



Выпускная квалификационная работа по курсу Data Science

**Тема: прогнозирование конечных свойств композитных
материалов**

Махров Андрей Сергеевич

Цели ВКР:

- Провести предварительный анализ и предобработку данных
- Применить различные алгоритмы машинного обучения к полученной после предобработки базе данных
- Сравнить результаты работы использованных алгоритмов друг с другом

Задачи ВКР:

- Обучить алгоритм машинного обучения для определения модуля упругости при растяжении, прочности при растяжении
- Написать нейронную сеть для рекомендации соотношения матрица-напонитель
- Разработать приложение

Исходные данные

Файл X_br.xlsx

1023 наблюдения параметров
композитных материалов

	count	mean	std	min	25%	50%	75%	max
Unnamed: 0	1023.0	511.000000	295.458965	0.000000	255.500000	511.000000	766.500000	1022.000000
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, С_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

Файл X_nup.xlsx

1040 наблюдений параметров
композитных материалов

	Unnamed: 0	Угол нашивки, град	Шаг нашивки	Плотность нашивки
count	1040.000000	1040.000000	1040.000000	1040.000000
mean	519.500000	45.000000	6.911385	57.248399
std	300.366443	45.02165	2.555181	12.332438
min	0.000000	0.000000	0.000000	0.000000
25%	259.750000	0.000000	5.102256	49.970740
50%	519.500000	45.000000	6.938000	57.413594
75%	779.250000	90.000000	8.587662	65.107235
max	1039.000000	90.000000	14.440522	103.988901

Объединение данных

Тип INNER по индексу

17 данных не попадают в итоговый датасет

	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, С_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, град	1023.0	44.252199	45.015793	0.000000	0.000000	0.000000	90.000000	90.000000
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

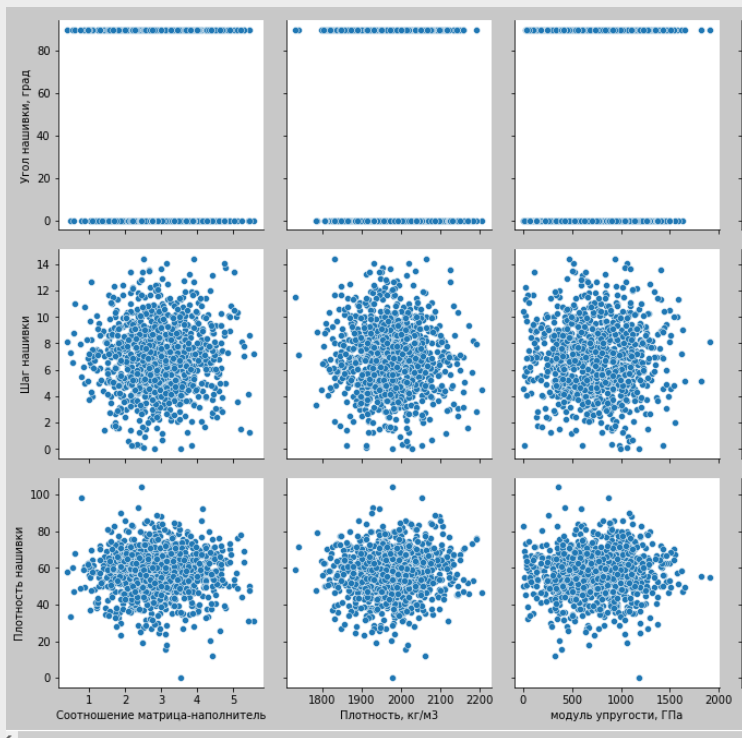
← Задача для
нейронной сети

← Задача для
методов ML

Разведочный анализ данных



Попарные графики (не все данные)*



Визуально взаимосвязи отсутствуют

Пропусков в данных нет, все значения числовые

```
df.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 1023 entries, 0 to 1022
Data columns (total 15 columns):
 #   Column                                                                 Non-Null Count  Dtype  
---  -
 0   Unnamed: 0_x                                                         1023 non-null  float64
 1   Соотношение матрица-наполнитель                                       1023 non-null  float64
 2   Плотность, кг/м3                                                       1023 non-null  float64
 3   модуль упругости, ГПа                                                  1023 non-null  float64
 4   Количество отвердителя, м.%                                           1023 non-null  float64
 5   Содержание эпоксидных групп,%_2                                       1023 non-null  float64
 6   Температура вспышки, C_2                                              1023 non-null  float64
 7   Поверхностная плотность, г/м2                                         1023 non-null  float64
 8   Модуль упругости при растяжении, ГПа                                  1023 non-null  float64
 9   Прочность при растяжении, МПа                                          1023 non-null  float64
10   Потребление смолы, г/м2                                               1023 non-null  float64
11   Unnamed: 0_y                                                         1023 non-null  float64
12   Угол нашивки, град                                                    1023 non-null  float64
13   Шаг нашивки                                                           1023 non-null  float64
14   Плотность нашивки                                                     1023 non-null  float64
dtypes: float64(15)
memory usage: 127.9 KB
```

Уникальные значения

```
df.nunique()

Unnamed: 0_x                1023
Соотношение матрица-наполнитель  1014
Плотность, кг/м3            1013
модуль упругости, ГПа       1020
Количество отвердителя, м.%  1005
Содержание эпоксидных групп,%_2  1004
Температура вспышки, C_2    1003
Поверхностная плотность, г/м2  1004
Модуль упругости при растяжении, ГПа  1004
Прочность при растяжении, МПа  1004
Потребление смолы, г/м2     1003
Unnamed: 0_y                1023
Угол нашивки, град          2
Шаг нашивки                 989
Плотность нашивки           988
dtype: int64
```



