ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА по курсу «Data Science»

ТЕМА: ПРОГНОЗИРОВАНИЕ КОНЕЧНЫХ СВОЙСТВ НОВЫХ МАТЕРИАЛОВ (КОМПОЗИЦИОННЫХ МАТЕРИАЛОВ)

Слушатель: Терещук В. О.

Актуальность темы

- Теоретический анализ полимерных композиционных материалов путём построения моделей на основе методов вычислительной механики и прогнозирование их эффективных характеристик с завершающей оценкой их надёжности является актуальным.
- ► Созданные прогнозные модели помогут сократить количество проводимых испытаний, а также пополнить базу данных материалов возможными новыми характеристиками материалов, и цифровыми двойниками новых композитов.

характеристики анализируемого датасета

▶ В задании представлены два файла «X_bp.xlsx» и «X_nup.xlsx» в формате таблицы Microsoft Office Excel. Для формирование единого массива данных, произведено сведения обоих файлов в один.

• Объём и характеристики датасета: в сведённом датасете 1023 записи по каждому показателю,

пропуски отсутствуют (нет пустых значений),

```
df all.info()
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1023 entries, 0 to 1022
Data columns (total 13 columns):
     Column
                                         Non-Null Count
                                                        Dtype
                                         1023 non-null float64
    Соотношение матрица-наполнитель
    Плотность, кг/м3
                                         1023 non-null float64
    модуль упругости, ГПа
                                         1023 non-null float64
    Количество отвердителя, м.%
                                         1023 non-null float64
                                         1023 non-null float64
    Содержание эпоксидных групп,% 2
    Температура вспышки, С 2
                                         1023 non-null float64
    Поверхностная плотность, г/м2
                                         1023 non-null float64
    Модуль упругости при растяжении, ГПа 1023 non-null float64
    Прочность при растяжении, МПа
                                         1023 non-null float64
    Потребление смолы, г/м2
                                        1023 non-null float64
                                         1023 non-null int64
    Угол нашивки, град
                                                        float64
     Шаг нашивки
                                         1023 non-null
                                         1023 non-null
                                                        float64
    Плотность нашивки
dtypes: float64(12), int64(1)
memory usage: 111.9 KB
```

