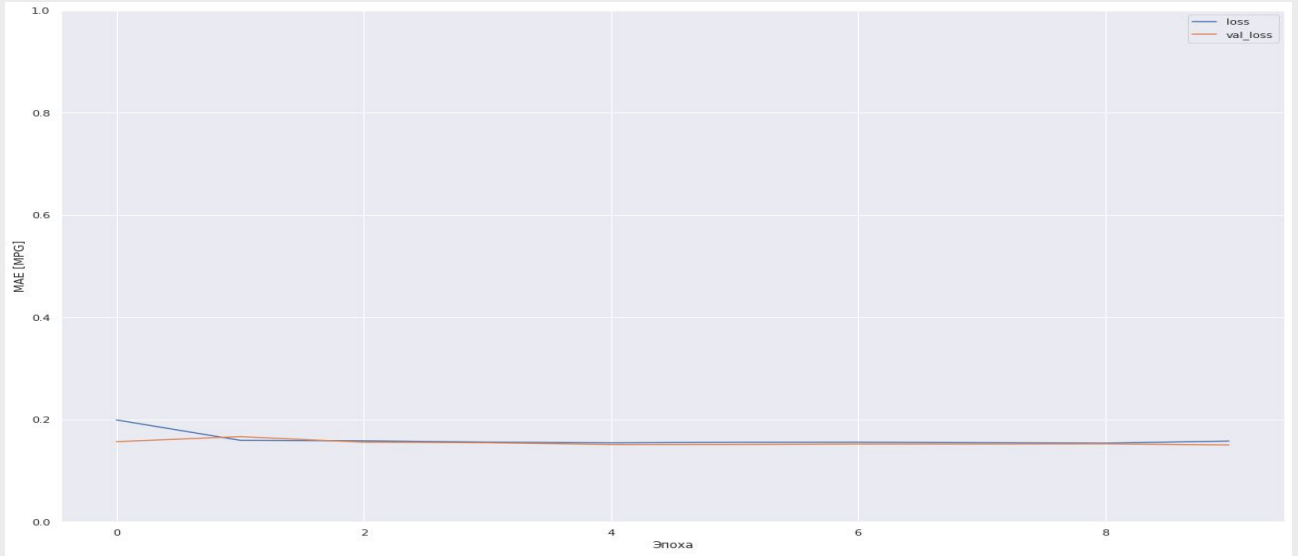
Random Forest



Нейросеть №1 для соотношения «матрица – наполнитель». Визуализация

```
model = Sequential()
model.add(layers.Dense(64, input_dim=X.shape[1], activation='tanh'))
model.add(layers.Dense(64, activation='tanh'))
model.add(layers.Dense(32, activation='sigmoid'))
model.add(layers.Dense(1))
```

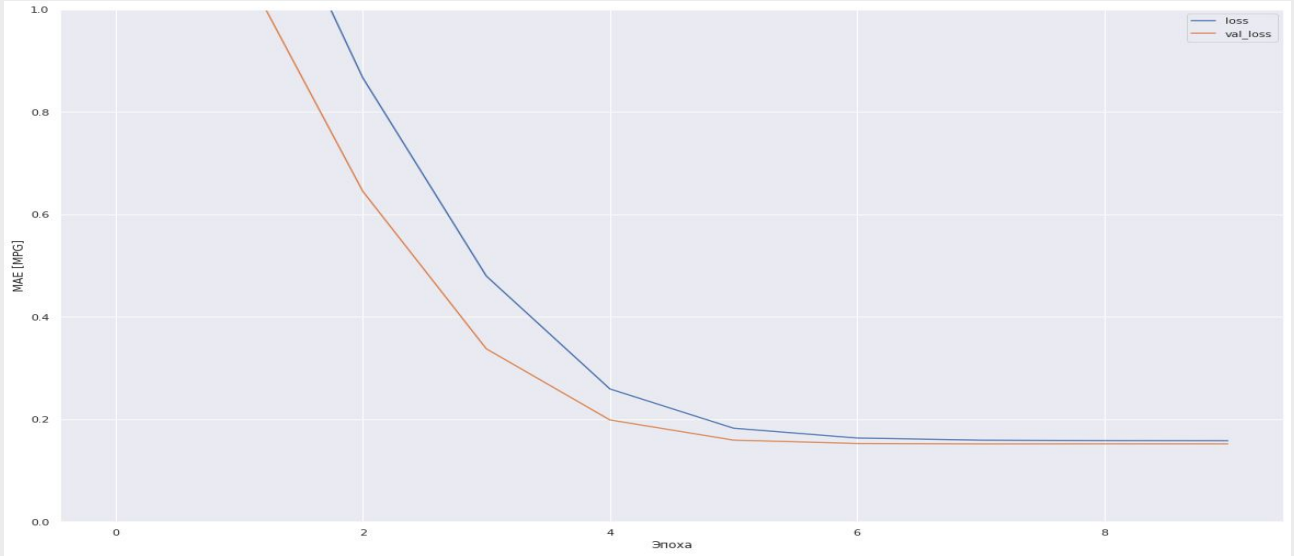
```
model.summary()
dfmodel = model.compile(optimizer='adam', loss='mae', metrics=['mae'])
history = model.fit(
    X_train,
    y_train,
    validation_split=0.2,
    verbose=1, epochs=10)
```