Угол нашивки 2 значения
LabelEncoder не применяется

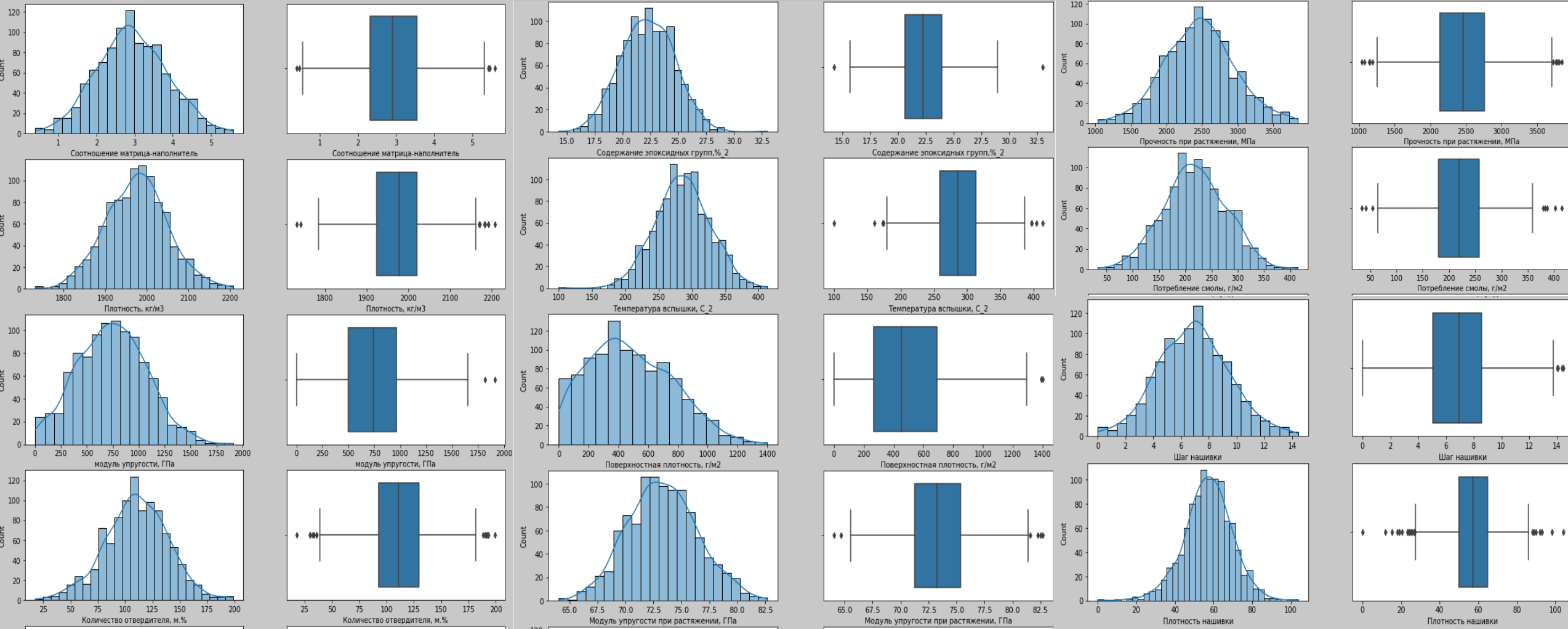
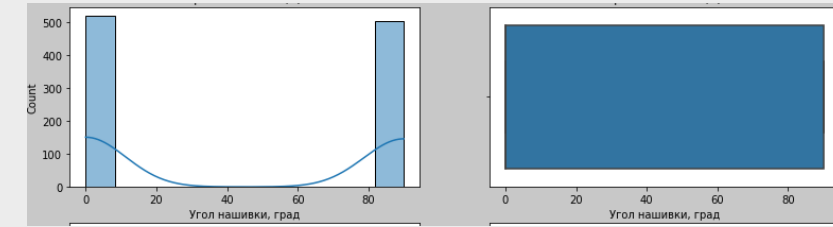
* График со всеми параметрами в пояснительной записке и тетради jupyter (результаты похожи)

Графики распределения данных

Все данные имеют нормальное распределение, кроме параметра «угол нашивки»