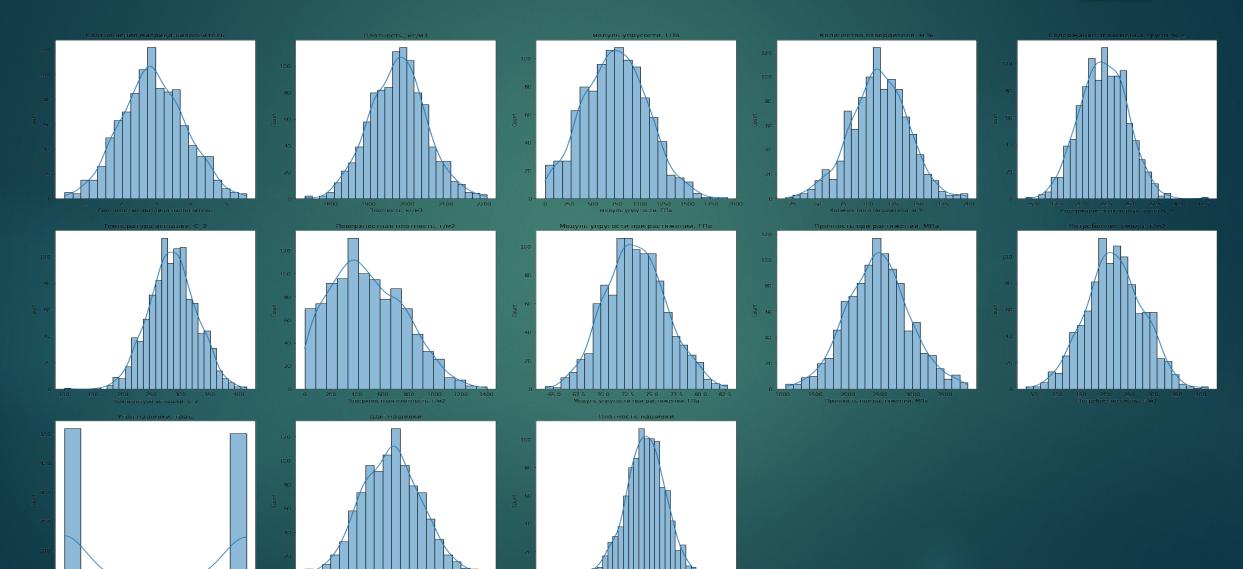
Матрица, анализирующая мощность выборки

1	_11	_	escr:		-/	_
αT	all	- 0	escr	1 D	ρ	•
-	~ ~ ~			_	_ \	,

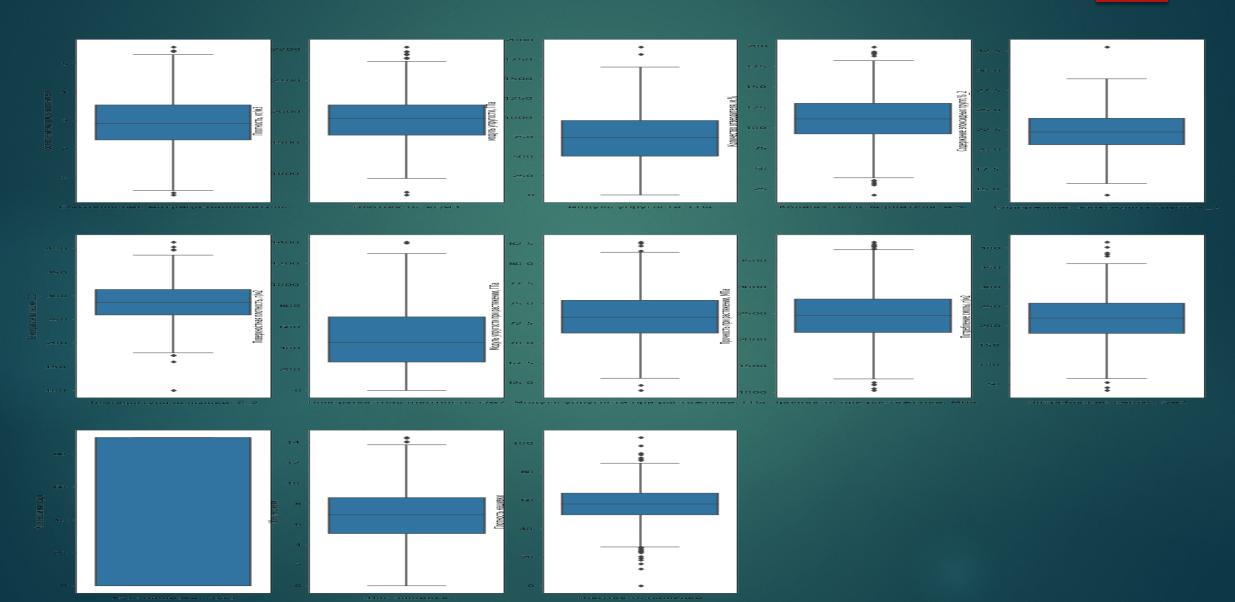
	Соотношение матрица- наполнитель	Плотность, кг/м3	модуль упругости, ГПа	Количество отвердителя, м.%	Содержание эпоксидных групп,%_2	Температура вспышки, С_2	Поверхностная плотность, г/ м2	Модуль упругости при растяжении, ГПа	Прочность при растяжении, МПа	Потребление смолы, г/м2	Уг нашив⊧ гр:
count	936.000000	936.000000	936.000000	936.000000	936.000000	936.000000	936.000000	936.000000	936.000000	936.000000	936.0000
mean	2.925683	1974.040023	738.247627	110.916216	22.209030	286.040414	482.993901	73.305127	2467.488822	217.613374	46.0576
std	0.893712	70.808120	328.708665	27.037891	2.394871	39.400677	280.190377	3.037381	463.838911	57.827255	45.0116
min	0.547391	1784.482245	2.436909	38.668500	15.695894	179.374391	0.603740	65.553336	1250.392802	63.685698	0.0000
25%	2.321931	1923.443748	498.577158	92.523816	20.571516	259.184486	266.004099	71.248823	2146.936034	179.489091	0.0000
50%	2.904731	1977.258043	738.736842	111.113175	22.184713	286.024118	457.732246	73.259230	2457.959767	218.388715	90.0000
75%	3.546650	2020.158764	958.418993	130.001450	23.961818	312.991425	695.900862	75.310788	2755.169485	256.396777	90.0000
max	5.314144	2161.565216	1649.415706	181.828448	28.955094	386.067992	1291.340115	81.417126	3705.672523	359.052220	90.0000
4											+

