

Практика по теме 1.3. Методы машинного обучения (часть 1)

Результат выполнения каждого пункта должен быть выведен на экран. В коде должны присутствовать комментарии и выводы.

Пункты задания:

1. Постройте и оцените модель регрессии по датасету 1.

– Выполните разведочный анализ данных (EDA): выведите первые и последние строки датасета, тип данных, размер датасета, проверьте наличие пропусков, категориируйте нечисловые признаки, рассчитайте описательную статистику, выведите на экран тепловую карту на основе таблицы корреляции, постройте график попарных значений признаков.

Сделайте вывод, между какими признаками есть взаимосвязь.

– Выполните стандартизацию признаков.

– Обучите модель линейной регрессии: в качестве целевого признака возьмите признак «Weight», разбейте датасет на обучающую и тестовую выборку, обучите модель линейной регрессии, выведите на экран коэффициенты модели линейной регрессии.

– Оцените качество модели на обучающей и тестовой выборке (метрики качества: MSE, R2, MAE, MAPE).

2. Постройте и оцените модель деревьев решений по датасету 1.

– Обучите модель деревьев решений: в качестве целевого признака возьмите признак «Species», разбейте датасет на обучающую и тестовую выборку, обучите модель деревьев решений при различных значениях гиперпараметров модели (минимум, два разных варианта **max_depth**).

– Оцените качество модели на обучающей и тестовой выборке (метрики качества: Accuracy, Precision, Recall, F1 Score) для различных вариантов.

– Сделайте вывод, модель с какими параметрами дает лучший результат.