Выбросы незначительные

Выбросы удаляются методом 3-х сигм

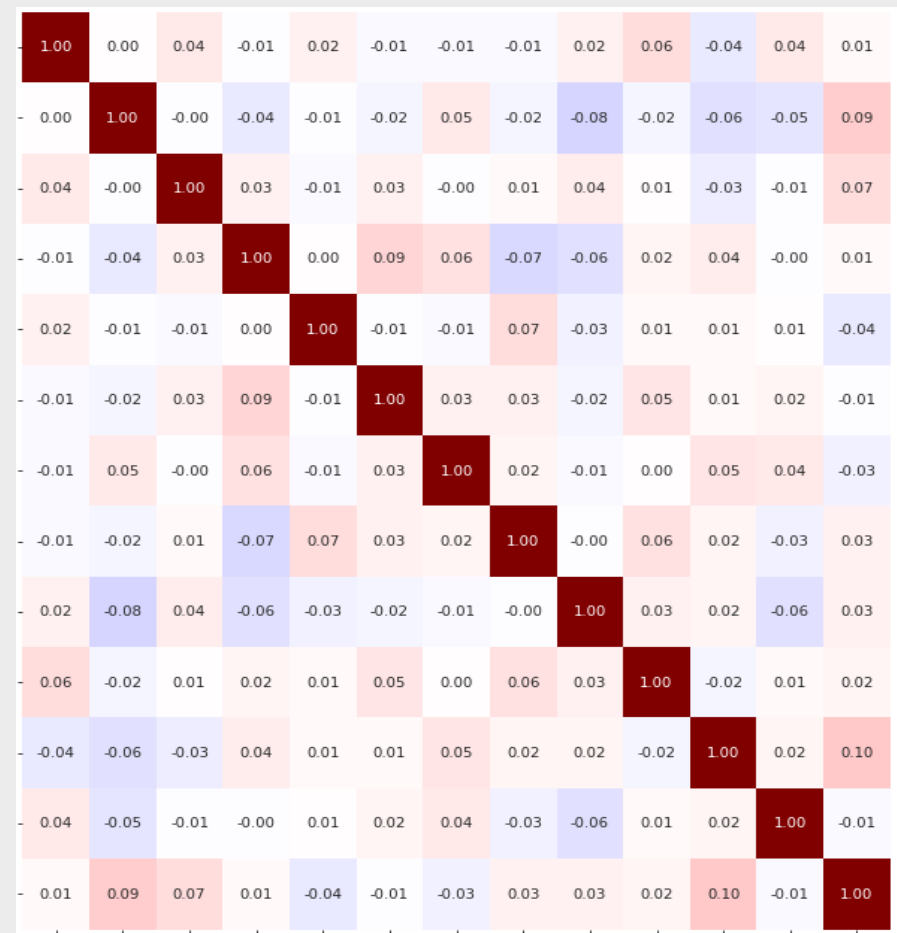


Корреляционные матрицы

Исходные данные



После удаления выбросов



Взаимосвязь между данными отсутствует

Алгоритмы ML для определения модуля упругости при растяжении, прочности при растяжении

Предпосылки:

Необходимо найти прогнозное значение одной переменной по значениям другой – применяются модели регрессии

Все значения имеют разный масштаб – применяется нормализация данных

Нет четкой взаимосвязи между данными – применяется несколько моделей регрессии и выбирается лучшая

Некоторые модели регрессии имеют различные параметры - оптимизация через GridSearchCV

Применяемые методы регрессии в ВКР



линейная регрессия

метод k-ближайших соседей

регрессия дерева решений

метод опорных векторов

Методы оценки качества моделей



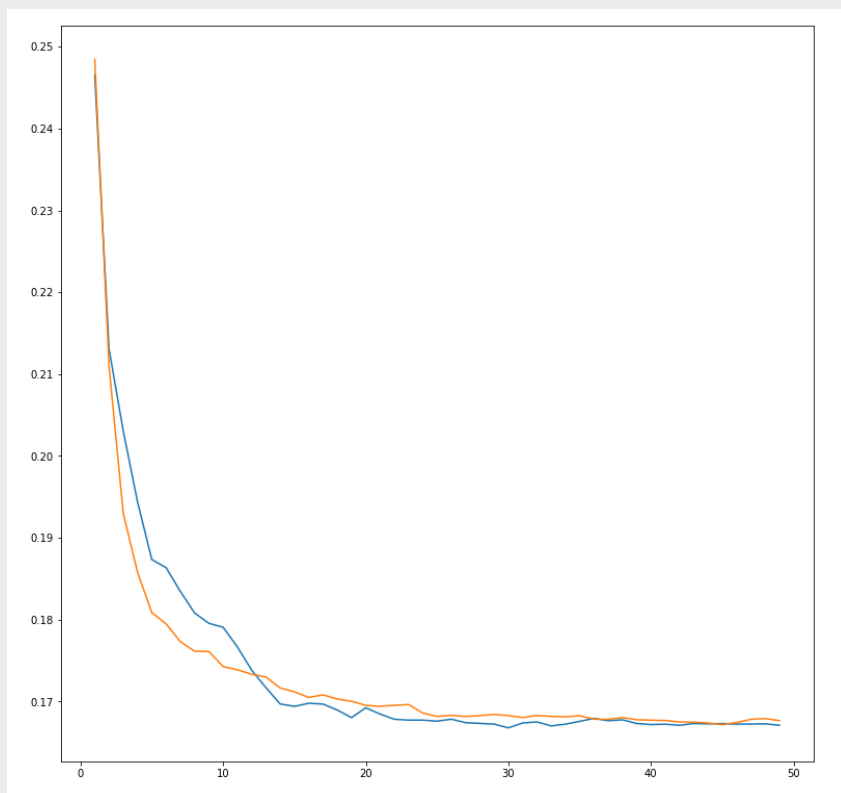
Средняя абсолютная ошибка (MAE)

Средняя квадратичная ошибка (MSE)

Корень из средней квадратичной ошибки (RMSE)

Подбор гиперпараметров моделей

Метод	Прогнозируемое значение	Лучшие параметры
K-nn	Модуль упругости	{'n_neighbors': 150}
K-nn	Прочность	{'n_neighbors': 150}
Tree	Модуль упругости	{'max_depth': 1, 'min_samples_split': 5, 'splitter': 'random'}
Tree	Прочность	{'max_depth': 2, 'min_samples_split': 2, 'splitter': 'random'}
SVR	Модуль упругости	{'degree': 5, 'gamma': 'auto', 'kernel': 'poly'}
SVR	Прочность	{'degree': 3, 'gamma': 'auto', 'kernel': 'poly'}



Градиентный спуск для k-nn

Итоговые данные для оценки качества моделей

(нормализованные данные)