• представлены основные характеристики параметров датасета: количество элементов, средние значения параметров, минимальные и максимальные значения, а также медианные значения (наименование строки 50%).

Гистограммы распределения



Ящик с усами



Матрица корреляция

Соотношение матрица-наполнитель	1	0.0038	0.032	-0.0064	0.02	-0.0048	-0.0063	-0.0084	0.024	0.073	-0.031	0.036	-0.0047
Плотность, кг/м3	0.0038	1	-0.0096	-0.036	-0.0083	-0.021	0.045	-0.018	-0.07	-0.016	-0.068	-0.061	0.08
мадуль упругости, ГПа	0.032	-0.0096	1	0.024	-0.0068	0.031	-0.0053	0.023	0.042	0.0018	-0.025	-0.0099	0.056
Количество отвердителя, м.%	-0.0064	-0.036	0.024	1	-0.00068	0.095	0.055	-0.066	-0.075	0.0074	0.039	0.015	0.017
Содержание эпоксидных групп,%_2	0.02	-0.0083	-0.0068	-0.00068	1	-0.0098	-0.013	0.057	-0.024	0.015	0.0081	0.003	-0.039
Температура вспышки, С_2	-0.0048	-0.021	0.031	0.095	-0.0098	1	0.02	0.028	-0.032	0.06	0.021	0.026	0.011
Поверхностная плотность, г/м2	-0.0063	0.045	-0.0053	0.055	-0.013	0.02	1	0.037	-0.0032	0.016	0.052	0.038	-0.05
Модуль упругости при растяжении, ГПа	-0.0084	-0.018	0.023	-0.066	0.057	0.028	0.037	1	-0.009	0.051	0.023	-0.029	0.0065
Прочность при растяжении, МПа	0.024	-0.07	0.042	-0.075	-0.024	-0.032	-0.0032	-0.009	1	0.029	0.023	-0.06	0.02
(Івтребление смолы, г/м2 -	0.073	-0.016	0.0018	0.0074	0.015	0.06	0.016	0.051	0.029	1	-0.015	0.013	0.012
Угол нашивки, град	-0.031	-0.068	-0.025	0.039	0.0081	0.021	0.052	0.023	0.023	-0.015	1	0.024	0.11
Шаг нашивки -	0.036	-0.061	-0.0099	0.015	0.003	0.026	0.038	-0.029	-0.06	0.013	0.024	1	0.0035
Плотность нашивки	-0.0047	0.08	0.056	0.017	-0.039	0.011	-0.05	0.0065	0.02	0.012	0.11	0.0035	1
	оотнашение матрица-наполнитель –	Juothacte, Kr(#3 =	мадуль упругости, ГПа –	Количества отвердителя, и % =	одержание эпоксидных групп,%_2 -	Температура вспышки, С_2 =	Пверхностная плотность, г/м2 –	ль упругасти при растяжении, ГПа –	Прочнасть при растяжении, МПа –	Потребление смолы, тум2 –	Угал нашизки, град –	Шаг зашивки =	Плотность нашивки –

