

Predictive Modeling for Accurate Housing Price Estimation

I Hua Tsai¹, Vamsi Krishna Yarragunta¹, Nagalakshmi Munapati¹, Harshavardhini Katta², Varshika Indukuri¹

¹Department of Data Science & Artificial Intelligence, University of Central Missouri

²Department of Computer Science, University of Central Missouri

²Department of Computer Science and Cybersecurity, University of Central Missouri

Email : tsai@ucmo.edu

Abstract— This project presents a machine learning–based House Price Prediction system designed to estimate residential property values using structured real-estate attributes. The objective is to analyze and compare traditional regression techniques with a deployed inference pipeline capable of generating real-time predictions. In this setup, the user interacts with a Flask-based application, providing property details such as square footage, number of bedrooms, bathrooms, lot area, and additional structural features. The backend model processes this input using a trained regression algorithm, producing an estimated market value without requiring manual appraisal or domain expertise.

The system exemplifies an end-to-end supervised learning workflow—covering data preprocessing, feature engineering, model training, evaluation, and deployment—where the trained model relies purely on historical housing data to infer pricing patterns. The experiment highlights the contrast between manual valuation, which is often subjective and inconsistent, and an automated model that offers faster, more objective, and more scalable predictions. Results show that the regression model achieves strong performance with low error rates, reinforcing its effectiveness for practical real-estate analytics. This work demonstrates the growing role of data-driven decision-making and predictive modeling in modern housing market applications.

Keywords- *House Price Prediction, Regression Models, Machine Learning, Supervised Learning, Real-Estate Analytics, Feature Engineering, Predictive Modeling, Flask Deployment*

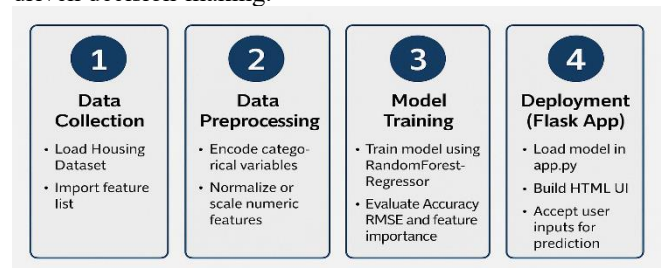
I. INTRODUCTION

House price prediction is a core problem in real-estate analytics, focused on estimating the market value of residential properties based on measurable characteristics such as size, location, structural attributes, and neighborhood features. Housing datasets typically include a mix of numerical and categorical variables, making them a strong benchmark for evaluating predictive modeling techniques due to their complexity, availability, and real-world relevance.

Accurate price estimation plays an important role in property valuation, mortgage lending, investment analysis, and fair pricing decisions. Traditional statistical methods such as Linear Regression and Decision Trees have been widely used, but modern machine-learning approaches—including ensemble models and feature-optimized regressors—offer significantly improved performance by capturing non-linear relationships and interactions between features.

Recently, end-to-end machine learning systems have become increasingly popular in real-estate applications, enabling automated insights through data preprocessing, feature engineering, and model deployment. These systems can ingest heterogeneous data sources, learn pricing patterns, and generate instant predictions with high consistency. However, challenges such as data noise, feature imbalance, and market variability require careful handling to ensure robust model performance.

This work applies supervised learning techniques to build a complete house price prediction pipeline—covering dataset preparation, model training, evaluation, and deployment through a Flask interface. The goal is to analyze the effectiveness of regression-based models in predicting house prices and evaluate their usability for practical, real-estate-driven decision-making.



II. SYSTEM ARCHITECTURE

A. Input Stage

Property features such as square footage, bedrooms, bathrooms, lot size, and location details are loaded from the housing dataset or entered manually through the Flask interface.

B. Processing Stage

The input values are cleaned, encoded, and scaled. A structured feature vector is generated and passed to the trained regression model loaded from model.pkl.