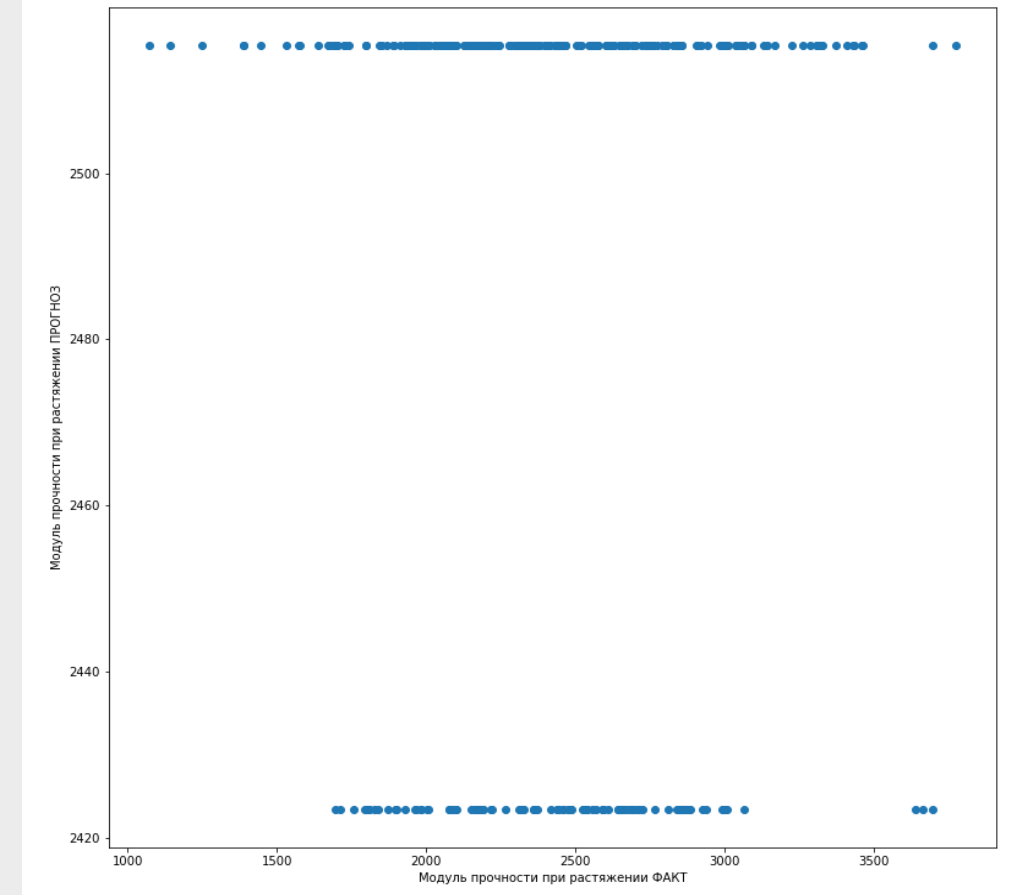
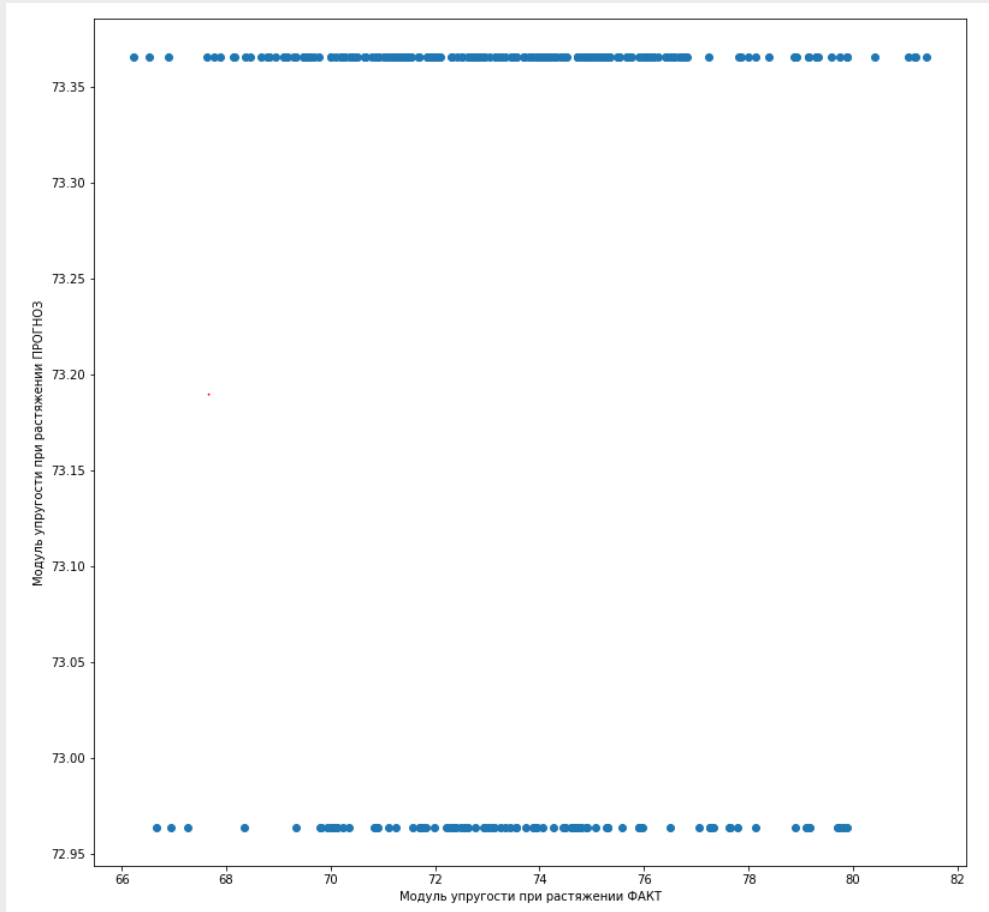
Метод опорных векторов и дерево решений дали одинаковые данные

Параметры качества моделей для данных методов наилучшие

Параметр	Линейная регрессия	К - ближайшие соседи	Дерево решений	Метод опорных векторов
MAE для упругости	0,133	0,136	0,134	0,134
MSE для упругости	0,028	0,029	0,028	0,028
RMSE для упругости	0,166	0,169	0,167	0,167
MAE для прочности	0,136	0,134	0,132	0,132
MSE для прочности	0,029	0,029	0,028	0,028
RMSE для прочности	0,171	0,170	0,168	0,168

