International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

# HOUSE PRICE PREDICTION USING MACHINE LEARNING

[1] K.Ramya Sree
Department of statistics, Sri Durga Malleswara Siddhartha Mahila Kalasala, Vijayawada.

[2] P.Maheswari
Department of statistics, Sri Durga Malleswara Siddhartha Mahila Kalasala, Vijayawada.

## ABSTRACT

House price prediction is an essential task in real estate analytics that helps buyers, sellers, investors, and financial institutions estimate property values accurately. Traditional pricing methods rely on manual assessment and market comparison, often resulting in inconsistencies and subjective bias. This paper proposes a machine learning–based model for predicting house prices using historical housing datasets. The system analyzes features such as location, area, number of rooms, amenities, and construction year to estimate prices. Multiple regression algorithms including Linear Regression, Decision Tree, Random Forest, Support Vector Regression, and Gradient Boosting are implemented and compared. The performance of the system is evaluated using accuracy, precision, recall, and F1-score metrics. Experimental results demonstrate that ensemble models provide higher prediction accuracy than traditional methods. The proposed approach improves reliability, reduces human effort, and enables real-time property valuation.

## Keywords

Machine Learning, House Price Prediction, Regression Models, Real Estate Analytics, Random Forest, Predictive Modeling

## 1. Introduction

The real estate market plays a significant role in economic development and investment planning. Property prices fluctuate due to multiple factors such as infrastructure, neighborhood development, population growth, and economic conditions. Accurate price estimation is challenging because these factors interact in complex ways.

Machine learning techniques provide an efficient solution by learning patterns from historical data and predicting future values. Unlike traditional methods, machine learning models can process large datasets, identify hidden relationships, and continuously improve prediction performance. This research aims to design and evaluate a machine learning system capable of predicting house prices with high accuracy.

## 2. Existing System

Traditional house price estimation methods include manual appraisal, comparative market analysis, and fixed-rate calculations. These approaches have several limitations:

- Dependence on human expertise
- Time-consuming analysis
- Limited scalability
- Inaccurate predictions for new locations
- Inability to process large datasets

Such drawbacks highlight the need for an automated prediction system capable of handling complex housing data.

## 3. Proposed System

The proposed system uses machine learning regression techniques to estimate property prices automatically. The system consists of data preprocessing, feature selection, model training, prediction, and evaluation modules.

Advantages:

- Automated prediction process
- Reduced human bias
- Higher accuracy
- Fast computation
- Adaptability to new data

## 4. Methodology

A. Data Collection

Housing datasets are collected from real estate listings and public datasets. Attributes include area, bedrooms, bathrooms, location, year built, and amenities.

B. Data Preprocessing

Data cleaning includes handling missing values, removing duplicates, encoding categorical variables, and normalization.

C. Feature Selection

Correlation analysis and feature importance ranking are used to identify influential features.

D. Model Training

Dataset is divided into training and testing sets (80% / 20%). Models are trained on historical data.

E. Prediction

The trained model predicts prices for unseen properties.

F. Evaluation

Model performance is evaluated using statistical metrics.

**5. Algorithms Used**

1. Linear Regression – Models linear relationship between independent variables and price.
2. Decision Tree Regression – Splits data into branches based on feature conditions.
3. Random Forest Regression – Combines multiple decision trees to improve accuracy and reduce overfitting.
4. Support Vector Regression – Uses hyperplanes for regression analysis.
5. Gradient Boosting – Builds strong predictors from multiple weak learners.

## 6. Performance Metrics

Accuracy = (TP + TN) / (TP + TN + FP + FN)

Precision = TP / (TP + FP)

Recall = TP / (TP + FN)

F1 Score = 2 × Precision × Recall / (Precision + Recall)

Sample Values:

TP = 90, TN = 85, FP = 10, FN = 15

Accuracy = 87.5%

Precision = 0.90

Recall = 0.857

F1 Score = 0.878

## 7. Results and Discussion

Experimental results show that ensemble methods outperform individual models. Random Forest achieved the highest accuracy due to its ability to handle nonlinear relationships and reduce variance. Gradient Boosting also performed well for complex datasets. Linear Regression showed acceptable performance only for smaller datasets with fewer features.

Prediction accuracy improved when dataset size increased, outliers were removed, and relevant features were selected.

## 8. Conclusion

This paper presented a machine learning–based approach for predicting house prices using multiple regression algorithms. The proposed system provides accurate, fast, and automated price estimation compared to traditional methods. Evaluation metrics confirm the effectiveness of ensemble learning models. The system can support real estate decision-making and investment planning. Future work may include integrating deep learning, geographic information systems, and real-time market data to further enhance prediction accuracy.

## References

[1] T. Mitchell, Machine Learning, McGraw-Hill, 1997.
[2] C. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.
[3] L. Breiman, Random Forests, Machine Learning Journal, 2001.
[4] J. Friedman, Gradient Boosting Machine, Annals of Statistics, 2001.
[5] P. Geurts, Extremely Randomized Trees, Machine Learning, 2006.
[6] A. Géron, Hands-On Machine Learning with Scikit-Learn, O'Reilly, 2019.