Нейросеть №2 для соотношения «матрица – наполнитель». Визуализация

```
model = Sequential()
model.add(layers.Dense(16, input_dim=X.shape[1], activation='elu'))
model.add(layers.Dense(32, activation='tanh'))
model.add(layers.Dense(8, activation='sigmoid'))
model.add(layers.Dense(1))
```

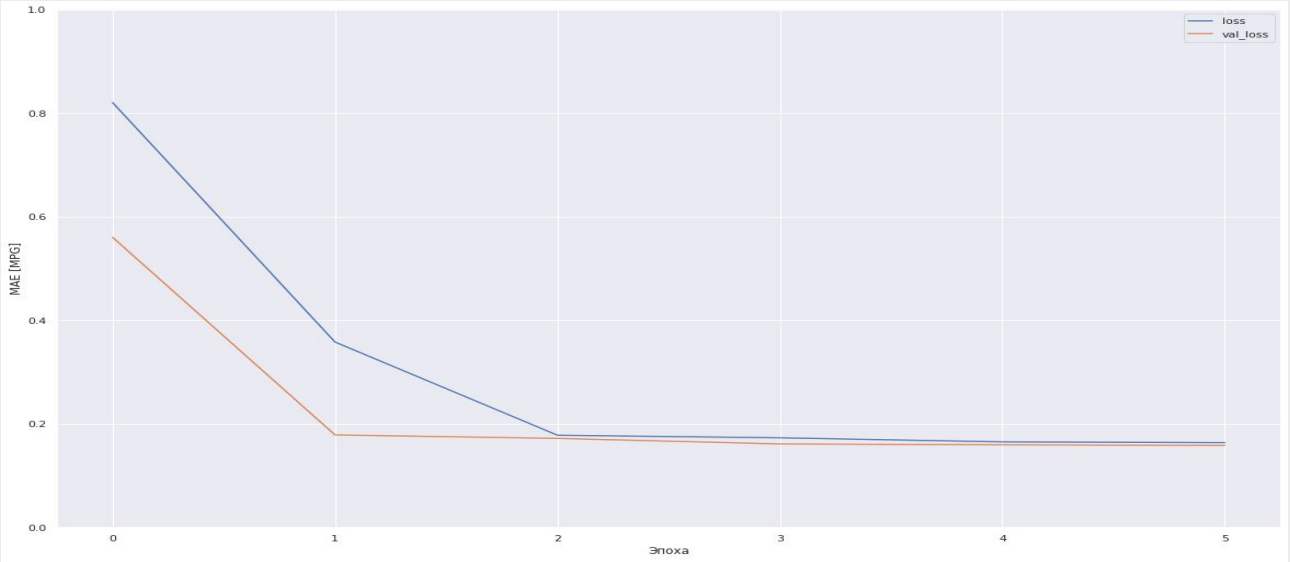
```
model.summary()
dfmodel = model.compile(optimizer='adam', loss='mae', metrics=['mae'])
history = model.fit(
    X_train,
    y_train,
    validation_split=0.2,
    verbose=1, epochs=10)
```



Нейросеть №3 для соотношения «матрица – наполнитель». Визуализация

```
model = Sequential()
model.add(layers.Dense(16, input_dim=X.shape[1], activation='relu'))
model.add(layers.Dense(32, activation='tanh'))
model.add(layers.Dense(8, activation='sigmoid'))
model.add(layers.Dense(1))
```