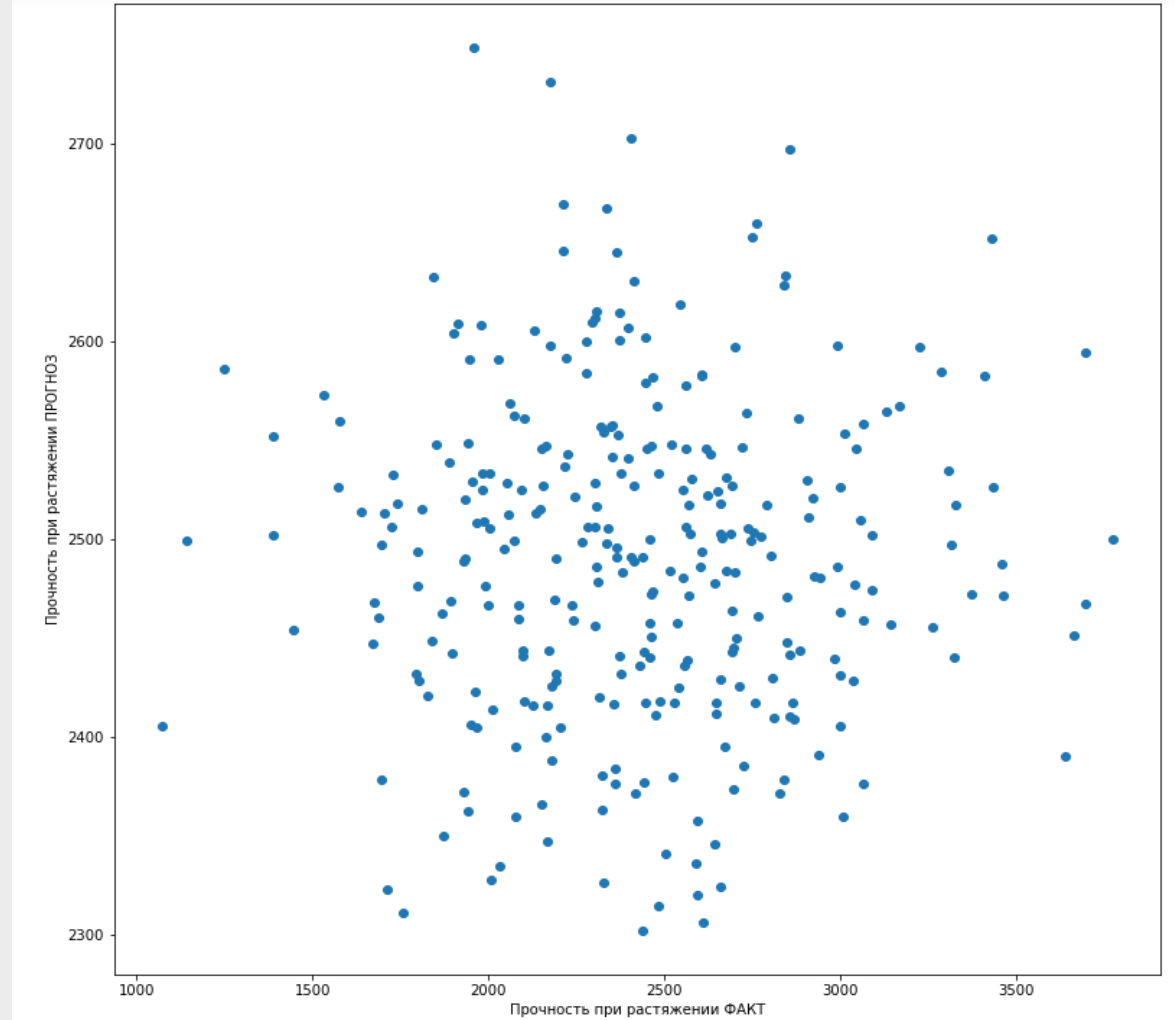
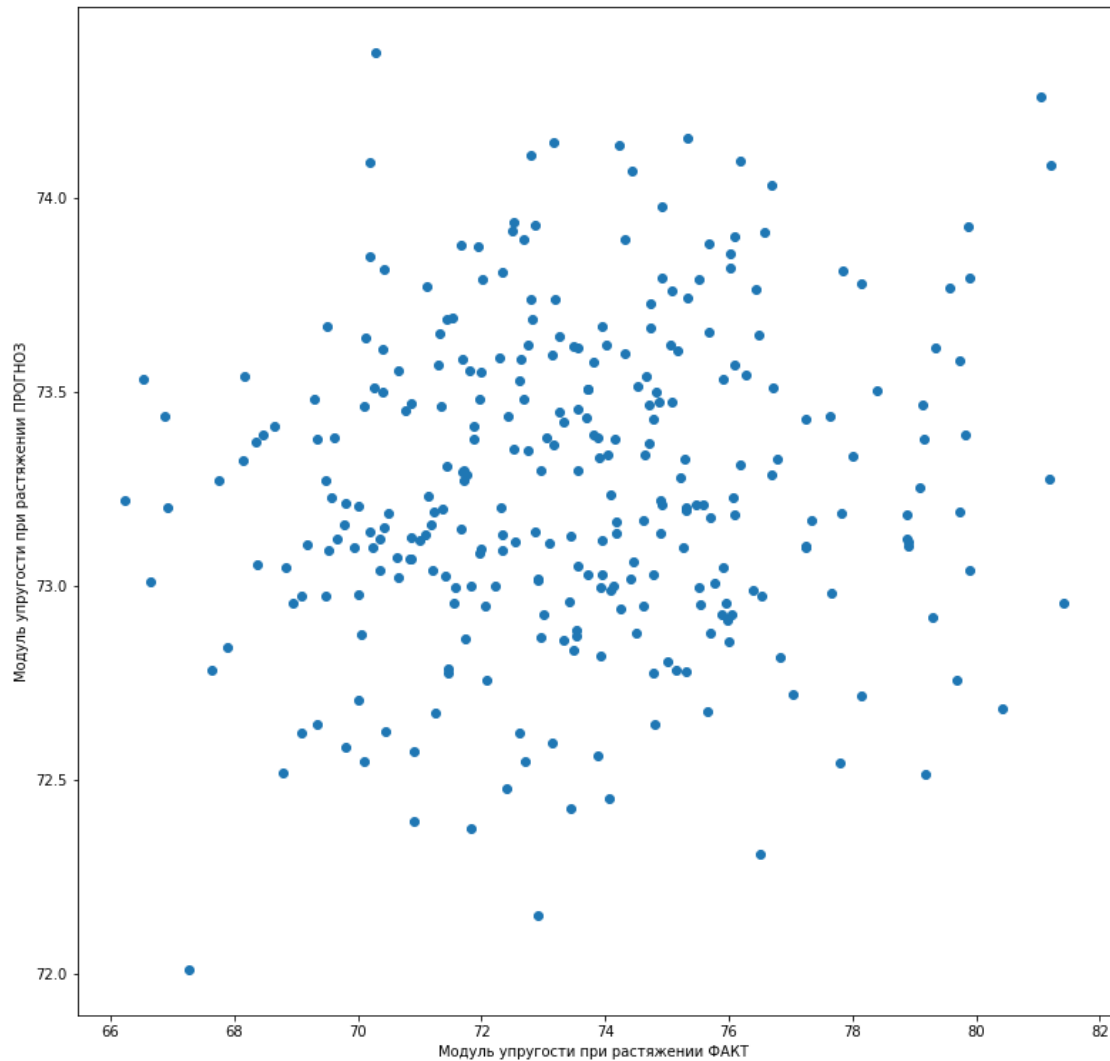
Фактические и спрогнозированные данные – метод опорных векторов (данные преобразованы в исходный размер)

