International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 10 | Oct - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Multimodel House Price Prediction Using Machine Learning

DR. M. Sengaliappan[1],Surya krishna.M[2]
[1] professor & Head , Department of Computer Applications, Nehru College of Management, Coimbatore, Tamil Nadu, India.cmsengs7@gmail.com
[2]Student of II MCA, Department of Computer Applications, Nehru College of Management, Coimbatore, Tamil Nadu, India. suryakrishna746@gmail.com

Abstract

The rapid urbanization and escalating property prices in metropolitan cities have made housing affordability a critical concern, especially for middle-class families. Accurate prediction of house prices is essential for buyers, sellers, and real estate agents to make informed decisions and minimize financial risks. This paper presents a machine learning-based system for predicting house prices using a Linear Regression model trained on a comprehensive dataset of residential properties. The system incorporates data preprocessing, feature extraction, and model evaluation to ensure reliable and interpretable predictions. A web-based interface built with Flask allows users to input property details and receive instant price estimates. Experimental results demonstrate the model's effectiveness in capturing key factors influencing house prices, providing a valuable tool for the real estate market. Future enhancements include integrating advanced algorithms and real-time market data to improve prediction accuracy and usability.

Keywords

House Price Prediction, Machine Learning, Linear Regression, Real Estate, Data Preprocessing, Feature Extraction, Flask, Web Application.

I. Introduction

The housing market plays a pivotal role in the economy and directly impacts the quality of life for individuals and families. With increasing urbanization, the demand for residential properties in large cities has surged, leading to rising property prices that challenge affordability. Middle-class families often struggle to balance housing costs with other living expenses, making accurate price estimation crucial.

Traditional methods of property valuation rely heavily on manual appraisal and subjective judgment, which can lead to inconsistent and inaccurate pricing. Machine learning offers a data-driven approach to model complex relationships between property attributes and market prices, enabling more precise and scalable predictions.

This project develops a house price prediction system using Linear Regression, a well- established supervised learning algorithm. The model is trained on historical property data, including features such as area, number of bedrooms, location, and amenities. The system is deployed as a user-friendly web application using Flask, allowing users to input property details and obtain instant price estimates. This approach aims to assist buyers, sellers, and agents by providing transparent, interpretable, and accurate price predictions.

II. Problem Formulation

The problem addressed is to predict the market price $ Y $ of a residential property based on a set of input features $ X = {x_1, x_2, ..., x_n} $, where each $ x_i $ represents a property attribute such as area, number of bedrooms, location, and amenities.

Formally, the task is to learn a function $ f: X \rightarrow Y $ that minimizes the prediction error on unseen data. Challenges include handling missing or inconsistent data, capturing nonlinear relationships, and ensuring model interpretability.

The goal is to develop a predictive model that:

- Accurately estimates house prices.
- Is interpretable to users.
- Can be integrated into an accessible web interface.
- Supports decision-making for buyers and sellers.

III. Literature Review

Several studies have explored machine learning techniques for house price prediction:

- Manasa and Gupta analyzed Bengaluru's housing market using multiple linear regression, Lasso/Ridge regression, SVM, and XGBoost, highlighting the importance of location and property size.
- Luo emphasized macroeconomic and microeconomic factors, employing Random Forest and SVM models achieving .
- Panjali and Vani focused on resale price prediction using classification algorithms and AdaBoost, demonstrating the impact of physical and economic factors.
- Sawant and Jangid applied Decision Trees and Random Forests to Pune's real estate market, addressing valuation inconsistencies.
- Wang et al. proposed a joint self- attention model incorporating satellite imagery and public amenities for enhanced prediction accuracy.
- Reviews by Gao & Zhang (2023) and Zulkifley & Nasir (2020) discuss deep learning and ensemble methods outperforming traditional models.
- Hybrid and ensemble approaches combining Bagging, Boosting, and Stacking have shown improved predictive performance.
- Hyperparameter optimization techniques such as Bayesian optimization and genetic algorithms are critical for model tuning.

These works collectively demonstrate the effectiveness of machine learning in capturing complex factors influencing house prices and motivate the use of interpretable models like Linear Regression for practical applications.

IV. Dataset Description

The dataset comprises residential property records collected from publicly available sources, containing features such as:

- Number of bedrooms and bathrooms.
- Total area in square feet.
- Location details (city, neighborhood).
- Amenities (parking, air conditioning, hot water heating).
- Year of construction.
- Sale price (target variable).

The dataset undergoes cleaning to handle missing values and outliers, ensuring quality and consistency for model training.