Используемые библиотеки

- **▶** NumPy
- Pandas
- Matplotlib
- **Seaborn**
- Sklearn
- **▶** Keras

МОДЕЛИ МАШИННОГО ОБУЧЕНИЯ

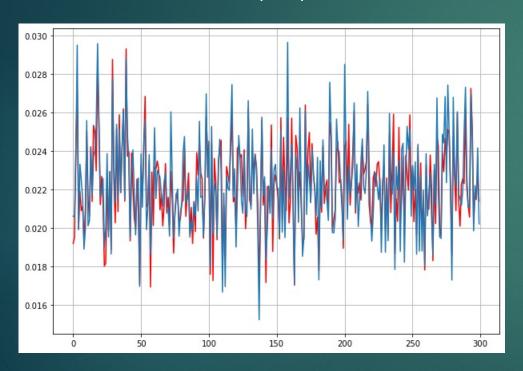
- >sklearn.linear_model.LinearRegression
- >sklearn.linear_model.SGDRegressor
- >sklearn.neighbors.KNeighborsRegressor
- >sklearn.ensemble.RandomForestRegressor
- >sklearn.tree.DecisionTreeRegressor
- >sklearn.ensemble.HistGradientBoostingRegressor

ЭТАПЫ ОБРАБОТКИ ДАННЫХ

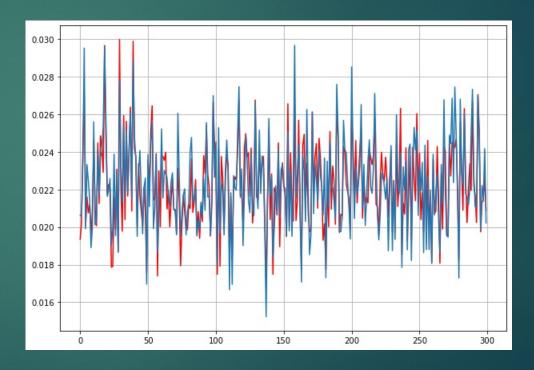
- ► На первом этапе были проанализированы все признаки для определения их максимальных, минимальных, средних и медианных значений, а также была проведена визуализация данных.
- ▶ После этого было проведено исключение выбросов данных, то есть точек данных, которые лежали вдали от обычного распределения данных. Диаграмма ящиков с усами является отличным способом визуализации таких значений.
- На заключительном этапе была проведена нормализация данных.
- ▶ После нормализации данных был также проведен анализ взаимосвязи переменных друг с другом.
- ▶ По результатам предобработки данных можно сделать следующий вывод. Между параметрами модели не наблюдается корреляций и очевидных связей. Число выбросов оказалось незначительным.
- ▶ Для рекомендации соотношения «матрица-наполнитель» была разработана простая модель глубокого обучения с помощью Keras.

Предсказание для признака «Модуль упругости при растяжении»

Линейная регрессия

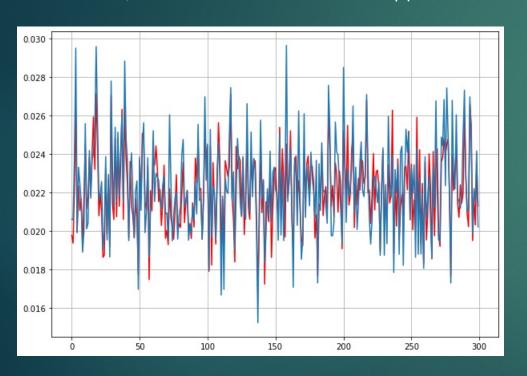


Случайный лес

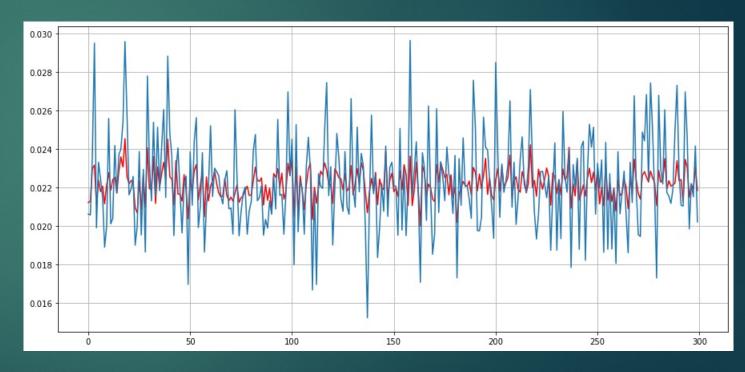


Предсказание для признака «Модуль упругости при растяжении»