Модель спрогнозировала только 2 значения



Фактические и спрогнозированные данные – линейная регрессия (данные преобразованы в исходный размер)

Линейная регрессия так-же плохо предсказывает данные



Нейронная сеть для рекомендации соотношения матрица-наполнитель

Количество слоев – 2 , 4

Данные предварительно нормализованы

Выходной слой - 1

Активационная функция «relu»

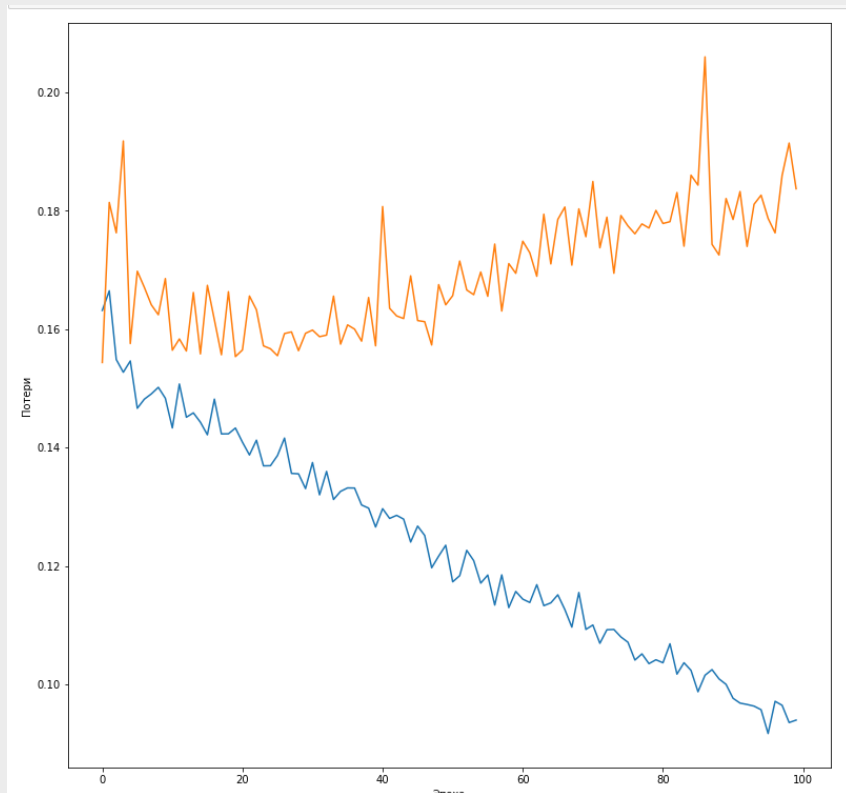
Оптимизатор «adam»

Эпох - 100

Оценка качества моделей

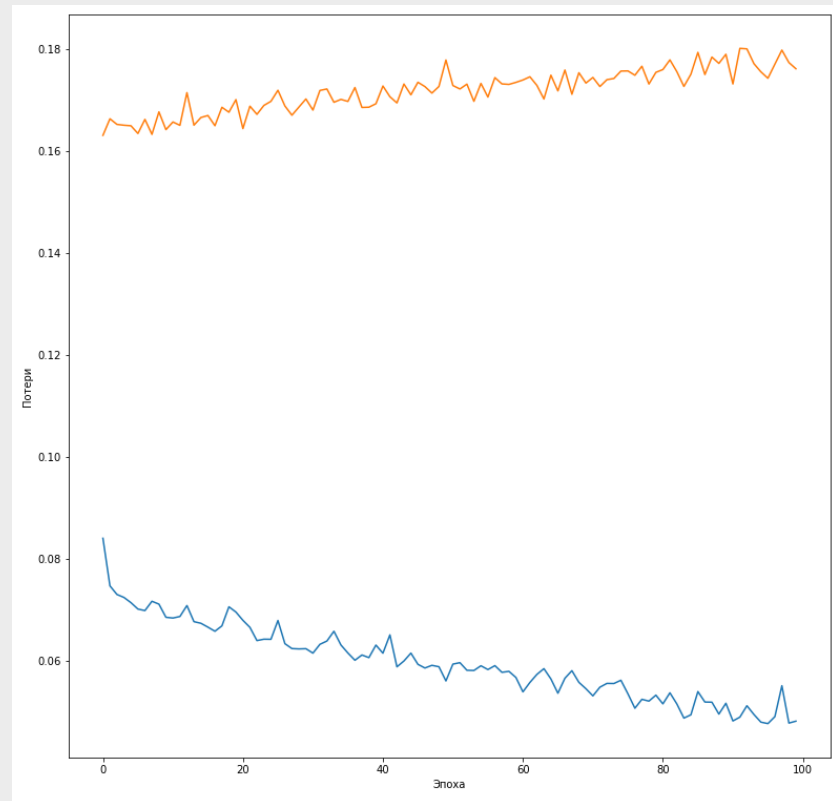
Параметр	2 слоя	4 слоя
MAE	0,115327	0,070819
MSE	0,020508	0,013829
RMSE	0,143205	0,117599