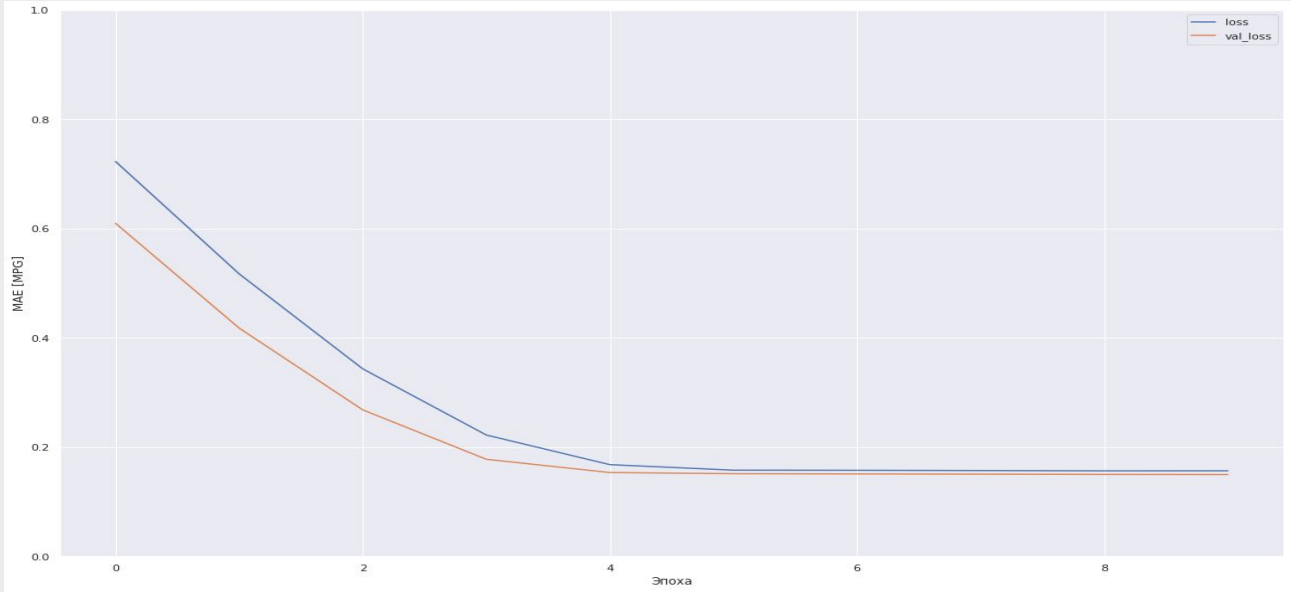
```
model.summary()
dfmodel = model.compile(optimizer='adam', loss='mae', metrics=['mae'])
history = model.fit(
    X_train,
    y_train,
    validation_split=0.2,
    verbose=1, epochs=6)
```



Нейросеть №4 для соотношения «матрица – наполнитель». Визуализация

```
model = Sequential()
model.add(layers.Dense(16, input_dim=X.shape[1], activation='tanh'))
model.add(layers.Dense(8, activation='sigmoid'))
model.add(layers.Dense(1))

model.summary()
dfmodel = model.compile(optimizer='adam', loss='mae', metrics=['mae'])
history = model.fit(
    X_train,
    y_train,
    validation_split=0.2,
    verbose=1, epochs=10)
```



Заключение

Данная исследовательская работа позволяет сделать некоторые основные выводы по теме. Распределение полученных данных в объединённом датасете близко к нормальному, но коэффициенты корреляции между парами признаков стремятся к нулю. Использованные при разработке моделей подходы не позволили получить сколько-нибудь достоверных прогнозов. Применённые модели регрессии не показали высокой эффективности в прогнозировании свойств композитов.

Был сделан вывод, что невозможно определить из свойств материалов соотношение «матрица – наполнитель». Данный факт не указывает на то, что прогнозирование характеристик композитных материалов на основании предоставленного набора данных невозможно, но может указывать на недостатки базы данных, подходов, использованных при прогнозе, необходимости пересмотра инструментов для прогнозирования.

Необходимы дополнительные вводные данные, получение новых результирующих признаков в результате математических преобразований, релевантных доменной области, консультации экспертов предметной области, новые исследования, работа эффективной команды, состоящей из различных специалистов.

В целом прогнозирование конечных свойств/характеристик композитных материалов без изучения материаловедения, погружения в вопрос экспериментального анализа характеристик композитных материалов не демонстрирует сколько-нибудь удовлетворительных результатов. Проработка моделей и построение прогнозов требует внедрения в процесс производных от имеющихся показателей для выявления иного уровня взаимосвязей. Отсюда, также учитывая отсутствие корреляции между признаками, делаем вывод, что текущим набором алгоритмов задача не решается, возможно, решается трудно или не решается совсем.

edu.bmstu.ru

+7 495 182-83-85

edu@bmstu.ru

Москва, Госпитальный переулок , д. 4-6, с.3

ОБРАЗОВАТЕЛЬНЫЙ ЦЕНТР МГТУ им. Н. Э. Баумана