Предметом выпускной квалификационной работы являются построение моделей для прогнозирования таких характеристик композиционных материалов, как модуль упругости при растяжении, прочность при растяжении и создание нейронной сети для рекомендации соотношения матрица-наполнитель.

Данные состоит из двух файлов первый из которых – это файл с данными о параметрах базальтопластика, а второй файл с данными о нашивках из углепластика.

Потребовалось объединить оба файла по индексу с типом объединения INNER.

После объединения количество строк 1023, столбцов 11

Полученные данные являются отправной точкой дальнейших исследований, требуют предварительный обработки.

В качестве инструментов разведочного анализа используется оценка статистических характеристик данных.

Для визуализации данных необходимо отобразить попарные графики рассеяния точек, гистограммы распределения каждой из переменной и ящика с усами, корреляционную тепловую карту Пирсона.

Из корреляционной тепловой карты Пирсона видно, что все коэффициенты корреляции близки к нулю. Это означает отсутствие линейной зависимости между признаками.

По форме разброса точек попарного графика рассеяния видно, что зависимости между переменными, на которых будет основываться работа модели, не обнаруживаются.

Гистограммы распределения переменных показали, что все признаки, кроме «Угол нашивки», имеют нормальное распределение и принимают неотрицательные значения. «Угол нашивки» принимает два значения: 0, 90 градусов.

При построении графика boxplot (ящик с усами) выявлено 25 выбросов методом 3-х сигм и 94 выброса методом межквартильных расстояний.

Для удаления выбросов используется метод 3х сигм так как при использовании этого метода исключается меньше полезной информации.

После удаления выбросов, проведя анализ данных видно, что значения находятся в разных диапазонах.

Для обучения моделей для прогноза модуля упругости при растяжении и прочности при растяжении необходимо данные нормализовать так как многие алгоритмы машинного обучения работают лучше, когда числовые входные переменные масштабируются до стандартного диапазона.

Для масштабирования данных используется метод MinMaxScaler который масштабирует все функции данных в диапазоне [0, 1].

Для прогноза модуля упругости при растяжении и прочности при растяжении использованы следующие модели: LinearRegression, SGDRegressor, Ridge Regression, LassoRegressor, BayesianRidge, DecisionTreeRegressor, LinerSVR, SVR, KNeighborsRegressor, AdaBoostRegressor, BaggingRegressor, ExtraTreesRegressor, GradientBoostingRegressor, RandomForestRegressor.

Данные разделены на тестовую и обучающую выборки: 30% данных оставлены на тестирование модели, на остальных данных происходит обучение моделей.

Для поиска наиболее подходящих критериев модели создана функция GridSearchCVCounter которая с помощью GridSearchCV выявляет наиболее подходящих критериев переданной модели. А также создана функция ErrorCounter которая считает метрики MAE Средняя абсолютная ошибка, MSE Средняя квадратичная ошибка, R2 Коэффициент детерминации.

Тестирование моделей. На графике видно, что все модели с использованием кросс-валидации показали неудовлетворительный результат. Для оценки качества регрессии использовался коэффициент детерминации.

Данный показатель является статистической мерой согласия, с помощью которой можно определить, насколько уравнение регрессии соответствует реальным данным. Коэффициент детерминации изменяется в диапазоне от 0 до 1. На графике видно, что коэффициент детерминации приближен к 0, это означает, что связь между переменными регрессионной модели отсутствует.

Модель DecisionTreeRegressor показала большой разброс и отрицательные значения результатов кросс-валидации. Это означает, что предсказания, сделанные данной моделью хуже, чем оценки на основе среднего значения.

Оригинальные и предсказанные значения по каждой модели прогноза модуля упругости и прочности при растяжении. По графику видно, что модели показали близкое значение к усредненному значению по выборке. Разработанные модели не дают достоверный прогноз

По диаграмме рассеяния видно, что корреляция между оригинальными значениями и предсказанными отсутствует

Результат кросс-валидация для каждой модели прогноза модуля упругости и прочности при растяжении

На графике видно, что все модели показали неудовлетворительный результат

Для рекомендации соотношения матрица-наполнитель написана нейронная сеть. Данные разделены на тестовую и обучающую выборки: 10% данных оставлены на тестирование модели, на остальных данных происходит обучение моделей.

Для обучения нейронной сети используется 100 эпох. Используя функцию EarlyStopping обучение остановится, когда val\_loss перестанет улучшаться. Нейронная сеть остановилась на 7 эпохе, так как с 3 эпохи val\_loss перестал улучшаться.

График оригинальные и предсказанные значения нейронной сети показывают близкое значение к усредненному значению по выборке.

По диаграмме рассеяния видно, что корреляция между оригинальными значениями и предсказанными отсутствует.

Демонстрация более простой модели.

График оригинальные и предсказанные значения нейронной сети показывают близкое значение к усредненному значению по выборке

По диаграмме рассеяния видно, что корреляция между оригинальными значениями и предсказанными отсутствует.

Нейронная сеть с 2 слоями показала эффективность не хуже, чем нейронная сеть с большим количеством слоев.

Разработаем приложение с интерфейсом командной строки, которое будет выдавать прогноз значения «Соотношение матрица-накопитель»

Разработаем приложение с интерфейсом командной строки, которое будет выдавать прогноз значения «Соотношение матрица-накопитель»

* С помощью keras.models.load\_model() загружена сохраненная модель.
* В цикле программа запрашивает входные данные для нейронной сети.
* После получения необходимых входных данных модель предсказывает искомое значение.