К-ближайших соседей

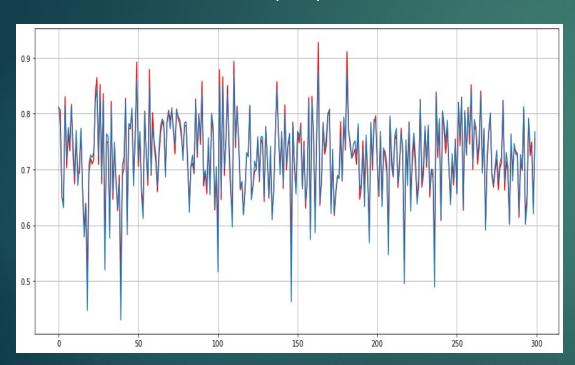


Стохастический градиентный спуск

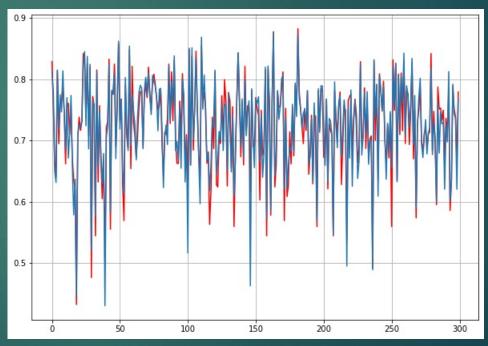


Предсказание для признака «Прочность при растяжении»

Линейная регрессия

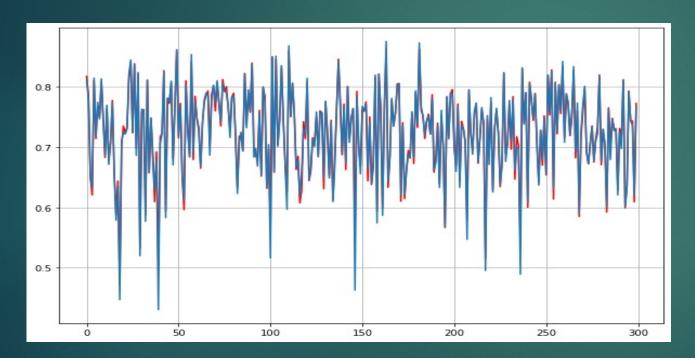


Деревья решений

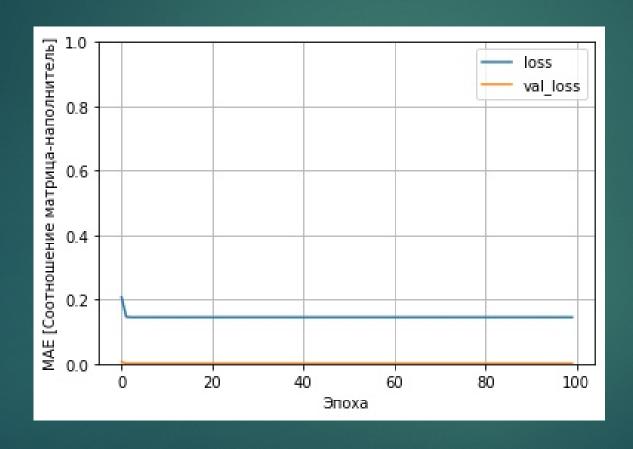


Предсказание для признака «Прочность при растяжении»

Градиентный бустинг



Предсказание для признака «соотношения матрица-наполнитель»



Удалённый репозитарий

▶ Репозитарием проекта избран сервис GitHub.

- Ссылка на репозитарий ВКР:
- ► https://github.com/vsevolod008/bmstu_qualifying_work
- https://gitlab.com/vsevolod008/bmstu_vkr

Заключение

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

Спасибо за внимание