График обучения модели 2 слоя



mae
val_mae

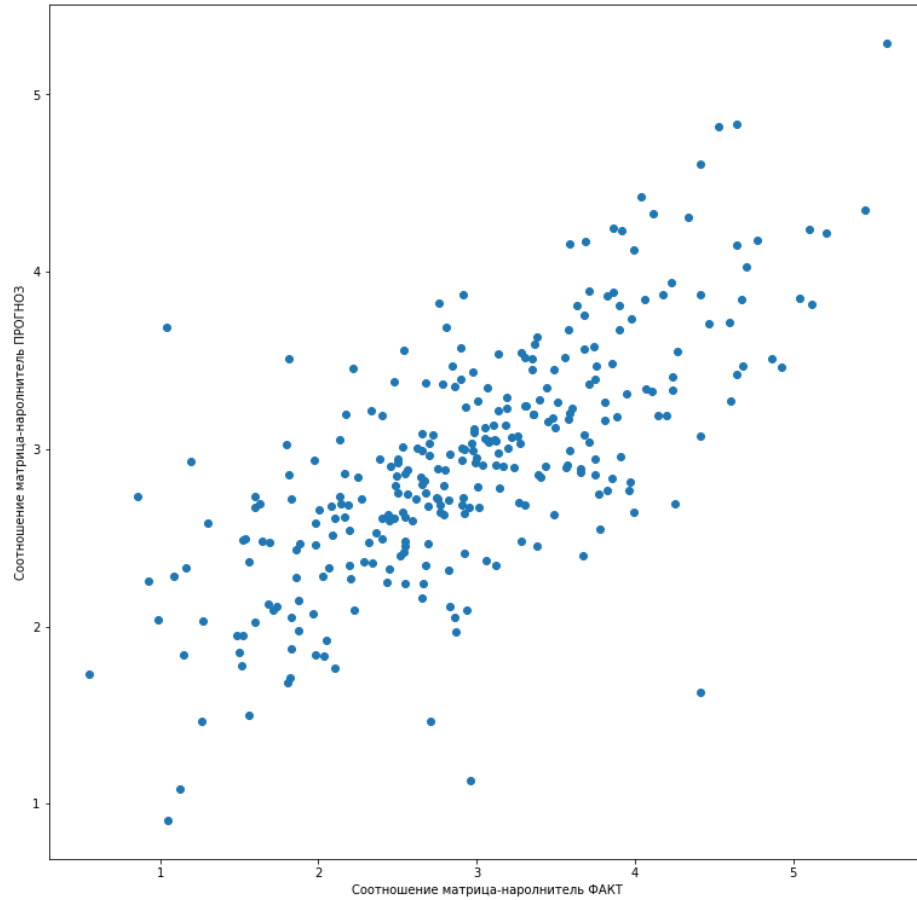
График обучения модели 4 слоя



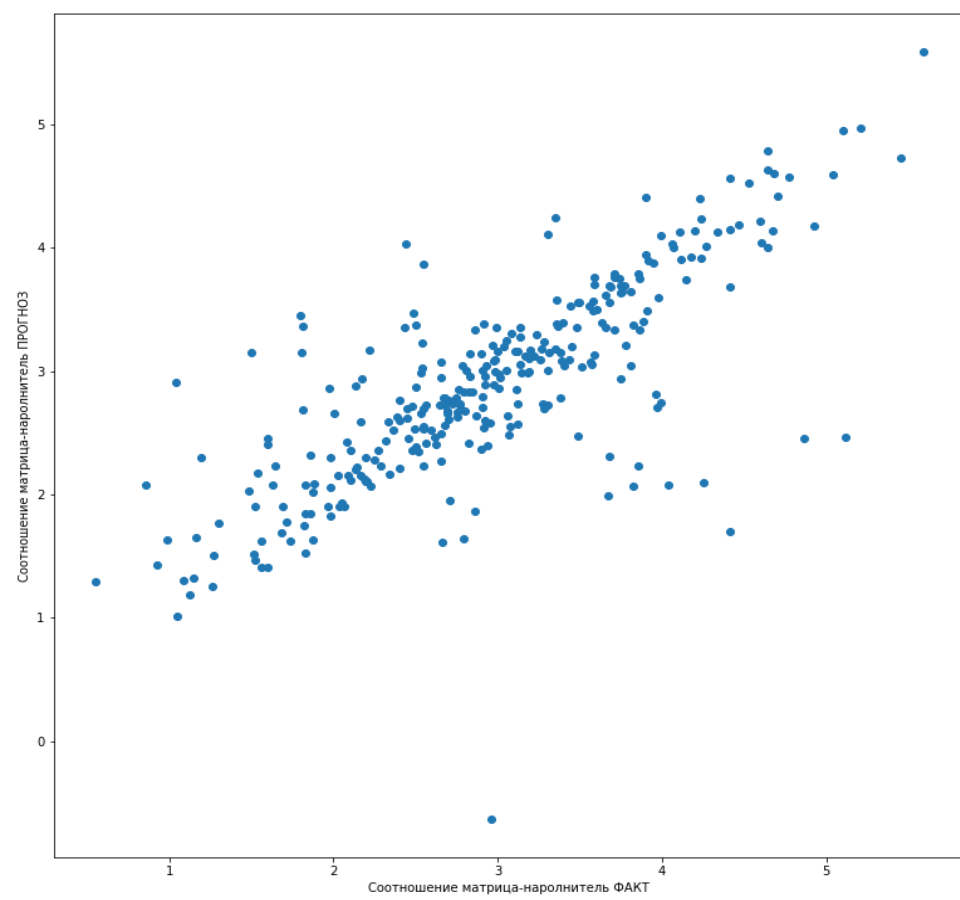
Фактические и спрогнозированные данные – нейронная сеть (данные преобразованы в исходный размер)

