ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ

РАБОТА

по курсу

«Data Science»

ТЕМА: ПРОГНОЗИРОВАНИЕ КОНЕЧНЫХ СВОЙСТВ НОВЫХ МАТЕРИАЛОВ (КОМПОЗИЦИОННЫХ МАТЕРИАЛОВ)

Слушатель: Новиков А.А.

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 папке «hw_data_composite» с данными 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
    Угол нашивки, град
                                                       int64
                                                        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

| df_all.describe() | | | | | | | | | | | |
|-------------------|--|---------------------|-----------------------------|-----------------------------------|---------------------------------------|--------------------------------|--------------------------------------|--|--|----------------------------|---------------------|
| | 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%).

Используемые библиотеки

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

МОДЕЛИ МАШИННОГО ОБУЧЕНИЯ

sklearn.neighbors.KNeighborsRegressor – регрессия на основе k ближайших соседей. Цель прогнозируется путем локальной интерполяции целей, связанных с ближайшими соседями в обучающем наборе.

sklearn.linear_model.LinearRegression — обычная линейная регрессия методом наименьших квадратов. Соответствует линейной модели с коэффициентами w = (w1, ..., wp), чтобы минимизировать остаточную сумму квадратов между наблюдаемыми целями в наборе данных и целями, предсказанными линейным приближением.

sklearn.svm.SVC – С-классификация опорных векторов. Реализация основана на libsvm. Время подгонки масштабируется квадратично с количеством выборок и может быть непрактичным за пределами десятков тысяч выборок. Поддержка мультиклассов осуществляется по схеме «один против одного».

sklearn.ensemble.GradientBoostingRegressor – используется для повышения градиента для регрессии. GradientBoostingRegressor строит аддитивную модель поэтапно вперед; он позволяет оптимизировать произвольные дифференцируемые функции потерь. На каждом этапе дерево регрессии аппроксимируется отрицательным градиентом заданной функции потерь.

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. Были построены графики попарного рассеяния переменных, а также была определена корреляция между переменными
- ▶ 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.

Удалённый репозитарий

▶ Репозитарием проекта избран сервис GitHub.

- Ссылка на репозитарий ВКР:
- ► https://github.com/zvhzvh/VKR

Заключение

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

Спасибо за внимание