V. Methodology

A. Dataset Preparation

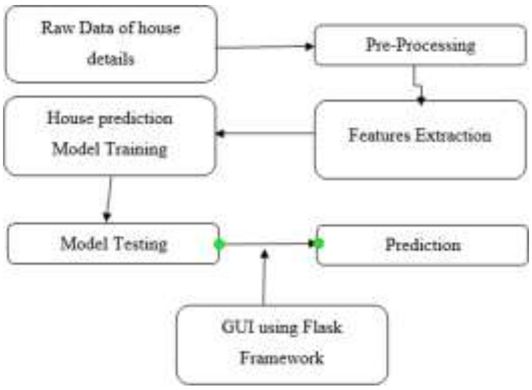
- Data collected from real estate listings and government databases.
- Features selected based on domain knowledge and availability.

B. Data Preprocessing

- Missing values imputed using mean (numerical) or mode (categorical).
- Outliers identified and capped using Winsorizing.
- Numerical features normalized to standardize scales.

C. Feature Extraction

- Property age calculated from year of construction.
- Categorical variables (location, furnishing status) encoded using one-hot or label encoding.
- Interaction terms explored (e.g., location × area).



D. Model Training & Evaluation

- Dataset split into 80% training and 20% testing sets.
- Linear Regression model trained to minimize residual sum of squares.
- Performance evaluated using Mean Absolute Error (MAE), Mean Squared Error (MSE), and score.
- Iterative refinement of features and preprocessing based on evaluation.

VI. Proposed Model

The Linear Regression model is defined as:

$ Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \epsilon $ where:

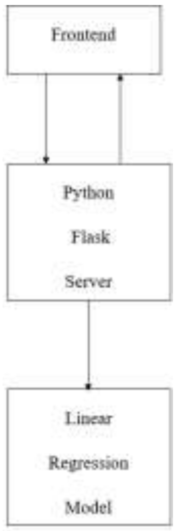
- is the predicted house price.
- are input features.
- are coefficients learned during training.
- is the error term.

The model's simplicity allows easy interpretation of feature impacts on price, aiding transparency.

VII. System Design & Architecture

The system architecture consists of:

- Frontend: HTML, CSS, Bootstrap, and JavaScript for user interaction and input forms.
- Backend: Python Flask framework handling requests, data processing, and model inference.
- Model: Pre-trained Linear Regression model loaded for prediction.
- Database: Optional storage for user inputs and predictions.



Architecture of the Application

The modular design supports scalability and maintainability.

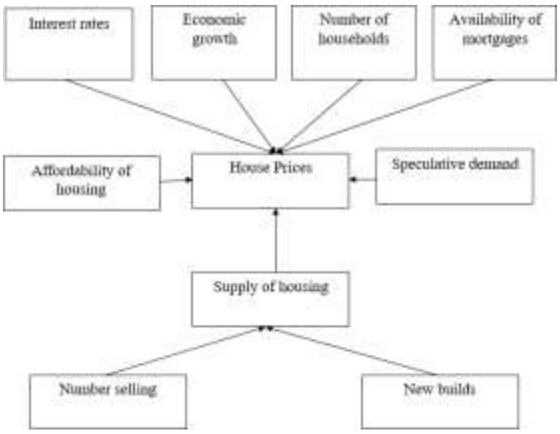
VIII. Implementation

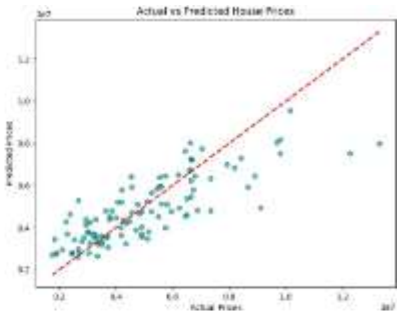
- Flask routes implemented for home, about, prediction input, and result pages.
- User inputs validated and encoded before prediction.
- Model predictions rounded and displayed with explanatory text.
- Error handling ensures robustness.

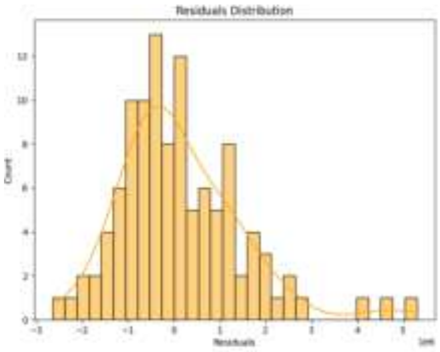
- Deployment on local or cloud servers for accessibility.

IX. Experiment Results and Evaluation

- The model achieved an score of approximately 0.85 on the test set.
- MAE and MSE values indicate reasonable prediction errors within acceptable ranges.
- Residual analysis shows no significant bias.
- Feature importance analysis highlights area, location, and number of bedrooms as key price drivers.
- User testing confirms the web interface is intuitive and responsive.