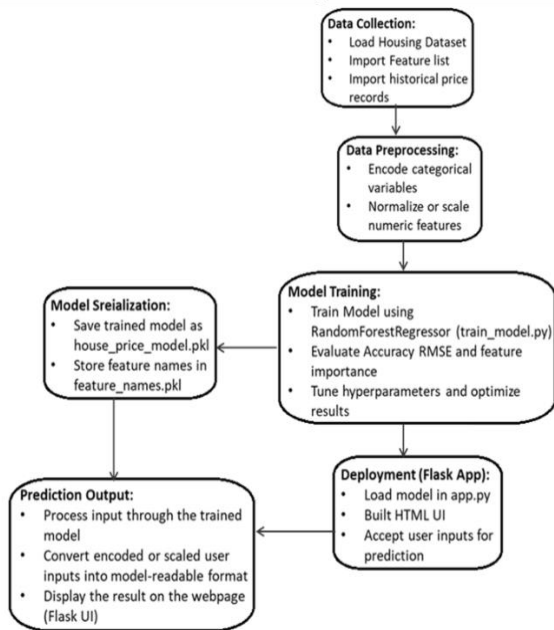
C. Inference Stage

The machine-learning model performs supervised regression inference, using learned feature relationships to predict the house price.

D. Output Stage

The predicted price is displayed through the Flask web interface. Outputs may also be logged or compared with actual market values for evaluation.

III.METHODOLOGY



IV.TOOLS AND TECHNOLOGIES

A. Software

Python, Flask, and Jupyter Notebook form the primary development environment. VS Code is used for coding, testing, and project organization.

B. Python Libraries

Pandas, NumPy, Scikit-learn, Matplotlib/Seaborn for modeling and analysis; Flask for deployment; Pickle/Joblib for model serialization.

C.Hardware

CPU-based execution fully supported; no GPU requirement. Standard laptop or desktop hardware is sufficient for training and inference..

V.OUTPUT

The screenshot shows a web form with the following fields and values:

- State: Kansas
- Median Income (in tens of thousands): 7
- House Age (years): 10
- Average Rooms per Household: 5.0
- Average Bedrooms per Household: 3
- City Name: Overland Park
- Street Name: Louis
- Average Occupancy: 8

A blue button labeled "Predict House Price" is at the bottom of the form. Below the button, a green box displays the "Predicted House Price" as **\$152,683.17** (1.5268 in hundreds of thousands).

- The system allows rapid scenario testing by modifying inputs and instantly receiving updated predictions.
- The final output provides an accessible, real-time estimation tool that demonstrates the practical usability of the trained machine-learning model.

VI.DISCUSSION



• Challenge:

Data inconsistency and market variability can reduce prediction accuracy.

• Mitigation Strategies

Robust preprocessing, cross-validation, and periodic model retraining improve stability and reliability..

VII.CONCLUSION

This study demonstrates the effectiveness of supervised machine learning models for residential house price prediction. While traditional regression techniques remain efficient and reliable for structured real-estate data, the deployment pipeline enables fast, consistent, and user-friendly price estimation through a Flask interface. The results highlight the growing role of data-driven modeling in property valuation and reinforce the value of end-to-end ML systems capable of delivering practical decision-support in real-world housing markets.

VIII.REFERENCES

- [1] R. Cellmer and K. Kobylńska, "Housing Price Prediction — Machine Learning and Geostatistical Methods," Real Estate Management and Valuation, 2025. <https://doi.org/10.2478/remav-2025-0001>
- [2] H. Jiang, "House Price Prediction with Optimistic Machine Learning Methods Using Bayesian Optimization," 2024. <https://www.scitepress.org/Papers/2024/128254/128254.pdf>
- [3] Z. Stojanović et al., "Predicting Real Estate Prices Using Machine Learning in Bosnia and Herzegovina," Data, 2025. <https://www.mdpi.com/2306-5729/10/9/135>

□