Фактические и прогнозные значения имеют некоторую взаимосвязь

Модель 2 слоя



Модель 4 слоя



Интерфейс приложения прогноз модуля упругости при растяжении

При расчете использовалась модель
линейной регрессии

Созданный репозиторий
<https://github.com/AndreyM123/vkr>

AndreyM123 / vkr Public

Unpin Unwatch 1 Fork 0 Star 0

Code Issues Pull requests Actions Projects Wiki Security Insights Settings

main 1 branch 0 tags

Go to file Add file Code

About

No description, website, or topics provided.

Readme 0 stars 1 watching 0 forks

Releases

No releases published
[Create a new release](#)

Packages

No packages published
[Publish your first package](#)

Languages

Jupyter Notebook 99.9% Other 0.1%

README.md

Приложение позволяет рассчитать прогноз модуля упругости при растяжении. Нужно ввести в соответствующие окна 12 параметров:

1. соотношение матрица-наполнитель
2. плотность, кг/м3;
3. модуль упругости, ГПа;

Введите значение

Соотношение матрица-наполнитель 3.444022

Введите значение

Плотность, кг/м3 1972.372865

Введите значение

модуль упругости, ГПа 807.000000

Введите значение

Количество отвердителя, м.% 111.860000

Введите значение

Содержание эпоксидных групп, %_2 23.957502

Введите значение

С Температура вспышки, C_2 300.000000

Введите значение

Поверхностная плотность, г/м2 350.660830

Введите значение

Потребление смолы, г/м2 641.468152

Введите значение

Угол нашивки, град 90.0

Введите значение

Шаг нашивки 6.313201

Введите значение

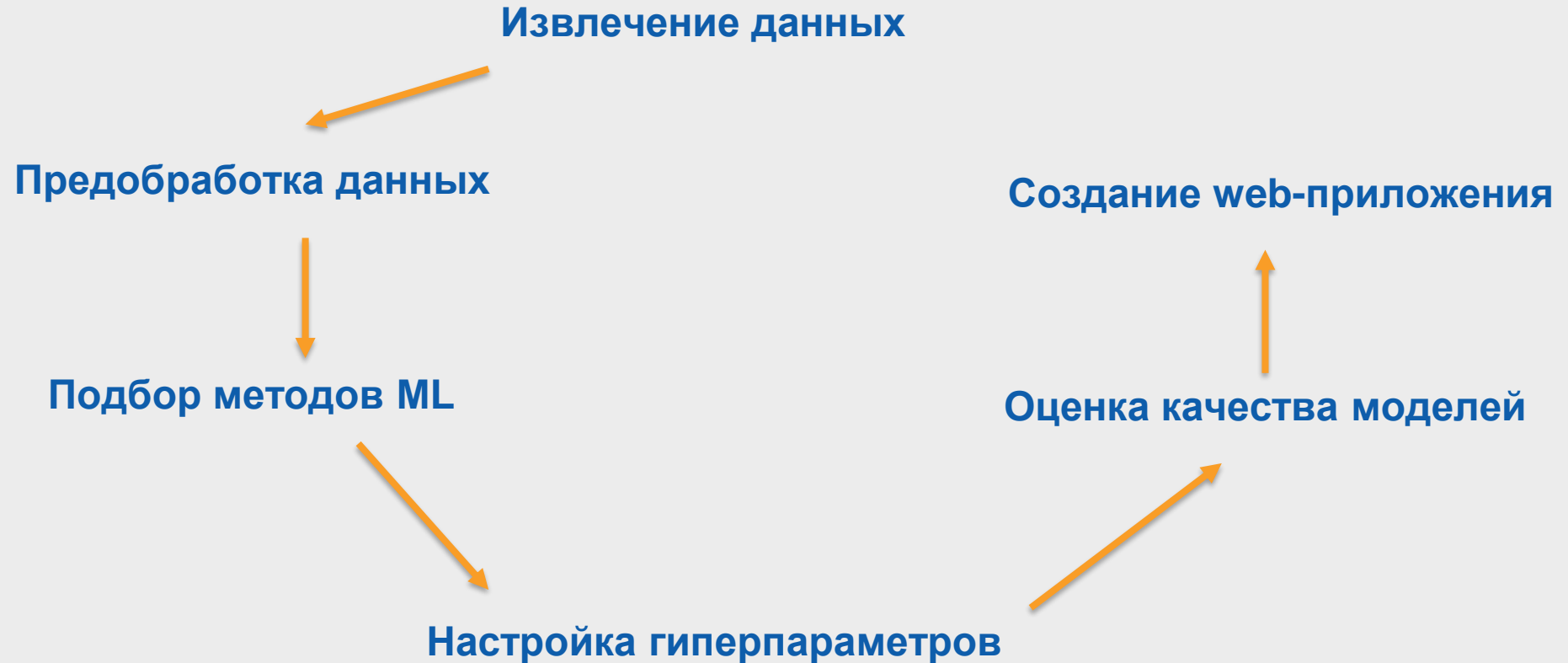
Плотность нашивки 77.434468

Отправить

Значение модуля упругости при
растяжении

[67.19748858]

Pipeline ML при выполнении ВКР





Спасибо за внимание !

Махров Андрей Сергеевич



edu.bmstu.ru

+7 495 182-83-85

edu@bmstu.ru

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