House Price Prediction Project

Project Overview

This project presents a machine learning–based house price prediction system built using the Ames Housing dataset. The goal was to create a clean, interpretable, and production-ready model suitable for real-world deployment and portfolio demonstration rather than leaderboard overfitting.

Dataset & Context

The Ames Housing dataset includes detailed residential property information such as size, quality, amenities, and sale conditions. The task is a regression problem predicting the final sale price of homes.

Preprocessing & Feature Engineering

Categorical features were encoded using one-hot encoding, missing values were handled systematically, and SalePrice was log-transformed to improve learning stability. Identifier columns were removed to prevent noise.

Modeling Strategy

Random Forest was used as a strong baseline. XGBoost Regression was selected as the final model due to superior validation performance and better generalization.

Performance Summary

The final model achieved a Kaggle RMSE score of approximately 0.13. This reflects a solid and well-regularized model appropriate for practical use.

Limitations & Scope

Predictions are based on historical data and cannot capture real-time market changes. Results should be interpreted as informed estimates rather than exact values.

Future Enhancements

This project intentionally serves as a strong base model. Further improvements may include advanced tuning, additional features, and external data integration. The developer is fully capable of implementing these enhancements.

Deployment & Application

The trained model is deployed using Streamlit as an interactive web application, allowing users to input property features and receive instant price estimates.

Developer Notes

This project demonstrates strong skills in data science, machine learning, and deployment practices, making it suitable as a professional portfolio project.