**Модель прогнозирования цены недвижимости**

**Обзор проекта**

Проект направлен на разработку модели прогнозирования цены на недвижимость на основе различных характеристик объектов. В работе используются несколько алгоритмов машинного обучения для обеспечения надёжности и точности предсказаний. Данные для проекта взяты из обширного набора данных о недвижимости, содержащего разнообразные признаки объектов.

**Признаки**

Для предсказания используются категориальные переменные, такие как штат, количество этажей и тип недвижимости, которые обрабатываются и кодируются для использования в моделях машинного обучения.

• state: Штат, в котором находится недвижимость.

• stories: Количество этажей в объекте недвижимости.

• propertyType: Тип недвижимости, например, квартира, кондоминиум и т.д.

• target: Цена объекта недвижимости, которая является целевой переменной для предсказания

**Модели**

В этом проекте используются следующие модели:

• CatBoostRegressor: Модель градиентного бустинга, которая может нативно работать с категориальными признаками.

• RandomForestRegressor: Ансамблевая модель, использующая алгоритм случайного леса, интегрирована в конвейер с предобработкой для One-Hot Encoding.

• LinearRegression: Базовая модель для оценки производительности.

**Визуализация**

• Гистограммы и диаграммы размаха используются для понимания распределения целевой переменной и числовых признаков.

• Pairplot помогает визуализировать взаимосвязи между числовыми признаками.

Использование

Для запуска предсказательных моделей выполните предоставленный скрипт Python, убедившись, что набор данных находится по указанному пути.

**Зависимости**

• pandas

• numpy

• sklearn

• matplotlib

• seaborn

• CatBoost

**Предобработка данных**

• Обработка пропущенных значений путём замены на заполнитель 'missing'.

• Кодирование категориальных переменных с использованием метода One-Hot Encoding.

• Разделение данных на обучающую и тестовую выборки для валидации модели.

**Оценка**

Модели оцениваются на основе среднеквадратичной ошибки (MSE) и коэффициента детерминации (R2) для определения точности и эффективности предсказаний.

**Визуализация**

• Гистограммы и ящики с усами используются для понимания распределения целевой переменной и числовых признаков.

• Парные графики помогают визуализировать взаимосвязи между числовыми признаками.

**Использование**

Для запуска предсказательных моделей выполните предоставленный скрипт Python после того, как убедитесь, что набор данных находится в указанном пути.