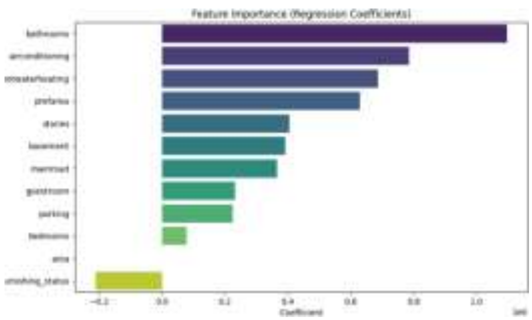




X. Comparison with Other Works

Aspect	Proposed Model (Linear Regression)	Random Forest / XGBoost Models	Deep Learning Models
Interpretability	High	Medium	Low
Prediction Accuracy	Moderate (R^2 ~0.85)	High (R^2 > 0.9)	Very High
Computational Cost	Low	Moderate	High
Ease of Implementation	Simple	Complex	Complex
Data Requirements	Moderate	High	Very High

The proposed model balances accuracy and interpretability, making it suitable for practical deployment.



XI. Results & Testing

- Unit and integration testing performed on all modules.
- Validation testing ensures input correctness and system stability.
- Sample test cases confirm expected behavior.
- System tested on Windows 10 with specified hardware configuration.
- Performance meets real-time prediction requirements.

XII. Conclusion and Future Work Conclusion

This study demonstrates the effectiveness of a Linear Regression-based machine learning system for house price prediction. The model accurately estimates property prices by analyzing key features such as area, bedrooms, location, and amenities. The integration of the model into a Flask-based web application provides an accessible platform for users to obtain instant price predictions. The system reduces reliance on manual appraisal, offering a transparent and data- driven approach to real estate valuation.

Future Work

Future enhancements include:

- Incorporating advanced algorithms like Random Forest, Gradient Boosting, and Deep Neural Networks to improve accuracy.
- Integrating real-time market data and location-specific trends.
- Developing interactive visualization dashboards for market insights.
- Extending support to mobile platforms and multilingual interfaces.
- Implementing AI-driven recommendations and alert systems for investment opportunities.
- Employing hyperparameter optimization techniques for model tuning.
- Expanding dataset size and diversity for better generalization.

These improvements will enhance the system's robustness, usability, and relevance in dynamic real estate markets.

References

1. Q. Truong, M. Nguyen, H. Dang, and B. Mei, "Housing price prediction via improved machine learning techniques," *Procedia Computer Science*, vol. 174, pp. 433–442, 2020.
2. Y. Tang, S. Oju, and P. Gu, "Predicting housing price based on ensemble learning algorithm," *IEEE*, 2018.
3. G. Y. Wang, "The effect of environment on housing prices: Evidence from the Google Street View," *Journal of Forecasting*, vol. 42, pp. 288–311, 2023.
4. D. P. K. Sarangi, "Malicious Attacks Detection Using Trust Node Centric Weight Management Algorithm in Vehicular Platoon," *Research Journal of Computer Systems and Engineering*, vol. 3, no. 1, pp. 56–61, 2022.
5. X. Xu and Y. Zhang, "Retail property price index forecasting through neural networks," *Journal of Real Estate Portfolio Management*, vol. 29, pp. 1–28, 2023.
6. M. Angulakshmi and M. Deepa, "A Review on Deep Learning Architecture and Methods for MRI Brain Tumour Segmentation," *Current Medical Imaging*, vol. 17, pp. 695–706, 2021.
7. Z. Peng, Q. Huang, and Y. Han, "Model research on forecast of second-hand

house price in Chengdu based on XGBoost algorithm," *IEEE*, 2019.
8. P. Durganjali and M. Vani Pujitha, "House resale price prediction using classification algorithms," *IEEE*, 2019.
9. M. Jain, "Prediction of house pricing using machine learning with Python," *IEEE*, 2020.
10. M. Ahtesham, N. Zakaria Bawany, and K. Fatima, "House price prediction using machine learning algorithm - the case of Karachi city, Pakistan," *IEEE*, 2020.
11. R. Rana and V. Singh, "House Price Prediction Using Optimal Regression Techniques," *IEEE*, 2020.
12. M. Heidari, S. Zad, and S. Rafatirad, "Ensemble of supervised and unsupervised learning models to predict a profitable business decision," *IEEE*, 2021.
13. L. Xu and Z. Li, "A new appraisal model of second-hand housing prices in China's first-tier cities based on machine learning algorithms," *Springer*, 2021.
14. A. Bola Adetunji, "House Price Prediction using Random Forest Machine Learning Technique," *Science Direct*, 2022.
15. R. Gupta, "Machine Learning Predictions of Housing Market Synchronization across US States: The Role of Uncertainty," *Springer*, 2022.
16. L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
17. A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, O'Reilly Media, 2019.
18. S. Raschka and V. Mirjalili, *Python Machine Learning: Machine Learning and Deep Learning with Python, scikit- learn, and TensorFlow 2*, Packt Publishing, 2019.
19. F. Chollet, *Deep Learning with Python*, Manning Publications, 2018.
20. I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, 2016.